Congesting the commons: A test for strategic congestion externalities in the airline industry*

Alejandro Molnar†
Stanford University

January 14, 2013

Abstract

Access to scarce runway capacity at most airports in the United States is allocated by queuing, creating the scope for congestion externalities. Unlike a classic "tragedy of the commons", congestion externalities from airline decisions can be strategic. Airlines regularly schedule more departures than an airport's runways can handle without delay, and delay to the average US passenger was 50 minutes in 2007. I develop an empirical framework for airline schedule choices that accounts for benefits from scheduling flights close together in time - such as enabling connections and serving demand at preferred times of day - and the effect of congestion, which has both a cost and the strategic benefit of deterring entry by competitors. I use an engineering model of runway capacity and queuing to construct measures of marginal congestion, but the measures are endogenous in an airline's scheduling choice because they depend on other schedules. I exploit variation in runway capacity due to weather patterns within the day and across seasons, together with excluded variation in the schedules of rival users of the runway to estimate the effect of marginal congestion on airline scheduling at hub and spoke airports. I find that airlines trade-off benefits from connections and passenger preferred times against the cost of increased congestion, but this cost is outweighed at hubs by the strategic entry deterrence benefit of congesting peak times. The effect is larger at times that are more valuable to competitors.

JEL: D61, D62, H23, H42, L21, L50, L51, L93, R40, R41, R42, R48

Keywords: Airline economics, airline scheduling, entry deterrence, congestion, externalities, congestion pricing.

*I am grateful to Tim Bresnahan, Jon Levin and Kalina Manova for guidance and encouragement. Special thanks to Tim Armstrong, Kyle Bagwell, Julian Dannevig, Liran Einav, Silke Forbes, Jakub Kastl, John Lazarev, Paul Ma, Aprajit Mahajan, Amedeo Odoni, Mar Reguant, Nicholas Rupp, Paulo Somaini, Kester Tong, Vikrant Vaze and Ali Yurukoglu for helpful comments and discussions. All remaining errors are my own.

†Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, e-mail: amolnar@stanford.edu.

1 Introduction

Every day across airports in the United States there are periods of time in which commercial airlines schedule far more flights than can be handled by the airport's runways. For example, airlines scheduled an average of 62 departures from Newark Liberty International Airport between 8.00 and 9.00 a.m. during June 2007. At most 48 flights, and on average 43, managed to actually depart within that hour.

This paper studies the incentives that lead airlines to make infeasible schedules, such as those at Newark and at many other highly congested airports in the United States. The main operational constraint on an airport's capacity for flights to take off and land without delay is its system of runways.[1] A key feature of the regulation on runway access at most US airports is that access is allocated by having airplanes queue on the day of the flight, rather than having flights pre-allocated to ensure runway capacity can meet demand.[2] This creates the scope for congestion externalities.

The standard model of bottleneck congestion (Vickrey, 1969) assumes runway users are atomistic like drivers on a highway. This model is unattractive for runway congestion because commercial airlines are large players at airports - Newark is a hub for Continental Airlines, which scheduled 79% of the flights in the opening example - and congestion is both self-imposed and imposed on airlines that are competitors for passenger demand. In this paper, I investigate the possibility that airlines have strategic incentives to bear large amounts of congestion at their hub airports. An action that lowers the benefit to competitors from serving passengers at a given time (such as scheduling a block of flights that ensures an entrant would face congestion and delay) can have a strategic benefit. Conversely, an airline that sought to reduce self-imposed congestion by removing flights from a peak time might see competitor entry, which could have a competitive cost as well as offset the improvement in congestion.

I develop the first econometric model of airline schedule choice and use it to estimate the tradeoffs involved in scheduling peaks of flights (referred to as "banks" in the industry). I use these estimates to test whether airlines schedule flights at their hub airports to create congestion at peak times, which would indicate that the strategic benefits of deterring entry by competitors outweigh the direct cost of congestion.

[1] Rather than other airport infrastructure such as taxiways, aprons and terminal gates (De Neufville and Odoni, 2003). The requirement to lease terminal gates can act as a barrier to providing service from an airport (Borenstein, 1989, Ciliberto and Williams, 2010), but access to gates plays a minor role in flight delays.

