

Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why All Delays Are Not Evil

Christopher Mayer
The Wharton School
University of Pennsylvania

and

Todd Sinai
The Wharton School
University of Pennsylvania
and NBER

This draft: October 23, 2001

First draft: April 6, 2001

*We are especially indebted to Jeff Butler of the US Department of Transportation for providing the flight data used in this paper. We also wish to thank Betsey Bailey, Jan Brueckner, Judy Chevalier, David Genesove, Ed Glaeser, Richard Golaszewski, Joseph Gyourko, Dorothy Robyn, Joel Waldfogel, Clifford Winston, and seminar participants at the NBER, University of Chicago, University of British Columbia, and the Wharton School for helpful comments. The excellent research assistance of Sam Chandan and James Knight-Dominick is appreciated.

Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why All Delays Are Not Evil

Abstract:

We examine two factors that might explain the extent of air traffic delays in the United States: network benefits due to hubbing and congestion externalities. Airline hubs enable passengers to cross-connect to many destinations, thus creating network benefits that increase in the number of markets served from the hub. Delays are the equilibrium outcome of a hub airline equating high marginal benefits from hubbing with the marginal cost of delays. Congestion externalities are created when airlines do not consider that adding flights may lead to increased delays for other air carriers. In this case, delays represent a market failure. Using data on all domestic flights by major US carriers from 1988-2000, we find that delays are increasing in hubbing activity at an airport and decreasing in market concentration but the hubbing effect dominates empirically. In addition, most delays due to hubbing actually accrue to the hub carrier, primarily because the hub carrier clusters its flights in short spans of time in order to maximize passenger interconnections. Non hub flights at hub airports operate with minimal additional travel time by avoiding the congested peak connecting times of the hub carrier. These results suggest that an optimal congestion tax would have a relatively small impact on air traffic delays since hub carriers already internalize most of the costs of hubbing and a tax that did not take the network benefits of hubbing into account could reduce social welfare.

1. Introduction

Over the last few years, air traffic delays have garnered increasing attention. The year 2000 produced record delays, with more than one-quarter of all flights arriving at their destination at least 15 minutes behind schedule. With infrastructure improvements being years away and conventional wisdom holding that delays are caused by air traffic congestion, proposed policy remedies have focused on economic solutions such as congestion pricing. However, selecting the appropriate remedy depends crucially on what is causing congestion and delays. In this paper, we try to determine the economic underpinnings of those delays.

One potential cause of greater travel times is the classic congestion externality, also known as the “tragedy of the commons.” Under this hypothesis, congestion occurs because most airports allow unlimited landings and take-offs and individual airlines add flights without valuing the fact that their traffic will create delay costs for other airlines.¹ Failure to internalize the true marginal cost of adding a flight leads to over-scheduling at airports and flights being delayed. The standard solution is to use a Pigouvian tax, such as pricing by time of day or the length of a queue, or to restrict traffic and assign property rights by selling ownership of scarce landing slots at congested airports. Previous empirical research has focused on these solutions, suggesting that a congestion tax would have substantial efficiency gains in reducing the level of delays. [Carlin and Park (1970); Morrison and Winston (1989); Daniel (1995); Daniel and Pahwa (2000)]

One problem with the congestion externality explanation for delays, however, is that it is

¹See models in Vickrey (1969) and Arnott (1979) as examples of transport systems with inefficient congestion.

not consistent with the delay pattern across all US airports. In the “tragedy of the commons,” it is usually assumed that there are multiple agents who do not take into account the externality that they create for others. While congestion externalities might explain why airports without a single dominant carrier, such as La Guardia, Los Angeles, JFK, or Boston, should have high delays, this model may not explain why airports that are dominated by one large carrier, such as Philadelphia, Newark, Atlanta, or Detroit, are consistently among the airports with the largest overall delays.²

We propose a second explanation for high air traffic delays: the network benefits associated with the hub and spoke system.³ Just one new round-trip flight from a hub where an airline already connects to n cities will create $2n$ additional connecting routes. Since the number of potential connections increases exponentially in the number of markets served by the hub carrier, the carrier has an incentive to serve an infinite number of markets. These increasing returns to scale are offset by the limited flight capacities of airports, so a hub airline must trade off the increasing benefits of serving additional markets against rising marginal congestion costs due to higher traffic, such as longer connecting times and greater flight delays. According to this simple model, longer delays at hub airports are the efficient equilibrium outcome of a hub airline equating high marginal benefits from hubbing with the marginal cost of delays. In addition, a hub airline with greater demand in outlying “spoke” cities will serve more markets and its hub

²Brueckner (2001) shows that a single dominant carrier will internalize much of the externality that would otherwise lead to greater delays.

³See Economides (1996) for a general explanation of the economics of networks and Saloner and Shepard (1995) for an example of empirical evidence in favor of internalized network benefits in the adoption of ATMs.

airport would have greater equilibrium flight delays.

Because hub airlines have increasing returns to scale only to the extent their passengers can connect between flights, they may have an incentive to cluster their inbound and outbound flights at “hubbing times” to keep connection times between flights short. If this is the case, hub carriers would face most of the cost of their own delays since they would generate a large fraction of the traffic at their hubbing times. Externalities might still exist if delays accrue not only to the hub carrier but also to the other air carriers serving the airport. Yet, non-hub carriers have the option to avoid peak hubbing times altogether and schedule their flights when the hub carrier’s aircraft have left the hub.⁴ Consistent with this hypothesis, flight patterns at many congested hub airports exhibit a sawtooth pattern, with a large peak of hub flights serving the airport at hubbing times, and smaller numbers of flights mainly by non-hub carriers at non-hubbing times.

Using U.S. Department of Transportation data on flights from 1988-2000 by all major air carriers with more than a one percent US market share, over 66 million flights in total, we find that delays are strongly related to hubbing. On average, a flight originating at a hub airport requires up to 7.2 minutes longer to travel to its destination than a flight originating at a non-hub airport. Planes flying to a hub airport are delayed up to 4.5 more minutes, on average. Delays at hub airports are increasing in the size of the hub, defined as the number of markets served by the hub carrier. However, most of the delays at hub airports are incurred by the hub carrier itself.

⁴Borenstein and Netz (1999) and Encaoua *et. al.* (1996) come to very different conclusions, suggesting that competition and demand peaks drive hub and non-hub airlines to cluster their flights at the same departure times. However, neither paper combines the beneficial effects from networking with negative costs of congestion in a system with an endogenous number of flights.

These facts are consistent with hub airlines clustering their flights to maximize the network benefits from passengers connecting between them and non hub airlines scheduling their flights to avoid these peak hubbing times.

The increase in delays associated with hubbing is partially offset by reduced congestion externalities at airports where the hub carrier has a dominant market share. However, the empirical impact of airport concentration, which we use as a proxy for the extent to which delay costs are internalized by the carriers at the airport, is much more modest than for hubbing.⁵ A 20 percentage point increase in airport concentration leads to a 0.3 to 1.2 minute decrease in travel time for all flights at the airport, depending on whether or not we include airport fixed effects. This effect is similar for both arriving and departing flights.

After 1995, we have more detailed data on travel times and are able to decompose the source of delays. All of the additional travel time due to originating at a hub is spent waiting at the gate or in line on a taxiway waiting to take off. If the destination airport is the airline's hub, some of the excess travel time occurs in the air, but the bulk of the additional delay comes from taxiing to the gate or waiting for a gate to become available. In addition, we can reject the hypothesis that hub carrier delays are the result of cascading delays due to late arriving aircraft on the previous inbound flight.

Our primary measure of delay, used in the empirical work described above, is the increase in travel time relative to the minimum feasible time on a route. An alternative definition of delay is schedule delay, or how much later a flight arrives than was promised. Schedule delays may be independent of the actual length of the flights since airlines can reduce

⁵Our definition of airport concentration is a Herfindahl index in market share.

them by choosing later scheduled arrival times. Nonetheless, we also show that airlines do not fully adjust their schedules to offset expected increases in travel time. On average, flights by a hub carrier arrive later relative to their scheduled arrival times than equivalent flights on non-hub carriers. A hub carrier's flights that originate at its hub are up to 6.7 percentage points less likely to arrive "on time," or within 15 minutes of schedule, and its flights that arrive at its hub are up to 1.5 percentage points less likely to be on time.

Alternative views of hub and spoke economics typically emphasize market power or economies of scale rather than the network effects we find. Previous empirical work has shown that hubbing gives the dominant hub carrier significant market power on non-stop flights to and from the hub airport.⁶ Some papers attribute the market power associated with hubs to barriers to entry imposed by a dominant airline, such as frequent flyer programs or computer reservation systems. Others argue that airlines benefit from economies of density, so that marginal costs decrease with number of markets served and the scale of service on those routes. [Brueckner *et. al.* (1992), Brueckner and Spiller (1994), Caves *et. al.* (1984)] Hubs may also increase the economic efficiency of an airline's operations. [Hendricks, Piccone, and Tan (1995, 1997); Brueckner and Zhang (2001)] While market power and cost efficiencies are important factors in hub and spoke networks and could explain some delays at hub airports, neither explains why, in the absence of increasing returns to network connections, the hub carrier would accept high

⁶See Borenstein (1990, 1991, 1992, 1993), Borenstein and Rose (1994), Hergott (1997), Kahn (1993), Kim and Singal (1993), Singal (1996), and Zhang (1996) for a discussion of the impact of hubs and having a dominant carrier at an airport on fares.

delays on its own hub flights relative to non-hub carrier flights to or from the same airport.⁷

The next section discusses the impact of network benefits and congestion externalities on the scheduling decisions of a hub and non hub carrier and the resulting impact on air traffic delays. Section 3 describes the data and discusses our measures of delay. Section 4 presents our empirical specification and results and Section 5 concludes with a policy discussion and an agenda of future research.

2. Hubbing, Network Benefits, and Flight Delays

In this section, we illustrate how network benefits, in the form of increasing returns to scale from hubbing, and congestion externalities due to airlines not internalizing the effect of their flights on others', lead to greater delays. We also present a series of graphs of scheduled flights at Dallas-Fort Worth airport as an example of scheduling practices by hub and non-hub carriers.

We use a simple model to generate four basic empirical predictions. First, hub airports should be more congested than non-hub airports since hub airlines have higher marginal benefits of additional flights and thus are willing to accept greater marginal delay costs. These higher marginal benefits come from increasing returns to scale through network effects in hubbing.

While a new flight on any route will create non-stop passenger traffic, a new flight to or from a

⁷While large enough declines in average cost with additional markets would generate a positive correlation between markets served and willingness to accept delays, it is inconsistent with hub carriers choosing to concentrate their flights at hubbing times. A carrier that was concerned with gaining low costs associated with serving additional cities, but was not interested in network benefits from connections, would evenly space its flights over the day to reduce congestion costs. Monopoly power raises the benefits of serving all cities, but the marginal benefit of serving any additional city still declines without considering network benefits.

hub at a hubbing time will generate additional passengers because of the opportunity to travel to all of the other airports served by the hub carrier. As the number of possible connections by a hub carrier at its hub increases, the value of an additional flight to or from that hub rises commensurately. An airline that serves n markets from its hub will add $2n$ connections with each new round-trip flight and is thus willing to accept greater delays to add such a flight than an airline that serves fewer markets.

Second, the bulk of the delays at the hub airports should be borne by the hub airlines' flights. Since hub airlines gain increasing returns to scale in connections, they have an incentive to cluster their flights close in time to maximize cross-connections by passengers and keep connection times short. Non-hub airlines, which don't have a connection benefit, schedule their flights to avoid the congested hubbing times but do not fully offset the hub airline's clustering. These behaviors lead to peak loads of flights, and hence delays.

Third, hub airlines will have greater delays for departures than arrivals. All airplanes must be on the ground at the same time so that passengers can connect with all possible outgoing flights. An airline facing high congestion costs from clustering flights will smooth its arrivals, as long as the time passengers spend waiting for their connections is not too high. Nonetheless, once the connections have occurred, all the flights are ready to leave at the same time, leading to significant clustering of departures from a hub airport. Smoothing departures does not generate any benefits, because all of the passengers are already on board waiting to depart.

Finally, we show that airlines' failure to internalize the delays caused to other airlines' flights can lead to overscheduling of the airport. Each airline only takes into account the cost of delays for its own flights, neglecting the fact that scheduling a flight may impose a delay on the

flights of other airlines. Thus airports that are more concentrated should exhibit fewer delays since the airlines that fly there each internalize a greater share of the delays.

To show how these empirical predictions can arise, we postulate a simple model of one airport with two airlines and two consecutive time periods.⁸ In each period, flights arrive (A) and depart (D), with departures occurring after arrivals. Total number of flights (F) in a period equals arrivals plus departures: $F = A + D$. There are two types of airlines at this airport. A hub carrier, denoted by subscript H , uses the airport as a connecting point for passengers from outlying airports. A non-hub airline, N , offers only point-to-point service. An airline of type i (N or H) maximizes its profits as follows:

$$\begin{aligned} \max_{A_{i,1}, D_{i,1}, A_{i,2}, D_{i,2}} \quad & B(D_{i,1}) + B(D_{i,2}) \\ & - (A_{i,1} + D_{i,1}) \cdot L(F_1 / K) - (A_{i,2} + D_{i,2}) \cdot L(F_2 / K) \\ & - (A_{i,1} / D_{i,1}) \cdot M(D_{i,2}) \end{aligned}$$

Flights for airline i must meet two equilibrium conditions: 1) All aircraft that arrive at an airport must eventually leave, so total arrivals in periods one and two must equal total departures ($D_{i,1} + D_{i,2} = A_{i,1} + A_{i,2}$) and 2) aircraft must arrive before they depart so arrivals in period one must be greater than or equal to departures in period 1 ($A_{i,1} \geq D_{i,1}$). Condition (2) allows airlines to have some flights arrive in period one and depart in period two. Any passenger who arrives in period two can connect only to a departure in that period. Passengers who arrive in period one can connect to a departure in either period.

⁸From a casual inspection of the data, these time periods correspond to 90-minute blocks, which appears to be the typical time between hub peaks. Over the day, an airport may have as many as eight hub peaks. One could easily extend the model to eight periods but the insights would remain the same.

The first line of the maximization equation describes the benefits to each carrier from a given number of departures. For the hub carrier, hubbing generates increasing returns to scale in a given period ($BM_H(D_{H,t}) > 0, BQ_H(D_{H,t}) > 0$). This condition implies that the marginal benefit from an additional flight is increasing in the number of hub departures in that period. Hubbing benefits are measured in terms of departures as that is the number of possible connections available to hub passengers in a given period. The non-hub airline, subscripted with N , does not obtain any hubbing benefits at this airport, so its marginal benefit function is either flat or decreasing in the number of flights in a given period ($BM_N(D_N) > 0, BQ_N(D_N) \neq 0$).

The second line in the maximization equation describes the expected delay cost in each period. This delay cost increases in the aggregate number of scheduled flights at the airport by both carriers. We assume for simplicity that delays do not spill over into other flying periods. In this illustration, the airport has a capacity, K , which is the number of flights the airport can handle even in bad weather without any expected delays in a given period. Once total flights (F), including arrivals and departures by the hub and non-hub carriers, exceeds capacity K , marginal delay costs begin to rise ($LMF - K > 0, LQF - K > 0$, if $F > K$). Delays affect all other flights scheduled during that time period.

Airline i perceives the marginal cost of an additional flight as equal to the direct marginal delay cost ($LMF - K$) plus the additional delay faced by its own flights in that time period ($LQF - K$)*($A_i + D_i$).⁹ Airlines do not consider the impact of an additional flight on the delay cost of other carriers, leading them to underestimate the marginal social cost of an additional flight. In

⁹Presumably there is also a marginal resource cost for each flight. We assume it to be constant and thus uninteresting, so we suppress it to focus more on hubbing.

equilibrium, the airport has “too many” flights relative to the social optimum, leading to excessive delays. It is immediately apparent that conditional on their number of flights, airlines with low market shares, $(A_i+D_i)/F$, will internalize less of the delay externality than high market-share airlines since less of the increase in average delay due to an additional flight accrues to their aircraft. At the airport level, this implies that more concentrated airports should be less overscheduled.

The final term represents the cost of connecting across periods. We assume that passengers prefer to depart in the same period that they arrive rather than arriving in period one and leaving in period two. Thus, the hubbing benefit for each long connection is reduced at rate M to reflect lower passenger demand for the route, for a total cost of $(A_{H,1} - D_{H,1}) \cdot M(D_{H,2})$. Since non-hub carriers do not have connecting flights passengers do not face long connections and $M_N=0$. However, holding an aircraft at the airport across two periods incurs a resource cost since the capital is sitting idle. Thus the non-hub carrier optimally chooses to set the number of departures equal to the number of arrivals ($D_{N,1} = A_{N,1}$ and $D_{N,2} = A_{N,2}$) in a given period to avoid that cost as it receives no hubbing benefits in exchange.

Both hub and non-hub carriers maximize total revenue by setting marginal revenue equal to marginal cost.¹⁰ For exposition, and to make the model more realistic, we focus on equilibria in which hub and non-hub carriers operate a positive number of flights at the airport in each

¹⁰We implicitly assume that the benefits of additional flights and the costs of delays and long connections can be expressed in dollar terms so the airline is maximizing a simplified profit function. Airlines internalize benefits from hubbing such as greater revenue from connecting passengers and higher load factors on connecting flights. Carriers also realize the costs from hubbing in that passenger demand depends on scheduled travel time for connections and on-time performance.

period. For hub carriers this assumption is fairly straightforward. However, this assumption also makes sense for non-hub carriers as well. Non-hub airlines will have very high marginal benefits from their first flight in each period, which typically connects the airline's largest hub to this airport (i.e., $BM(I)$ is typically quite large).¹¹ Subsequently, the marginal benefits from other flights diminish rapidly after the non-hub carrier has enough flights to service its hubs in other cities. When competing with the hub carrier on a route to a city that is not its hub, the non-hub carrier faces a much lower marginal benefit than the hub carrier, who receives strong network benefits. Thus the marginal benefit for non-hub carriers drops quickly (i.e., $BQ_N(D_N) \neq 0$). Finally, we assume that the marginal cost of additional flights rises faster than the marginal benefit for the hub carrier once the airport exceeds capacity.¹²

Given this setup, hub airports will have more traffic and greater delays than non-hub airports of equivalent size and with equal local demand. This conclusion follows from hub carriers having an increasing marginal benefit from additional flights. In addition, holding the size of the airport constant, the extent of delays increases in the demand for hubbing by the outlying markets. To see this result, consider two otherwise identical airports, one located in a market with many surrounding cities with high demand (airport 1) and the other located in an area with a smaller demand for connecting flights (airport 2). This assumption means that

¹¹In fact, as we note below in Table 1, about three-quarters of all flights in the US either originate or land at an airline's own hub.

¹²On one hand, capacity at the airport is fixed, so LM rises quite quickly when $F > K$. On the other hand, when airlines add service to additional markets, each market is smaller than the market before it, so BM falls with additional flights. The assumptions that $LQ(K) > BQ(K)$, $LM(K) = 0$, and $BM(K) > 0$ guarantees that the airport is served by a finite, positive number of flights each period.

$B^1 N_H(D_H) > B^2 N_H(D_H)$ for all values of D_H . Given that the hub carrier will set $B N_H(D_H) = L N_H(D_H)$, $B^1 N_H(D_H) > B^2 N_H(D_H)$ implies that $L^1 N_H(D_H) > L^2 N_H(D_H)$. Also note that in this framework, the non-hub carrier decreases its flights less than one-for one with the increase in hub carrier flights. In other words, in equilibrium, the high demand airport (1) will have more flights and have higher delays than the low demand airport (2), even accounting for the behavioral response of the non-hub carrier.

In addition to selecting the total number of flights, each airline must decide how to schedule their flights during the two periods. We discuss three equilibria, that differ depending on the size of the connecting costs between period one and period two, $M(\cdot)$. The most straightforward case is when the cost of inter-period connections, $M(\cdot)$, is high. Then both carriers will treat periods one and two as independent and choose identical arrival and departure levels in each period. As we will see below, that equilibrium does not appear to describe the behavior of hub airlines at most airports.

The second equilibrium occurs for sufficiently small values of $M(\cdot)$ so that the hub airline faces little or no cost to having passengers wait to connect between periods. In this scenario, the hub airline minimizes delay costs by evenly distributing flights between periods one and two but maximizes connection benefits by lumping all arrivals in period one and all departures in period two.¹³ Facing equal delays in periods one and two, the non-hub carrier will also choose an equal

¹³To demonstrate the stability of this equilibrium, consider what would happen if the airline moved a departure from period two to the first period. The existence of network benefits ($BQ_H(D_H) > 0$) suggests that the marginal benefit of an additional flight in period two, when the hub airline already serves $D_H - 1$ markets, is much higher than the marginal benefit of the first flight in period one (i.e., $BN_H(1) < BN_H(D_H - 1)$). A second alternative would be for the hub airline to shift an arrival from period one to period two. That strategy would reduce the delay cost in period one but increase it by a larger extent in period two ($LQ_H(D_H) > 0$ implies that $L_{H,2}(D_{H,2} + 1)$

number of arrivals and departures in each period. As with the case of very high connection costs, this smoothing equilibrium is empirically less interesting because it does not adequately describe the behavior of flights in most hub airports.

The third and most realistic equilibrium corresponds to a moderate connecting cost and generates an interior solution. As compared to the case with low connection costs, the hub airline will not fully separate its arrivals and departures. Higher connection costs ($M(.)$) offset the increased delay costs from moving some arrivals from the first to the second period, leading the hub airline to have arrivals in both periods. However, strong network benefits continue to encourage the hub airline to cluster its departures only in period two. In this equilibrium, the hub's departures spike in period two while the arrival distribution is more smooth, so total hub flights in period two are greater than in period one. As before, the non-hub carrier will partially offset the behavior of the hub carrier by concentrating its flights in period one and reducing flights in period two. Nonetheless, total delays will still be higher in period two than in period one.