[2] A small number of airports is under slot-control, discussed in Section 2.2. Some airports follow noise mitigation rules, which restrict capacity during certain hours but do not affect how aircraft are allocated to that capacity over the day.

To evaluate airline response and planning around congestion, I develop an engineering model of the incremental congestion that an airline should expect to arise from the addition of a marginal flight. The components of this model are runway capacities, which vary over the day in response to weather, the schedules of all airlines and a model of how capacity and runway demand interact: when scheduled demand for runways exceeds capacity, queues form and scheduled flights depart or arrive with delay. An airline's schedule choice has a marginal impact on excess demand for runway capacity, and through long-lasting queues can affect delays for flights scheduled at other times of the day. The model provides an estimate of the marginal congestion that an airline can expect to arise when it adds a flight to a peak time.

Airlines trade off the operational cost of congestion against benefits from scheduling banks of flights. Peaks arise to supply departures and arrivals at passengers' preferred times, and to coordinate departures to follow shortly after complementary arrivals and create the connections between multiple city-pair markets that are the reason for hub and spoke networks. The estimation framework allows for a rich model of an airline's benefits from scheduling operations at hub airports into peaks. The model accounts for potential connections to other domestic and international flights and codeshares with other airlines, and values connections under alternative layover times in an unrestricted way.

A hurdle to estimating the effect of marginal congestion and potential connections on airline schedule choices are that both of these quantities are functions of other schedules (by the same airline, and by its competitors), making them endogenous in a schedule choice equation. I exploit two exogenous sources of variation: First, weather patterns affecting airport capacity over the day are plausibly excluded from airline choices other than through their effect on congestion. Second, the paired-structure of airline schedules (a flight's scheduled time is determined by conditions at both origin and destination) can be used to instrument for the timing of a schedule at one endpoint of the flight, using (appropriately excluded) variation from the other endpoint. I use both of these ideas to generate instruments for the congestion measures, and a version of the second that additionally exploits the geographic structure of airline route networks to instrument for connections.

I find that conditional on connection and time-of-day benefits from scheduling banks of departures or arrivals, the strategic benefit of adding congestion to a peak time outweighs the operational costs for an airline scheduling flights at its hubs. For concreteness, consider how this result is obtained in terms of the weather instrument: if airport- and season-specific variation in intra-day weather patterns makes some times of the day likelier to result in additional congestion from a marginal flight, we would expect a non-strategic airline to reduce scheduling at such times. I find that airlines respond to congestion in this manner at the airports that are spokes in their network, but at hubs they add flights at times when weather enables the marginal flight to create entry-deterring congestion. An interpretation for this finding is that airlines over-congest attractive scheduling times at their hubs because the benefit of reducing congestion would be undone by competitor entry.

The results in this paper are consistent with a broader literature on the hub premium and market power arising from economies of density or limited airport resources,[3] but this paper is the first in that literature to investigate the use of runway capacity as a common-pool resource that can be congested to raise entry barriers for non-hub airlines. The paper also contributes to the literature on airport congestion and delays, of which two empirical strands reach conflicting results. One approach relates delay performance (Brueckner (2002), Mayer and Sinai (2003), Rupp, 2009) or a measure of bank characteristics (Ater, 2012) to market structure. This literature finds mixed evidence that the most concentrated airports do not appear to have the highest levels of delay, and this is attributed to large airlines "internalizing" the congestion they impose on themselves. The second approach is based on the only preceding empirical model of the timing of aircraft operations, the equilibrium-bottleneck model of Daniel (1995) and Daniel and Harback (2008). This approach finds the opposite result on internalization, which it attributes to strategic incentives.[4] This paper is the first in the airport congestion literature to explicitly estimate airline schedule choices, introduce a measure of ex ante congestion externalities and propose an empirical strategy to estimate the effect of congestion on schedule choice. An additional strength of the empirical strategy is that it allows for flight-specific determinants of schedule choice, such as airline connections and time-of-day effects at both endpoints of a flight. Taken together, these components provide a flexible model of the hub airline's non-strategic motives for scheduling flights into the peak of a bank.