It is important to note that, from the point of view of consumer welfare, the "excess" delays due to network benefits from hubbing are socially optimal and would be chosen by a social planner who faced the same cost and benefit functions of the hub carrier. By scheduling more flights, the hub carrier offers passengers a greater variety of destinations. By compressing departures, the hub carrier increases the number of possible connections relative to an equilibrium in which the hub carrier mixed arrivals and departures in both periods. By loading more of the traffic into period two, the hub carrier reduces connection costs relative to the

$! L_{H,1}(A_{H,1}-I) > L_{H,2}(D_{H,2}) ! L_{H,1}(A_{H,1}) = 0$ when $A_{H,1} = D_{H,2}$.

$M(.)=0$ equilibrium where all arrivals are in period one.

This type of clustering behavior is apparent when one looks at the flight schedule from Dallas-Fort Worth (DFW) airport. Figure 1 plots scheduled flights by hub and non-hub carriers at 15 minute intervals from 6 a.m. to midnight for an arbitrary date, Friday October 20, 2000. DFW has two hub carriers, American and Delta, although American operates the bulk of the flights at the airport. DFW is amongst the most congested airports in the country. Two facts are immediately apparent from this figure: 1) flights are clustered into peaks and 2) most of the clustering is due to the hub carriers who bunch their flights together. In addition, Figure 2 shows that the hub carriers at DFW smooth their arrivals more than departures. Both figures are consistent with the third equilibrium in which hub airlines face a moderate cost of long connections.

This pattern of clustering flights is not evident at non-hub airports. For comparison, Figure 3 plots total flights at Boston Logan Airport (BOS) on the same date. While Delta and US Airways have fairly large market shares at BOS, neither operates a hub at the airport. Total flights at BOS have many more small ups and downs than at DFW, but DFW flights exhibit much more pronounced peak to trough variability than at BOS.

Clustering by the hub carrier leads to peak flight loads at the airport and delays around hubbing times since non-hub carriers do not fully offset the hub's spikes in departures. Figure 4 plots the total density of flights for hub and non-hub carriers at DFW. With the exception of the 9 a.m. and 1:30 p.m. peaks, the non-hub carriers choose travel times that differ from the peaks of the hub carriers. It is also important to note that the size of the hub peaks are much larger in absolute flights than the non-hub peaks. Finally, not only do the non-hub carriers at DFW avoid

scheduling flights at the same time as the hub carriers, the two hub carriers who use the airport also avoid scheduling flights against each other. Figure 5 graphs departure density for American and Delta at DFW. The fact that the major hubs avoid each other's hubbing times suggests that the clustering of hub flights is more likely due to network benefits rather than airlines choosing to schedule at desirable peak times. We explore these hypotheses further in the empirical work that follows.

3. Data

In 1988, the US Department of Transportation began requiring all airlines with at least one percent of all domestic traffic to report flight-by-flight statistics on delays for the top 27 airports in the US.¹⁴ This rule was passed as a result of a public outcry over the growth in air traffic delays in the 1980s. In addition, the major carriers covered by this rule agreed to voluntarily report data on all of their flights to domestic airports. Originally, the data included the scheduled arrival and departure time of the flight, the actual arrival and departure time, whether the flight was canceled or diverted, and the flight number. From 1988-1994, airlines excluded information on flights that were delayed or canceled due to mechanical problems. Beginning in 1995, major carriers began reporting information on all scheduled flights, regardless for the reason for a delay or cancellation.¹⁵ In that year, the data was expanded to include the time spent taxiing from the gate to the runway, actual flight time, time spent taxiing to the gate after landing, and the tail number of the aircraft. Our sample includes 66.4 million

¹⁴A flight is defined as a nonstop segment.

¹⁵We address this issue in more detail in the empirical section.

flights, which is all data over this time period with the exception of flights in 5 months that had substantially missing or corrupted data files.¹⁶

Figure 6 reports annual averages for the most widely reported measure of airline on-time performance: the percentage of flights that arrive within 15 minutes of scheduled arrival time. For the purposes of on-time statistics, canceled and diverted flights are treated as late arrivals. As has often been reported in the popular press, the percentage of on-time flights declined from 81 percent to 74 percent between 1988 and 2000. However, some of this decrease may be due to the omission of late or canceled flights with maintenance problems in earlier years. For example, the percentage of on-time arrivals fell from 81.3 in 1994 to 78.6 percent in 1995, the first year of expanded reporting. Even with an adjustment for reporting differences, however, on-time performance has clearly deteriorated, especially in the last 3 years.

One problem with on-time performance as a measure of true delay is that airlines can manipulate it by adjusting their scheduled flight times to compensate for expected delays. However, the total delay cost to passengers from congestion or hubbing depends on how much additional travel time these factors impose. Thus we construct a measure of delay that is unaffected by airline scheduling: actual travel time minus minimum feasible travel time. Minimum travel time is defined as the shortest observed travel time on a given nonstop route in a particular month. We consider the minimum feasible time to be a useful benchmark for what travel time would be if airports were sufficiently uncongested and weather were equally favorable. This statistic controls for possible changes over time in the types of routes flown or in

¹⁶ The missing months are July and August 1993, March 1994, May 1999, and December 2000.

the performance of the air traffic control system that could affect average flying times. Routes are directional to allow for prevailing winds and other physical differences in travel, so we consider Philadelphia to Los Angeles to be a different route than Los Angeles to Philadelphia. Travel time is computed as the actual arrival time minus the scheduled departure time and thus includes delays in the flight leaving the gate.¹⁷

Figure 7 plots average minimum travel time, scheduled travel time, and actual travel time. For consistency, the data used in Figure 7 includes only routes where we observe flights in each month of the entire sample period. Actual travel time exceeds minimum travel time by more than 32 minutes in the year 2000. This number has increased more than 10 percent over the sample period, although as we mentioned earlier, changes in reporting between 1994 and 1995 could account for some of that growth. In addition, minimum travel time increased from 89 to 94 minutes over the time period, possibly due to greater traffic system wide.¹⁸ Clearly carriers do not choose their schedules to have a mean delay of zero. The average delay from schedule of 9.9 minutes is both positive and large and has grown over time as airlines only increased scheduled travel time by about two-thirds of the growth in average travel time between 1988 and 2000.

We decompose the excess travel time into its component parts in Table 1. Over our

¹⁷ Such delays are often caused by late arriving aircraft. Regressions below will show that late arriving aircraft do not impact our conclusions as to the impact of hubbing and concentration on overall delays.

¹⁸ Since the average route had over 150 flights even in 1988, our lowest-volume year, we believe we measure the minimum time with good accuracy. In principle, however, we are more likely to observe the true minimum travel time on routes with more flights and could overestimate the minimum time on sparse routes.

sample period, the average flight required about 30.5 minutes more than the minimum feasible travel time on a route. Nearly 10 minutes of that excess is due to a late push-back from the gate. For flights after 1995, about one-half of the total excess travel time on the flight is spent mid-air, though much of that 16 minutes is probably due to less-than-favorable winds and weather en route. Overall, more than one in four flights is canceled or arrives at least 15 minutes late.

Following Section 2, the measure of the size of the hub and thus the extent of network benefits should be the number of possible connections for a traveler through the hub. We define this variable as the number of other airports that an airline flies to from a given airport in a particular month. Airport concentration is defined as the Herfindahl index on the share of flights by the various airlines that serve that airport over each one-month period and proxies for the extent to which delays are internalized by the carriers.

The bulk of flights in the US are associated with hubs. Table 1 shows that nearly two-thirds of all flights in the sample originate at an airport that is a hub, with the hub carrier itself originating a little more than one-half of hub flights (39 percent total flights). In all, 83 percent of flights either originate or land at some carrier's hub and almost three-quarters of all flights occur on an airline flying to or from its own hub. With the strong prevalence of hubbing, it is not surprising that the typical airport has a moderately high HHI of 0.40, although there is substantial variation across airports.

Table 2a identifies the hub carriers and reports airport concentration for all airports with at least one percent of the total flights in November 2000. Cincinnati, dominated by a Delta hub, was the most concentrated large airport that month at 0.91. Charlotte and Pittsburgh, both US Airways hubs, were close behind at 0.81. Not every airport with a hub carrier is highly

concentrated. Many single-hub airports are only moderately concentrated, such as Newark (0.38 with a large Continental hub) and Salt Lake City (0.48 with a large Delta hub). Hubs with less connection activity, such as United in San Francisco, have much lower concentrations (0.33). Some airports have multiple hub airlines, such as Chicago's O'Hare with United and American and only 0.38 concentration. Also, some busy airports do not have hubs with significant connecting activity: Chicago Midway with a concentration of 0.71 or New York's La Guardia which has a concentration of 0.22.

Overall, there has been substantial consolidation since 1988, especially in the early 1990s, when mergers and bankruptcies reduced the number of major carriers in the sample from 14 to 10.¹⁹ The remaining airlines have continued to expand their hub and spoke systems, although a few carriers abandoned previous hubs. As a result, many airports looked quite different in 1988 than they do in the year 2000. Table 2b presents the same snapshot of all airports with at least one percent of the total flights in November 1988. For example, relative to 1988, Denver and Atlanta each lost one of their hub carriers. Miami, Washington National, and Cleveland have gained a single hub carrier, Las Vegas and Los Angeles gained two hub carriers, and Phoenix added a second hub carrier. JFK, Orlando and, Raleigh Durham lost their hubs altogether. Several cities had a change of hub airline or a change in the size of the hub. Finally, airport concentration has varied over this time period, with many airports exhibiting a general increase in concentration, a few airports exhibiting a strong rise in concentration as a single carrier consolidated its hubbing at that airport, and several airports showing a decline in

¹⁹See Morrison (1996) for a discussion of the policy issues relating to the merger trend in the airline industry.

concentration as hub carriers pulled out. In many regression specifications in Section 4, we will use this variation in hub size and concentration within an airport over time to identify their effects on delays.

4. Estimation and Results

The discussion in Section 2 makes several empirical predictions regarding the impact of network benefits and congestion externalities on delays. First, flights operating at hub airports should face delays that increase with the size of the hub. Second, hub airline flights should cluster at hubbing times and non-hub carriers should avoid the delays associated with hub carrier flights. Thus most delays at hub airports will be incurred by the hub airline itself, and these delays will also be increasing in the size of the hub. Third, delays should be longer for flights that originate at a hub than flights arriving at a hub, as hub airlines cluster their departures more than their arrivals. Finally, congestion externalities should cause higher delays at less concentrated airports, holding the extent of hubbing constant.

To examine these predictions, we estimate the following base empirical specification:

$$\begin{aligned}
 \text{DELAY}_{ijkmt} = & \alpha + \beta_1 \text{CONC}_{\text{org}(k),t} + \beta_2 \text{CONC}_{\text{dest}(m),t} + \\
 & \theta_1 (\text{HUB AIRPORT}_{\text{org}})_{kt} + \theta_2 (\text{HUB AIRPORT}_{\text{dest}})_{mt} + \\
 & \gamma_1 (\text{HUB AIRLINE x HUB AIRPORT}_{\text{org}})_{jkt} + \gamma_2 (\text{HUB AIRLINE x HUB AIRPORT}_{\text{dest}})_{jmt} + \\
 & + \Psi_1 (\text{DEMAND}_{\text{org}})_{mt} + \Psi_2 (\text{DEMAND}_{\text{dest}})_{mt} + \delta_1 \text{YEAR}_t + \delta_2 \text{MONTH}_t \\
 & + \delta_3 \text{AIRLINE}_j + \delta_4 \text{AIRPORT}_{\text{org}(k)} + \delta_5 \text{AIRPORT}_{\text{dest}(m)} + \varepsilon_{ijkmt}
 \end{aligned}$$

where DELAY is a measure of travel time or on-time performance of flight i on airline j from airport k to airport m on date t . CONC refers to the airport concentration of the origin (k) or destination (m) airport. HUB is measured both at the airport level (whether airport k is a hub for any airline) and the airline level (whether airline j has a hub at airport k). An airline's hub is defined as a function of the number of airports airline j flies to from airport k . We generate dummy variables for three different ranges of the number of destination airports: 26 to 45, 46 to 70, and 71 or more.²⁰ Although our illustration in Section 2 simplified the flight scheduling decision by considering the impact of additional flights only for one endpoint, we recognize that conditions at both the arrival and departure airports determine flight decisions and delays. Thus concentration and hub are included separately for both the origin and destination airports to allow for separate effects for each end of the flight.

We also include DEMAND variables to control for changes in local demand for air travel over time and across Metropolitan Statistical Areas (MSAs) that might lead to greater flight delays. All equations include annual population, employment, and per-capita income. For airports in a MSA, we include their MSA values, but also interact the economic variables with a dummy variable that equals one if the airport is the largest airport in the MSA, a proxy for the likeliest airport to be a hub. For airports not in a MSA, we interact a non-MSA dummy or Alaska airport dummy with national values of the economic variables.²¹ Most of these economic

²⁰Our results are robust to alternative functional form assumptions, but we find that the categories provide a better fit than a linear function and are more easily interpretable than a higher-order polynomial.

²¹Almost all airports in our data set that are not in an MSA and not in Alaska are located at destination vacation spots. Many are airports at ski resort locations.

variables are of the expected sign and are statistically significant.

All specifications have dummies for the year and month of travel to control for unobserved time and seasonal factors that may affect system wide delays, and for the airline, j , to control for unobserved airline quality. Finally, most specifications are run with a full set of fixed effects for the airport the flight originates from (k) and the airport it arrives at (m) to control for unobserved airport heterogeneity that may affect delays, such as capacity.

Given that we have data on more than 66 million flights, we take two steps to make estimation more manageable. First, we narrow our data to all flights on Fridays.²² We have done some preliminary estimation on Saturdays, the least busy day of the week, and obtain the same basic results. Second, in our basic specification we generate cells of flights by each airline on every route for all months in every year, a total of more than 617,150 airline-route-month/year cells. Within each cell we compute the mean of the dependent variable and all independent variables, and use these cell means in the regressions that are reported in this paper. These regressions are weighted by the number of flights within the cell. Since none of the independent variables in our basic specification vary within the cells, the coefficient estimates are identical to what we would obtain using OLS estimates on the entire sample of Friday flights. We compute robust standard errors, allowing the residuals to be correlated over time within a route.

We examine three basic measures of delays. Our preferred measure of delay is the excess travel time above the minimum feasible time since we think a longer travel time is the

²²We construct the independent variables in our regressions using all data, not just Fridays.

best indicator of the social cost of hubbing and congestion. As an alternative, we use schedule delays based on the number of minutes between actual arrival time and scheduled arrival time. Schedule delay reflects the delay relative to expected arrival time and can be negative—if the flight arrives early-- as well as positive. This measures delays to the extent the airlines do not completely adjust their schedules. Our third estimate of delay is the Department of Transportation’s widely reported “on time” arrival, which is flights arriving within 15 minutes of schedule. In this measure, canceled or diverted flights are counted as late.

Airport level findings:

Overall, we find large and significant effects of hubbing and moderate effects of concentration on delays. Our initial evidence is presented in Table 3. The dependent variable is excess travel time above the minimum feasible travel time. Consistent with our characterization of network benefits from hubbing, hub airports have more delays. In column 1, flights originating and arriving at hub airports face delays of up to 7.2 and 4.5 minutes, respectively. In addition, hub delays increase monotonically in the size of the hub. Flights that originate from the smallest hubs are delayed four minutes more than flights departing from non-hub airports, 6.7 minutes at medium size hubs, and 7.2 at the largest ones. A similar pattern holds for flights flying to hub airports.

We also find evidence that airports with low concentration have higher delays, possibly because carriers do not fully internalize the costs their flights impose on other carriers. In column (1), higher concentration has a small but beneficial impact on delays. Controlling for the extent of hubbing, a one standard deviation increase in concentration (0.20) leads to a modest 1.2

minute decline in delay at both origin and destination airports. Even an increase from the mean concentration level of 0.40 to an airport with just one airline leads to just a 3.6 minute decrease in delays, smaller than the effect of hubbing.

One potential problem with this regression is the possibility that our income, employment, and population variables might not fully control for local demand. In particular, airports with high unobserved local demand for air travel might have a greater number of flights and also have a hub that serves a large number of destinations. Thus high levels of congestion may be due to local demand rather than hubbing. To address this issue, we take two approaches. In column (2), we instrument for the probability that an airport is a hub with variables that are based on the demand for connections by surrounding communities, rather than by the hub city. We compute the distance from a given airport to all of the other airports in our sample, counting the number of airports within 500 miles, 500-1000 miles, and 1000-1500 miles, and also sum up the population and per-capita income for each of the airports within those rings. This gives the total demand for connections around each airport, both in terms of number of connecting airports and economic buying power of the potential connections. The demand variables are also interacted with a dummy variable that indicates the primary airport within each MSA. Such an interaction is important to differentiate the largest airport from smaller secondary airports within an MSA. These instruments are significant in the first stage and are moderately successful in isolating the hub delay effect from local demand. For origin airports, the hub variables are still individually and jointly significant and nearly as large as the OLS coefficients, suggesting that hubs are associated with greater congestion. However, the destination hub variables are much smaller than the OLS coefficients and are not statistically significant. While the instruments can

successfully identify hubs, separate origin and destination effects may be harder to pin down. Overall, the estimation is consistent with greater origination delays at hub airports.

In the third column we include airport fixed effects. By effectively looking only at changes over time within airports in hubbing, concentration, and delays, we absorb time-invariant airport level factors such as capacity or local demand. However, by including fixed effects, we eliminate a large source of variation—differences in hubbing and delays across airports. The fixed effects estimates present the same pattern as the earlier estimation, except the coefficients are considerably smaller. The biggest hubs have the largest delays and delays increase with the size of the hub. The coefficient on the smallest category of origin and destination hubs is negative, suggesting that these hubs appear to have slightly better performance than non hub airports. However, the negative coefficients themselves are small. Concentration also reduces delays, although the coefficients are again much smaller than in the estimation that does not include fixed effects. To the extent that cross sectional variation is required to identify airport level hubbing and concentration effects, the fixed effects estimates might provide a lower bound on the true effects.

Within-airport clustering of flights:

In column (4), we move on to consider an important implication of our model of network benefits: that the hub carrier should have greater delays than non-hub carriers at hub airports. In this case we include separate covariates for flights by the hub carrier to or from its own hub airport. With airport fixed effects, the hub/non-hub carrier effect is identified based on differences between hub and non-hub airlines within each airport, so all carriers at the airport are

subject to identical capacity constraints and face the same level of local demand. The results suggest that the dominant hub carrier incurs most delays at hub airports. Relative to non-hub airlines at the same airport, hub airlines have excess travel time of up to 5.5 minutes at origin airports and 2.8 minutes at destination airports. Estimated delays accruing to the hub airline increase monotonically in the size of the hub and are larger for origin airports. All of these results are consistent with the existence of strong network benefits that lead to more delays for hub airlines.

In this regression, airport-level hub variables indicate the extent of delays by non-hub airlines at a hub airport. In column (4), the coefficients on the airport level variables become much smaller in magnitude, and are even negative in some cases. The negative coefficient indicates that the non-hub airlines have slightly lower delays when they operate at smaller hubs rather than non-hub airports. These coefficients are consistent with the hypothesis that non-hub airlines benefit from scheduling their flights at times when there are few hub airline flights. Even with these changes, the results suggest that the performance of non-hub carriers deteriorates monotonically with hub size and that non-hub carriers at the largest hubs face worse delays than carriers that operate at airports without a hub.

Airport concentration remains negative and statistically significant, but its coefficient in the fixed effects specifications is small compared to the magnitude of the hub variables. A large increase in the HHI from 0.40 to 1 only leads to a 1.1 to 1.6 minute decrease in delays, less than one-third of the increase in delays associated with the largest hub airlines.

The evidence that non-hub airlines' flights at hub airports have fewer delays than hub airlines' flights might at first seem surprising, but it emphasizes the importance that hubbing

plays in the operations of most airlines. The vast majority of airline service involves flights to or from a carrier's hub. Given that the economic returns from adding a flight on a given route are much higher for a hub carrier, either because of network benefits or increased market power by the hub carrier, non-hub carriers are often reluctant to add flights to or from another carrier's hub that do not travel to its own hub. Thus at the non-hubbing times, flight levels fall off substantially for most hub airports.

While hub carriers need to cluster their departures to maximize network connection benefits, Section 2 shows that they can smooth their arrivals somewhat as long as the cost of long connections is not too high. Our results support this conclusion since the effect of hubbing on originating flights is much larger than on arriving flights in every specification. If delays come from peak loads of traffic at hubs, arrivals should exhibit lower hub-induced delays than the more clustered departures, as we observe. The delays due to concentration do not depend on clustering of flights and thus should not exhibit a systematic pattern of being larger for origin or destination airports. Indeed, the estimated concentration effects are similar for both types of airports.

Further exploration:

The results in Table 3 suggest that hub carriers are willing to accept substantial delays on their own flights, even at airports where they control a large proportion of the total flights. We investigate the reasons behind these delays and the robustness of these results using more detailed data beginning in 1995 that allows us to track the movement of individual aircraft and to decompose overall travel time into time spent taxiing out to the runway, in the air, and taxiing in

to the gate at the destination.