The paper proceeds as follows: Section 2 describes costs and benefits of scheduling

[3] Borenstein (1989), Brueckner, Dyer, and Spiller (1992), Borenstein and Rose (1995), Berry, Carnall, and Spiller (2006), Hendricks, Piccione, and Tan (1997), Chen and Ross (2000), Ciliberto and Williams (2010), Lederman (2008), Aguirregabiria and Ho (2010), Snider (2009), Snider and Williams (2011).

[4] Daniel (1995) introduced the notion that a hub airline may choose to impose inefficient levels of congestion on itself if actions to reduce congestion could be undone by congestion from entry by other flights. However, the (markedly different) empirical framework in that paper relies on the assumption that all departures from an airport within a given time window share the same ideal departure time, regardless of destination or the identity of the airline. That assumption leads to rationalizing the actions of a hub airline that schedules its departures into a bank (and in particular into the peak of the bank) as evidence that hub airlines congest peak times strategically.

choices in the airline industry, as well as relevant institutional details on how access to runways is allocated in the United States. Section 3 describes the data and estimation sample. Section 4 describes how airline schedules are set up as a discrete choice. Sections 5 and 6 detail the two models for determinants of airline schedules: congestion and connections. Section 7 describes the empirical strategy, and introduces the instruments I use for congestion and connection terms. Sections 8 and 9 present results and conclude.

2 Scheduling in the airline industry

2.1 Costs and benefits from schedule choices

Scheduling is the component of airline strategy that involves the frequency and timetables of an airline's flights. It is one of many component parts of an airline's business strategy, which includes decisions over capital (e.g. choices of aircraft fleet and ground facilities) and labor (e.g. hiring, training and the homebase of crews), the segments to fly and therefore the origin-destination markets to serve, the assignment of aircraft and crews to specific flights as well as decisions over pricing and revenue management. Major national airlines also decide on vertical integration (e.g. aircraft ownership, maintenance or travel agent services) and horizontal integration (e.g. flights on specific routes), while smaller regional airlines decide whether to market flights directly to consumers or act as subcontractors for the major national airlines.

An airline's decisions over routes and schedules define a network structure of available passenger travel service to many origin-destination city-pair markets. To differing degrees, airlines adopt a hub-and-spoke network structure to share overhead costs and to connect routes (or segments) at hubs, allowing the airline to jointly supply more city-pair markets than the number of segments flown. The aggregation of demand from many origin-destination city-pair markets into each constituting segment may allow both for segments to cities that would otherwise not support point-to-point service, as well as higher frequencies on each segment.

Just as hubs act as a point in space at which to bring together and connect passengers from many inbound and outbound spokes, airlines seek to coordinate connections at a single point in time by scheduling their flights into so-called "banks". Banks are periods of time at an airport which have a high number of arrivals or departures. Departure banks follow arrival banks to enable connections with layovers that are as short as possible (see Figure 5 for profiles of scheduled departures at major US airports, including many with distinct departure banks). Scheduling is a component part of the network decision problem because it determines the length of the layover between connecting flights (and therefore the utility to passengers from the route provided to the connecting airline, relative to alternatives), as well as whether a connection is feasible.[5]

Airlines select schedules as a component part of their overall business strategy.[6] On the revenue side, timetable choice trades off factors that create residual demand for each segment: i) enabling origin-destination city-pair market service by creating potential connections, ii) the quality of the connections created in terms of travel time, both time in flight and over the layover, affecting the utility of the connection to passengers relative to rival alternatives and outside options, iii) proximity to passenger ideal travel times (e.g. at the start and end of the business day), and iv) cannibalization of demand for other segments in the airline's network that serve the same city-pairs. Airlines may also select departure times in response to competitor times on the same segment, e.g. closer to competitor times for business stealing, or further apart to create differentiation and weaken price competition (Borenstein and Netz, 1999).