To begin, we consider the possibility that increased hub airline delays at the origin airport relative to the destination are due to late arriving aircraft from previous flights, so called “cascading” delays. For flights after 1995, we use data on the aircraft tail numbers to compute a variable called buffer, defined as the difference between the actual arrival time of the airplane from its previous flight and its next scheduled departure time. Buffer measures the time the plane has to be “turned around” once at the gate. By scheduling more time on the ground, an airline can maintain a sufficient buffer even if the aircraft arrives late, mitigating cascading delays from late-arriving flights.

In Table 4 we present the same base regression with fixed effects from column (4) of Table 3 with the addition of spline terms for various levels of buffers: greater than 120 minutes, 41 to 120 minutes, 20 to 40 minutes, 1 to 20 minutes, 0 to -120 minutes, and less than -120 minutes. The last two categories reflect aircraft that arrived after their next scheduled departure. The first column includes the base regression with the spline terms, and the second column is the base regression run only on the 1995 to 2000 sample period for comparison.

Our primary focus in these regressions is the extent to which the within airport hub airline variables change when we control for the buffer from the inbound aircraft. Once we control for buffer, the results in column (1) suggest that late arriving aircraft do not explain greater delays for hub airlines at either origin or destination airports. Controlling for the buffer, the hub airline coefficients nearly double in size and significance. This result is not surprising given that hub carriers have a greater incentive to have large buffers for inbound aircraft. To the extent that hub airlines face higher costs from late arrivals due to potential missed connections,

we should expect that hubs allow more time connecting. Comparing columns (1) and (2), when we condition on the length of the plane's layover, hub airlines at hub airports create more new delays than non-hub airlines but partially offset that effect by being better able to mitigate the effect of previous delays.

The coefficients on the buffer spline terms are all of expected sign and significance. The coefficient on each spline term is interpreted as the impact of an aircraft arriving within a given buffer time period on the departure time of the aircraft on its next flight. The results show that buffer has little additional impact on flights already arriving more than 40 minutes before their next departure. Airlines can make up about one-third of the delay ($1 - 0.622$) when aircraft arrive between 21 and 40 minutes prior to the next scheduled departure by turning the aircraft around quickly, but travel time increases about one-for-one (1.012) for flights whose inbound aircraft had a buffer of less than 20 minutes. The fact that delays do not increase one-for-one for the latest arriving category, more than 120 minutes late, may be due to airlines deciding to either cancel or substitute another aircraft for some very late flights.

The airport concentration effects are much lower in both specifications that use the 1995 to 2000 sample period. However, with fixed effects, the coefficient on concentration is identified only by changes in concentration within airports over a six year period. While there were many significant changes in concentration and hubbing within airports between 1988 and 2000, most of those changes had already taken place by 1995, leaving little variation with which to identify the concentration coefficient.

To further examine the robustness of our results, we decompose excess travel time into its various components, including delay in departure from the gate, time spent taxiing to the

runway, travel time mid air in excess the monthly minimum on that route, and taxi time to the gate at the destination airport. To the extent that the delays on hub carriers are due to hubbing, we should be able to isolate these effects at the origin and destination airports.²³

Table 5 contains regression results using these four dependent variables in our base specification. Given that the short time period with the airport dummies makes it difficult to identify airport level effects, we will focus on the within-airport hub airline coefficients. The results are consistent with the network benefits view of delays. For hub carrier flights originating at their hub, most delays involve a late departure from the gate or increased time on the taxiway, with the same increasing delays with hub size that we saw in Table 3. In fact, the sum of hub airline coefficients in column (1) for departure delay and column (2) for taxi out time is nearly exactly equal to our total estimated delay for flights originating at a hub in Table 4, column (2). Originating from a large hub accounts for 30 percent of the average departure delay and 16 percent of the average delay on the tarmac waiting to take off. Originating at a hub has virtually no effect on excess flying time or time spent on the taxiway at the destination airport.

Airline flights to their own hub require about a minute more time mid air, possibly due to air traffic queues into the hub airport, and two minutes of additional inbound taxi time, which is one-third of the average, with the delays increasing with the size of the hub. Once again, flying to a hub does not affect delays at the origination airport as the departure delays and taxi out time effects are almost always indistinguishable from zero.

²³Several seminar participants have suggested that variation in aircraft speed may explain differences in travel time relative to minimum travel time on various routes. Below, we show that most hubbing related delays occur on the ground at either the origin or destination airport, casting doubt on the likelihood that differences in aircraft types by hub and non-hub carriers can explain these results.

These regressions also provide some insights about the exact source of hub delays. The fact that inbound hub aircraft spend additional time on the taxiway after landing suggests that peak airport capacity, not just runway capacity, may be a problem at some hub airports. Such delays could be due to congestion on the taxiway or time spent waiting for a gate to become available. Excess departure delays and taxi out time at hubs might also reflect such congestion as the “alleyways” between gates may be blocked by other departing aircraft. Given that these excess delays are imposed only on the hub carrier’s flights, the hub carrier has a strong incentive to invest in additional airport capacity.²⁴

Our final set of specifications in Table 6 examine alternative measures of delay. The first column uses on-time performance as defined by the Department of Transportation. These regressions use our whole sample period from 1988 to 2000. Interestingly, we continue to find that hub airlines face much greater delays than other carriers even when we measure delays relative to schedule rather than against minimum feasible travel time. Flights originating at the largest hubs have an on-time performance that is 6.7 percentage points worse than flights on non-hub carriers. Arriving hub flights have a worse on-time performance as well, but to a lesser extent. Non-hub flights perform slightly better at the smallest hub airports, while concentration only impacts on-time performance at destination airports. Overall, these results suggest that airlines appear to take into account a portion of expected delays in setting schedules, but that hub carriers do not fully account for the effect of hubbing on the likelihood that their flights are late. We see the same basic results in column (2) when we use the average delay from schedule as our

²⁴While airlines cannot easily increase runway capacity due to political constraints, many hub carriers have invested heavily in recent years to increase the number of gates and provide additional taxiways.

dependent variable. Future work will investigate the factors that lead carriers to compensate for expected delays in choosing scheduled arrival times.²⁵

Finally, column (3) investigates cancellations. This is one area where hub carriers perform better than non-hub airlines. Hub airlines have fewer cancellations on flights that originate at their own hubs, while non-hub carriers have more cancellations at hub airports. The improved performance of hub carriers may reflect two factors. First, hub airlines may have extra aircraft available at their own hubs that can be substituted for a disabled plane in the event of a severe mechanical malfunction. Second, hub airlines face stronger network-based disincentives to cancel flights because canceling a flight inbound to a hub will severely inconvenience the passengers who need to make their connections and leave the airline without an aircraft at the hub to service an outbound route. Canceling a flight outbound from a hub may strand away from home many passengers waiting to connect to their final destination. Thus hub airline cancellations have a negative impact on more passengers than non-hub cancellations do. That hub flights arriving at large hubs have nearly the same low cancellation rate as hub flights departing from a hub is evidence in favor of the network interpretation of these results.

5. Conclusion

Over the last 13 years, air traffic delays have grown considerably. From a policy perspective, it is important to distinguish between the two potential causes of delays: network benefits from hubbing, which lead hub carriers to accept higher equilibrium levels of delays, and

²⁵For example, accurate scheduling may be related to the extent of competition on a route, the type of travelers on a route—business versus leisure—or the ease of obtaining public information about delays on specific flights.

congestion externalities, which lead to higher delays for all carriers at an airport. Although we find evidence that congestion externalities lead to modest levels of air traffic delays, our results suggest that hubbing is the primary economic contributor to air traffic congestion. Flights departing from hub airports require between 4 and 7 minutes of excess travel time, while flights arriving at a hub require 1.5 to 4.5 minutes of additional delay. However, nearly all of the delays associated with hubbing are incurred by the hub airline itself. Non-hub airlines operating at hub airports face minimal delays at all but the largest hubs. Within hubs, delays increase monotonically with the size of the hub and flights originating at the hub face greater delays than flights arriving at a hub. Finally, hub carriers are less likely to cancel flights to or from their hubs. All of these findings are consistent with a model in which the hub carrier receives large network benefits that increase with the number of markets served from a hub. These interconnection benefits encourage the hub carrier to bunch its flights at hubbing times, even at the cost of additional delays. Non-hub airlines operating at the hub airport optimally schedule their flights to avoid these peak hubbing times.

From a social perspective, these results suggest that there is little area for public policy intervention and even the appropriate Pigouvian tax might have a small impact on delays at most airports. Delays are primarily the outcome of hub airlines trying to provide a large variety of potential destinations with short connection times. With a very high market share during peak times when most delays occur, dominant hub carriers already appear to internalize most congestion costs at their own hub airports. While hub carriers may impose a cost on non-hub airlines by crowding them out of certain flight times, a social planner who recognized the network benefits from hubbing would choose a similar outcome.

To some extent, our finding of a small congestion externality effect reflects the fact that the four airports most likely to suffer from it -- La Guardia, JFK, Washington National and Chicago O'Hare -- already face FAA limits on the number of hourly flights at the airports. In a well-publicized policy shift, the FAA recently removed the hourly caps on service at La Guardia, an airport that has a low concentration, no single dominant hub carrier, small capacity, and large local demand. Airlines immediately moved in to increase service. By November of 2000, La Guardia accounted for about 25 percent of all delays of more than 15 minutes for the entire nation. This incident suggests that congestion externalities are important at some airports, and would be more important were the FAA to remove the caps at any of the slot constrained airports. Congestion pricing may be an appropriate solution for the inefficiency at these airports.

Nonetheless, our findings also suggest that the imposition of arbitrary caps on airport takeoffs and landings that do not account for the network benefits of hubbing might result in social losses. In the presence of hubbing, the optimal policy does not just minimize delays without considering interconnection benefits. Delays are not necessarily evidence of a socially inefficient outcome, but might instead reflect the optimal use of scarce runway capacity.²⁶

Finally, this paper raises a number of issues that may form the basis of future research, such as the finding that airlines do not choose their schedules to set average delay equal to zero or to correct for the impact of hubbing on schedule time. That passengers purchase tickets based on scheduled travel time rather than actual expected travel time might contribute to an airline's

²⁶In addition, these results do not measure the extent to which hub carriers might take steps to limit entry by other carriers. Debates over access to gates at the new airports in Denver and Pittsburgh suggest that market power over local passengers is an important consideration by airlines regarding their willingness to invest in additional capacity.

unwillingness to fully account for expected delays in their schedules. We can also investigate how the extent of competition, the publication of overall airline on-time percentage rankings, or the provision of actual flight information over the internet has impacted the incentives for airlines to publish accurate schedules.

References:

- Arnott, Richard. 1979. "Unpriced Transport Congestion," *Journal of Economic Theory*, 21, 294-316.
- Bailey, Elizabeth and John Panzar. 1985. *Deregulating the Airlines*, Cambridge: MIT Press.
- Borenstein, Severin. 1989. "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry," *Rand Journal of Economics*, 20(Autumn), 344-65.
- _____. 1990. "Airline Mergers, Airport Dominance, and Market Power," *American Economic Review Papers and Proceedings*, 80(May), 400-4.
- _____. 1991. "The Dominant-Firm Advantage in Multi-Product Industries: Evidence from the U.S. Airlines," *Quarterly Journal of Economics*, 106(November), 1237-66.
- _____. 1992. "The Evolution of U.S. Airline Competition," *Journal of Economic Perspectives*, 7 (Spring), 45-73.
- Borenstein, Severin and Nancy Rose. 1994. "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 103(August), 653-83.
- Borenstein, Severin and Janet Netz. 1999. "Why do all flights leave at 8 am?: Competition and departure-time differentiation in airline markets," *International Journal of Industrial Organization*, 17, 611-40.
- Brueckner, Jan K. 2001. "Airport Congestion when Carriers Have Market Power," University of Illinois mimeo, May.
- Brueckner, Jan K., Nichola J. Dyer, and Pablo T. Spiller. 1992. "Fare Determination in Airline Hub-and-Spoke Networks," *Rand Journal of Economics*, 23(3), 309-34.
- Brueckner, Jan K. and Pablo T. Spiller. 1994. "Economies of Traffic Density in the Deregulated Airline Industry," *Journal of Law and Economics*, 37, 379-415.
- Brueckner, Jan and Yimin Zhang. 2000. "A Model of Scheduling in Airline Networks: How a Hub and Spoke System Affects Flight Frequency, Fares, and Welfare," forthcoming in *Journal of Transportation Economics and Policy*.
- Carlin, Alan and R. E. Park. 1970. "Marginal Cost Pricing of Airport Runway Capacity," *American Economic Review*, 60(3), 310-19.

- Caves, Douglas, Lauritis Christensen, and Michael Tretheway. 1984. "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *Rand Journal of Economics*, 15(Winter), 471-89.
- Daniel, Joseph I. 1995. "Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues," *Econometrica*, 63(2), 327-70.
- Daniel, Joseph I. And Munish Pahwa. 2000. "Comparison of Three Empirical Models of Airport Pricing," *Journal of Urban Economics*, 47, 1-38.
- Economides, Nicholas. 1996. "The Economics of Networks," *International Journal of Industrial Organization*, 14(6), 673-99.
- Encaoua, David, Michel Moreaux, and Anne Perrot. 1996. "Compatibility and Competition in Airlines: Demand Side Network Effects," *International Journal of Industrial Organization*, 14(6), 701-26.
- Hendricks, Ken, Michelle Piccione, and Guofu Tan. 1995. "The Economics of Hubs: The Case of Monopoly," *Review of Economic Studies*, 62(1), 83-99.
- _____. 1997. "Entry and Exit in Hub-Spoke Networks," *Rand Journal of Economics*, 28(2), 291-303.
- Hergott, Matthew. 1997. "Airport Concentration and Market Power: An Events Study Approach." *Review of Industrial Organization*, 12(December), 793-800.
- Kahn, Alfred . 1993. "The Competitive Consequences of Hub Dominance: A Case Study," *Review of Industrial Organization*, 8(August), 381-405.
- Keeler, J P and John Formby. 1994. "Cost Economies and Consolidation in the U.S. Airline Industry," *International Journal of Transport Economics*, 21(February). 21-45.
- Kim, E Han and Vijay Singal. 1993. "Mergers and Market Power: Evidence from the Airline Industry," *American Economic Review*, 83(June), 549-69.
- McShan, Scott and Robert Windle. 1989. "The Implications of Hub-and-Spoke Routing for Airline Costs and Competitiveness," *Logistics & Transportation Review*, 25(September), 209-30.
- Morrison, Steven. 1996. "Airline Mergers: A Longer View," *Journal of Transport Economics & Policy*. 30(September), 237-50.
- Morrison, Steven and Clifford Winston. 1989. "Enhancing the Performance of the Deregulated Air Transportation System," *Brookings Papers on Economic Activity*, Microeconomics, 61-123.

Saloner, Garth and Andrea Shepard. 1995. "Adoption of Technologies With Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines," *The Rand Journal of Economics*, 26(3), 479-501.

Singal, Vijay. 1996. "Airline Mergers and Competition: An Integration of Stock and Product Price Effects," *Journal of Business*, 69(April), 233-68.

Vickrey, William S. 1969. "Congestion Theory and Transport Investment," *American Economic Review*, 59(2), 251-60.

Zhang, Anming. 1996. "An Analysis of Fortress Hubs in Airline Networks," *Journal of Transport Economics & Policy*, 30(September), 293-307.

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	
Actual elapsed route time minus minimum + departure delay	30.6	15.0	
Departure delay: actual push-back time minus scheduled departure	9.9	10.2	
Taxi out time: time from push-back to wheels-off-the-ground*	15.2	5.7	
Actual flight time minus minimum feasible flight time*	16.2	9.4	
Taxi in time: time from landing to reaching the gate*	6.0	2.5	
“On time:” arriving within 15 minutes of schedule	0.75	0.17	
Difference between actual arrival time and scheduled arrival	9.9	12.2	
Flight cancelled?	0.018	0.05	
Origin airport concentration	0.40	0.21	
Origin airport hub size:			
26 – 45 markets	0.19	0.39	
46 – 70 markets	0.24	0.43	
71 + markets	0.22	0.42	
Origin airline hub size:			
26 – 45 markets	0.09	0.28	
46 – 70 markets	0.15	0.36	
71 + markets	0.15	0.36	
Flight is traveling to or from a hub airport	0.83	0.38	
Flight is traveling to or from the airline’s own hub	0.74	0.44	
Buffer: Minutes between actual arrival and scheduled departure, spline terms:			
	Percent of <u>Total</u>	<u>Mean</u>	<u>Standard Deviation</u>
< -120 minutes*	0.004	-173	63
0 and –120 minutes*	0.061	-29	27
1 and 20 minutes*	0.093	13	6
21 and 40 minutes*	0.191	31	6
41 and 120 minutes*	0.403	62	17
> 120 minutes*	0.249	652	343

Sample includes all flights for major carriers on Fridays from January, 1988 – November, 2000 (N=9,956,576), except for rows with (*) which include Fridays from January, 1995- November, 2000 (N= 4,592,595). Also, data is missing for July and August, 1993, March, 1994, and May, 1999.

Table 2a: Hubbing and concentration for airports with at least one percent of flights in November, 2000

Airport	Airport Concentration	Hub carriers with...		
		71+ connections	46-70 connections	26-45 connections
Atlanta	0.72	Delta		
Baltimore-Washington	0.29			Southwest
Charlotte	0.81		USAirways	
Chicago O'Hare	0.38	United	American	
Cincinnati	0.91		Delta	
Cleveland	0.31			Continental
Dallas-Forth Worth	0.52	American		Delta
Denver	0.57		United	
Detroit	0.63	Northwest		
Houston Intercontinental	0.61	Continental		
Las Vegas	0.25			America West, Southwest
Los Angeles	0.19			American, United
Memphis	0.66			Northwest
Miami	0.37			American
Minneapolis-St. Paul	0.66	Northwest		
Nashville	0.31			Southwest
Newark	0.38		Continental	
Philadelphia	0.50		USAirways	
Phoenix	0.30		America West	Southwest
Pittsburgh	0.81		USAirways	
Saint Louis	0.52	Trans World		
Salt Lake City	0.48		Delta	
San Francisco	0.33			United
Washington Dulles	0.39			United
Washington National	0.23			USAirways

Airports with at least one percent of flights in November 2000 but without a large hub carrier are (concentration in parentheses): Boston (0.29), Chicago Midway (0.71), Houston Hobby (0.84), LaGuardia (0.22), Kansas City (0.23), Oakland (0.47), Orlando (0.19), Portland (0.19), San Diego (0.21), San Jose (0.24), Seattle (0.23), Tampa (0.18)

Table 2b: Hubbing and concentration for airports with at least one percent of flights in November, 1988

Airport	Airport Concentration	Hub carriers with...		
		71+ connections	46-70 connections	26-45 connections
Atlanta	0.43	Delta	Eastern	
Baltimore-Washington	0.47			Piedmont
Charlotte	0.80		Piedmont	
Chicago O'Hare	0.40	United	American	
Cincinnati	0.54		Delta	
Dallas-Forth Worth	0.44	American	Delta	
Denver	0.38		Continental United	
Detroit	0.45		Northwest	
Houston Intercontinental	0.51		Continental	
Memphis	0.63		Northwest	
Minneapolis-St. Paul	0.60		Northwest	
Nashville	0.40			American
Newark	0.27			Continental
New York JFK	0.18			TWA
Orlando	0.17			Delta
Philadelphia	0.23			USAirways
Phoenix	0.28			America West
Pittsburgh	0.69	USAirways		
Raleigh-Durham	0.41			American
Saint Louis	0.61	Trans World		
Salt Lake City	0.62		Delta	
San Francisco	0.20			United
Washington Dulles	0.47			United

Airports with at least one percent of flights in November 1998 but without a large hub carrier are (concentration in parentheses): Boston (0.12), Cleveland (0.21), Houston (0.41), LaGuardia (0.12), Las Vegas (0.24), Los Angeles (0.13), Miami (0.19), San Diego (0.13), Seattle (0.16), Tampa (0.14), and Washington National (0.12).

Table 3: The Effect of Airline Hubbing and Airport Concentration on Travel Time

		Dependent Variable: Travel Time in Excess of Minimum Feasible							
		(1)		(2)		(3)		(4)	
		OLS		IV		OLS		OLS	
		Origin	Destination	Origin	Destination	Origin	Destination	Origin	Destination
<u>Airline Hub Size</u>									
26 to 45 markets								3.37 (0.34)	0.28 (0.35)
46 to 70 markets								5.24 (0.39)	1.96 (0.41)
71 or more markets								5.57 (0.45)	2.81 (0.47)
<u>Airport Hub Size</u>									
26 to 45 markets		4.02 (0.27)	2.30 (0.28)	4.46 (0.19)	-0.31 (0.18)	-0.60 (0.23)	-1.05 (0.20)	-1.91 (0.27)	-0.91 (0.26)
46 to 70 markets		6.70 (0.33)	3.65 (0.34)	6.24 (0.23)	-0.76 (0.22)	1.90 (0.33)	0.81 (0.31)	-0.82 (0.39)	0.06 (0.40)
71 or more markets		7.24 (0.42)	4.50 (0.42)	5.45 (0.15)	0.88 (0.15)	4.03 (0.37)	1.84 (0.35)	0.98 (0.47)	0.42 (0.47)
<u>Airport Concentration</u>		-6.34 (0.57)	-6.94 (0.60)	-4.80 (0.24)	-1.45 (0.24)	-1.31 (0.56)	-2.94 (0.54)	-1.75 (0.55)	-2.66 (0.54)
Airport Fixed Effects	No								
R-squared	No	0.24		No		Yes		Yes	
		0.24				0.34		0.34	

Notes: Robust standard errors in parentheses. Regressions are based on the mean of the dependent independent variables for each airline on every route for all months in every year. (N= 617, 150 airline-route-month/year cells) Equations also include indicator variables for year, month, and airline and various economic demand variables that are described in the paper.