Schedule choice also affects operational costs, such as the efficient utilization of fleet, crew and airport installations. For each time of the day there is a probability distribution for the delay that a flight scheduled at that time could incur, and an additional scheduled flight may have an external effect on delays for other flights. To discuss the cost of delays to airlines, consider the classification by Ball et al. (2010): for each (directed) pair of airports, aircraft type and airline policy on flight speed there is an "unimpeded flight time", which is the minimum time required to move the aircraft from gate to gate in the absence of other users congesting the airspace system and under standard weather conditions.[7] Due to runway congestion and other sources of delay, airlines announce a flight's scheduled arrival time to include a "schedule buffer" (or "schedule

[5]In addition to requiring a minimum time for connection, the maximum connection time for two domestic segments to be eligible as a single ticket is capped at 4 hours in the US.

[6]Despite the fact that network structure, route selection and scheduling (both frequency and timetables) are all part of an integrated problem for the airline, mathematical optimization of this problem is intractable and airlines do not solve the problems jointly in practice. Major components of the route structure such as "hub presence" as well as fleet choice are decided over long planning horizons based on long run forecasts for demand, cost and long run competitor strategies. Routes and schedules (both frequency and timetable) are modified individually and incrementally over planning horizons of 6 to 18 months, with the assistance of cost metrics and demand forecast tools that predict market shares on passenger origin-destination markets based on the airline's extant network and assumptions for competitor supply (see Barnhart (2009), Belobaba (2009) and Jacobs et al. (2012) on how the scheduling problem fits within airline strategic planning). Given the choice of routes and schedules, operations research methods are applied to a number of operational problems, e.g. revenue management and the allocation of available fleets and crews to implement the schedule choice, ensure specific aircraft receive proper maintenance, etc. (Gopalan and Talluri, 1998).

[7]Weather has a minimal effect on take-off and landing time in the absence of congestion, but can still affect flight time, e.g. due to tailwinds.

padding") in excess of this unimpeded time. The duration of a flight from gate to gate is referred to as its block time (actual, or scheduled). An actual block time is realized on the day of the flight, which may be greater or less than the scheduled block time. The difference between these is the "flight delay against schedule", which is what passengers experience as delay and the metric used to classify whether a flight is delayed in the Bureau of Transportation Statistics' On-Time Performance database and on airline ticketing websites that report on-time performance.

Increasing schedule buffer has direct costs for airlines. Most pilot contracts specify that pilots are paid for the maximum of scheduled and actual block time, and increased flight time reduces the number of feasible flights per aircraft. Buffer also lowers passenger demand: longer flight times discourage passengers from air travel, and fewer and lengthier flights mean less potential connections and feasible passenger itineraries. To the extent that it cannot be predicted and incorporated into scheduled buffer, flight delay reduces fleet utilization by requiring fleet assignments that are robust to disruption, has a utility cost to passengers (Forbes, 2008) and a direct costs to airlines when passengers miss connections. For the remainder of the paper, I use the word "delay" to refer to the sum of schedule buffer and flight delay against schedule, and will employ a flight-specific measure of unimpeded flight time in the empirical application to measure this delay.

2.2 Runway allocation at airports in the United States

The mapping of a flight's schedule to its expected congestion externalities depends on how runway access is allocated to scheduled flights. It is an established fact in the operations research literature on airport operations that delays arise predominantly due to bottlenecks in runway throughput capacity and schedule demand in excess of that capacity (De Neufville and Odoni, 2003).[8]

The most common practice outside the United States for dealing with demand for runway access in excess of capacity is the rationing of access rights: airlines are assigned "slots", which are caps on the number of operations the airline can schedule over a given time period.[9] The use of slot control in the United States is limited: three airports were

[8]Rather than due to bottlenecks elsewhere at the airport, such as ramps, taxiways or terminal gates. Some delay is caused by weather en-route or airspace congestion, mechanical failure, late arrival of aircraft due to delay elsewhere in the system, etc.