Table 4: The Effect of Late Arriving Aircraft on Excess Travel Time Due to Airline Hubbing and Airport Concentration: 1995-2000

Dependent Variable: Travel Time in Excess of Minimum Feasible

	(1)		(2)	
	Origin	Destination	Origin	Destination
<u>Airline hub size</u>				
26 to 45	3.39 (0.57)	-0.60 (0.57)	2.44 (0.43)	-0.85 (0.43)
46 to 70	7.97 (0.60)	2.96 (0.64)	4.80 (0.56)	1.39 (0.60)
71 or more	9.46 (0.62)	4.98 (0.64)	4.95 (0.57)	2.25 (0.58)
<u>Airport hub size</u>				
26 to 45 markets	-2.63 (0.33)	-1.45 (0.35)	-3.56 (0.35)	-2.04 (0.36)
46 to 70 markets	-3.30 (0.55)	-0.83 (0.56)	-2.86 (0.63)	-1.13 (0.61)
71 or more markets	-3.10 (0.66)	-1.71 (0.67)	-1.79 (0.73)	-0.78 (0.71)
Airport concentration	-0.21 (1.15)	-2.02 (1.00)	-0.48 (1.401)	-1.17 (1.41)
<u>Buffer size</u>				
Less than – 121 minutes		-0.551 (0.072)		
0 to – 121 minutes		-1.080 (0.012)		
0 to 20 minutes		-1.012 (0.013)		
21 to 40 minutes		-0.622 (0.013)		
41 to 120 minutes		-0.086 (0.003)		
More than 120 minutes		0.025 (0.000)		
R-squared	0.40		0.07	

Notes: Robust standard errors in parentheses. Regressions also include indicator variables for year, month, airport, and airline and various economic demand variables that are described in the paper. Based on a 1-in-4 sample of all flights on Fridays. (N= 1,112,510)

Table 5: Decomposing Delays to Identify the Effects of Hubbing on Various Measures of Travel Time: 1995-2000

Dependent Variable:	(1)		(2)		(3)		(4)	
	Departure Delay		Taxi Out Time		Flight Time in Excess of Minimum Feasible		Taxi In Time	
	Origin	Destination	Origin	Destination	Origin	Destination	Origin	Destination
<u>Airline hub size</u>								
26 to 45 markets	1.71 (0.27)	-0.35 (0.28)	1.51 (0.15)	-0.53 (0.13)	-0.78 (0.42)	-0.39 (0.41)	0.11 (0.08)	0.41 (0.08)
46 to 70 markets	3.13 (0.33)	-0.19 (0.33)	2.28 (0.16)	-0.27 (0.17)	0.18 (0.48)	1.24 (0.47)	-0.22 (0.08)	1.50 (0.09)
71 or more markets	3.08 (0.35)	0.17 (0.33)	2.43 (0.18)	-0.01 (0.17)	0.12 (0.47)	1.20 (0.49)	0.20 (0.10)	2.17 (0.11)
<u>Airport hub size</u>								
26 to 45 markets	-2.40 (0.25)	-1.34 (0.28)	-1.61 (0.13)	-0.13 (0.13)	-0.29 (0.26)	-0.35 (0.31)	-0.09 (0.06)	-0.47 (0.05)
46 to 70 markets	-1.91 (0.38)	-0.75 (0.40)	-0.82 (0.22)	-0.42 (0.22)	-1.40 (0.41)	-0.04 (0.40)	-0.10 (0.08)	-0.83 (0.08)
71 or more markets	-0.95 (0.44)	-0.18 (0.47)	-0.80 (0.25)	-0.85 (0.24)	-1.30 (0.49)	-0.07 (0.49)	-0.47 (0.11)	-1.22 (0.10)
<u>Airport concentration</u>	-0.19 (0.95)	-1.07 (1.02)	-0.36 (0.38)	-0.63 (0.45)	-2.55 (0.969)	-2.86 (1.012)	0.19 (0.17)	0.68 (0.15)
R-squared	0.16		0.69		0.30		0.71	

Notes: Robust standard errors in parentheses. Regressions are based on the mean of the dependent independent variables for each airline on every route for all months in every year. (N= 262,285 airline-route-month/year cells) Equations also include indicator variables for year, month, and airline and various economic demand variables that are described in the paper.

Table 6: The Effect of Airline Hubbing and Airport Concentration on Alternative Measures of Air Travel Delays

Dependent Variable:	(1)		(2)		(3)	
	On Time		Difference Between Actual and Scheduled Arrival		Flight Cancelled	
	Origin	Destination	Origin	Destination	Origin	Destination
<u>Airline hub size</u>						
26 to 45 markets	-0.039 (-0.003)	-0.006 (-0.003)	2.05 (-0.21)	0.22 (-0.23)	-0.001 (-0.001)	-0.001 (-0.001)
46 to 70 markets	-0.052 (-0.003)	-0.012 (-0.003)	2.64 (-0.23)	0.46 (-0.22)	-0.004 (-0.001)	-0.004 (-0.001)
71 or more markets	-0.067 (-0.003)	-0.015 (-0.003)	2.69 (-0.25)	0.17 (-0.25)	-0.005 (-0.001)	-0.003 (-0.001)
<u>Airport hub size</u>						
26 to 45 markets	0.017 (-0.003)	0.009 (-0.003)	-1.32 (-0.21)	-0.97 (-0.22)	0.002 (-0.001)	0.002 (-0.001)
46 to 70 markets	0.004 (-0.005)	-0.004 (-0.005)	-0.33 (-0.31)	0.11 (-0.32)	0.008 (-0.002)	0.007 (-0.002)
71 or more markets	-0.002 (-0.005)	-0.01 (-0.005)	0.50 (-0.36)	0.40 (-0.37)	0.01 (-0.002)	0.009 (-0.002)
<u>Airport concentration</u>						
	0.001 (-0.007)	0.034 (-0.007)	-0.26 (-0.44)	-1.80 (-0.46)	-0.007 (-0.002)	-0.007 (-0.002)
R-squared	0.17		0.13		0.12	

Notes: Robust standard errors in parentheses. Regressions are based on the mean of the dependent independent variables for each airline on every route for all months in every year. (N= 617, 150 airline-route-month/year cells) Equations also include indicator variables for year, month, and airline and various economic demand variables that are described in the paper.

Figure 1: Hub versus airport's total flights at Dallas-Fort Worth (DFW)

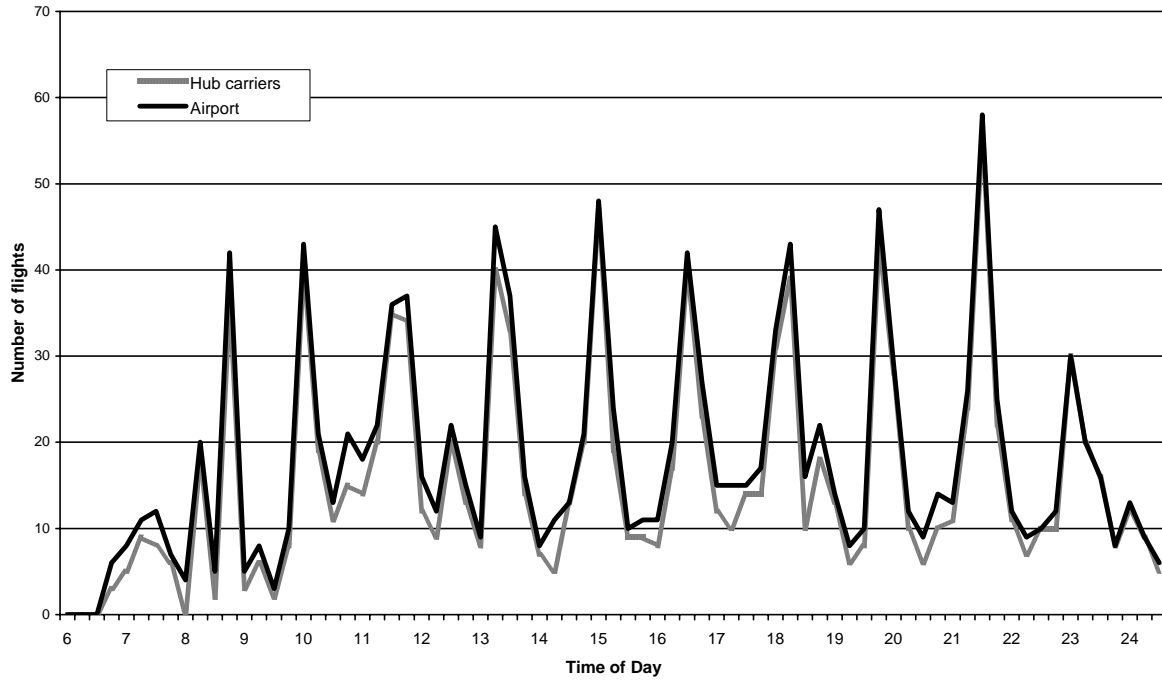


Figure 2: Hub carriers' departures and arrivals at Dallas-Fort Worth (DFW)

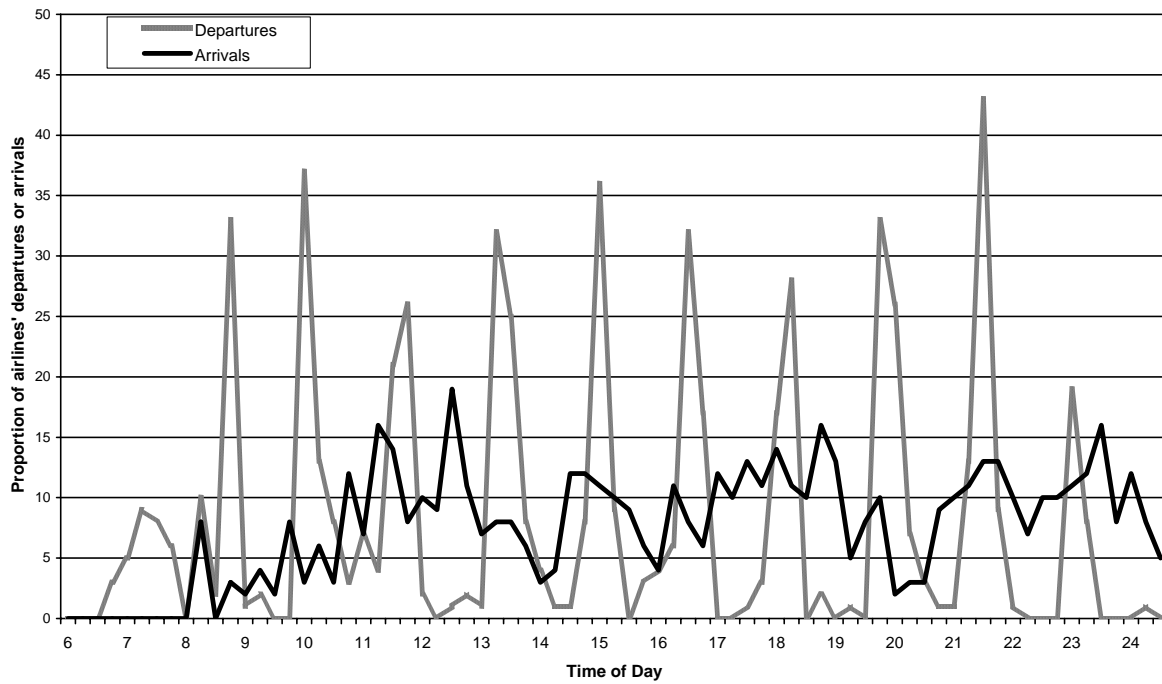


Figure 3: Total flights at Boston Logan Airport (BOS)

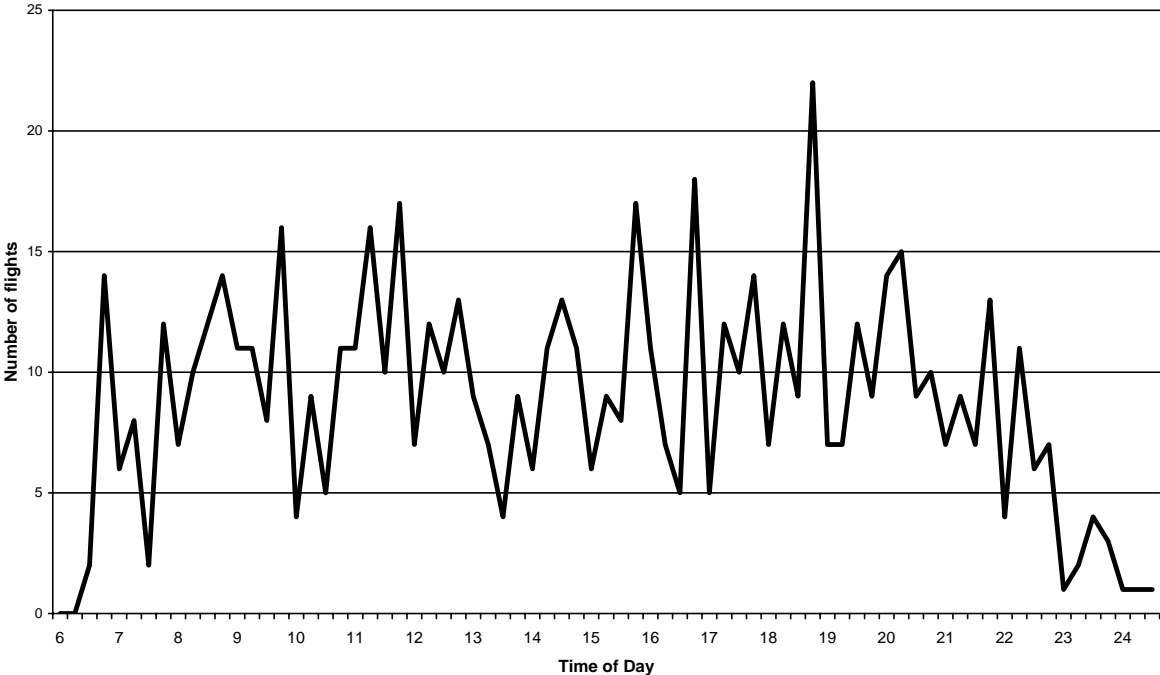


Figure 4: Departure density for hub and non-hub carriers at Dallas-Fort Worth (DFW)

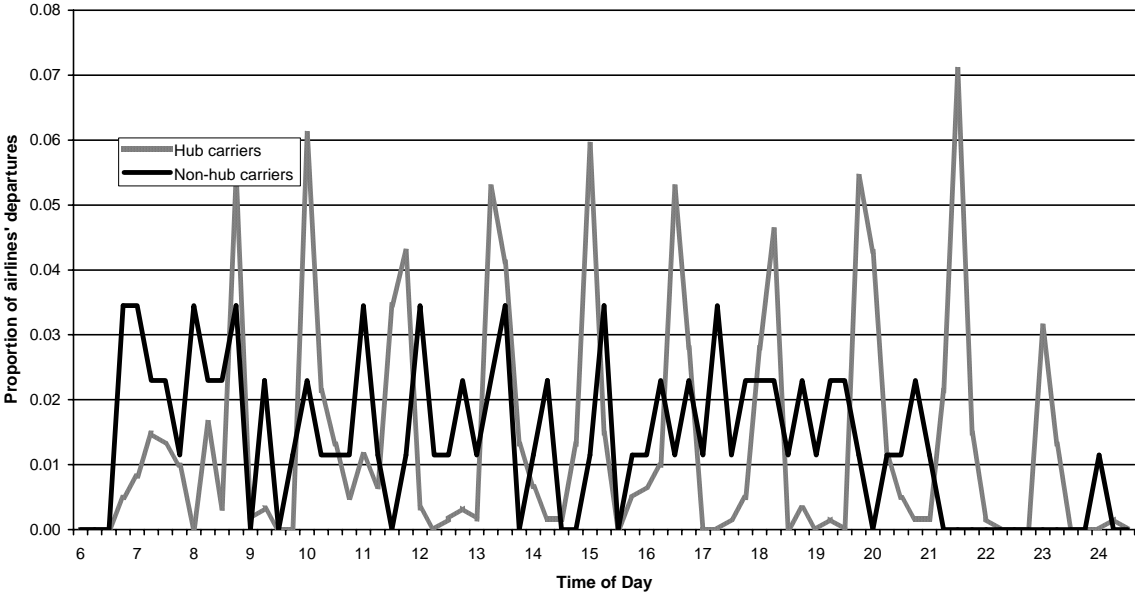


Figure 5: Departure density for American and Delta at Dallas-Fort Worth (DFW)

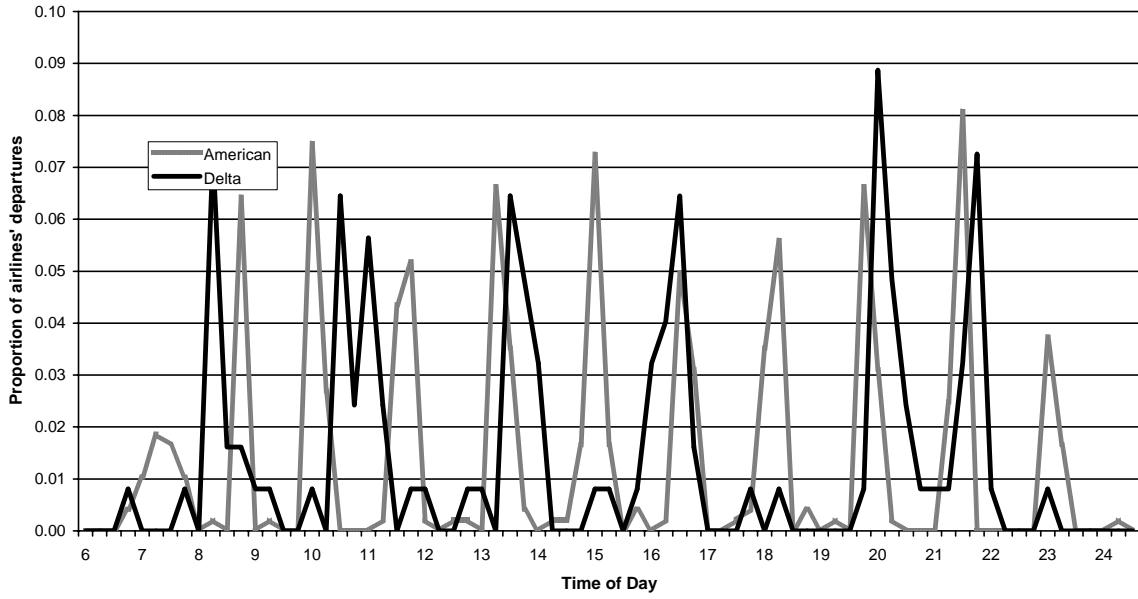


Figure 6: Percent On-time

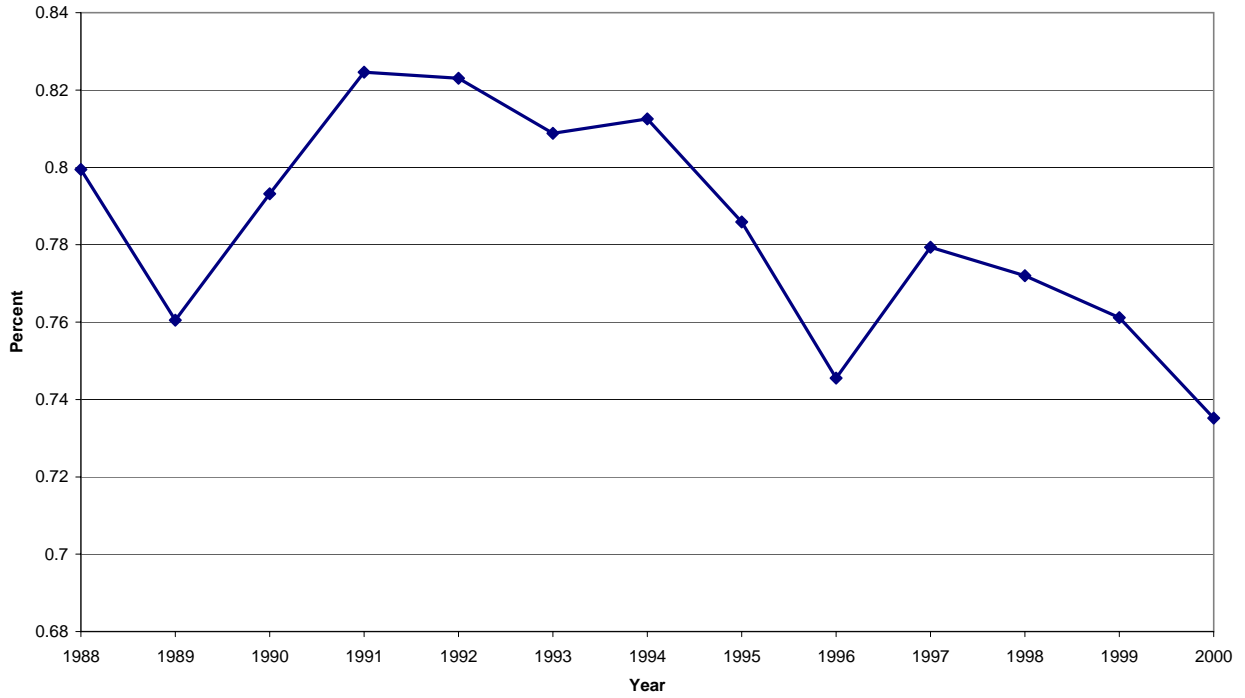


Figure 7: Minimum, Scheduled, and Actual Travel Times