[9]More than 150 airports outside the US employ some form of slot control negotiated by airport stakeholders following guidelines set down by IATA, the main trade group for the international airline industry.

slot controlled in 2007, and five airports as of November 2012.[10] At non-slot controlled airports there are essentially no airport or government-mandated restrictions on runway access (Odoni, 2009).[11] To schedule a commercial flight between two airports, airlines only need to lease gate space at terminals and other on-ground services, and pay landing and user fees, which increase with aircraft weight but do not vary by time of day. On the actual day of the flight, access to the runway for departure or arrival is allocated in real time by air traffic controllers according to first-come, first-serve priority, with no discrimination between airlines or, in principle, between airline and general aviation users.[12] A flight obtains its runway allocation by actual queueing: an aircraft that wants to claim an allocation for departure contacts ground control for permission to "push-back" from the gate and join a physical queue on the airport's taxiways, burning fuel while wasting aircraft and passenger time.[13] For periods of the day in which scheduled demand exceeds an airport's throughput capacity, every scheduled flight is subject to the possibility of delay and imposes an expected delay externality on other flights scheduled at the airport.

2.3 Congestion and entry deterrence

This section describes an airline strategy that can lead to inefficient congestion: an airline may find it profitable to schedule flights at a hub airport in a way that causes runway

[10]Washington National (DCA), New York La Guardia (LGA), New York JFK (JFK) and Chicago O'Hare (ORD) were subject to slot control under the FAA's High Density Rule (HDR) until March 2000, when that rule was replaced by the Aviation Investment and Reform Act for the 21st Century (AIR-21). The act called for gradual phasing out of slots at LGA, JFK, and ORD. Slot control was removed at ORD in June 2002 without an immediate increase in delays due to decline in traffic following the attacks of September 11, 2001. With increasing traffic and delays from 2004 onwards, the FAA attempted to broker voluntary cutbacks in flights by United Airlines and American Airlines, but cutbacks were undermined by new entrants and the FAA reinstated slot control in August 2006. AIR-21 caused slots to expire at LGA and JFK on January 1, 2007, but at LGA these were immediately replaced by temporary FAA slot authorizations, which are ongoing. Due to severe delays in the summer of 2007, slot control was reinstated at JFK in March 2008. Slot control has remained in place at DCA throughout the period, and was reinstated at Newark (EWR) in June 2008 after several decades in which the airport was exempted from the FAA High Density Rule.

[11]A further set of restrictions arise for noise abatement at some urban airports (e.g. LGB, HNL, JFK, SFO, PDX, TPA). Noise abatement restrictions reduce runway capacity at certain hours of the day (e.g. limiting the use of throughput maximizing runway configurations), or cap the total number of daily operations, but do not allocate time slots to specific airlines.

[12]This principle applies to all airports that have been developed or improved with federal grants, a condition that applies to all major commercial airports. See *49 United States Code § 47101.*

[13]Pilot programs for "demand metering" that place aircraft in a "virtual queue" at the gate have started in the last couple of years at BOS, JFK, MCO and MEM. Although allocation mechanisms vary across programs, virtual queues still require queuing based on the aircraft committing to a claim for a "virtual" push-back time, so efficiency gains are mostly limited to the reduction in fuel burn.

congestion if this action deters competitor entry. I discuss how this strategy can be credible, can deter entry by raising rival's costs, and how deterrence can be profitable to an airline by preserving market power at hub airports. Because rival costs are raised through the use of a congestible resource, the congestion cost incurred by an airline to implement the strategy is offset by any congestion avoided through deterrence of rival users of the airport's runways.

To motive the entry deterrence argument, Figure 3 plots the departures scheduled by commercial airlines for the morning of June 13, 2007 at Newark Liberty International. Newark is a heavily congested airport and was a hub for Continental Airlines at the time: the airline scheduled 79% of the airport's commercial flights in June 2007. The figure also plots a horizontal line at 48 departures per hour, which was the 99th percentile of actual hourly departures in June 2007. The figure shows two periods of sustained scheduled demand beyond the limits of the airport's departure capacity (approximately between 6.30 and 7.30 a.m., and between 8 and 9 a.m.), most of which was scheduled by Continental.

Scheduling in excess of airport capacity, as in these two periods, occurred every weekday during June 2007 at Newark, and routinely resulted in high levels of delay, a large share of which accrued to flights by Continental Airlines.[14] A fraction of Continental's flights scheduled for the two peak periods that exceeded airport capacity invariably departed in the "valleys" that followed the peaks: during June 2007, the median Continental flight required 25 minutes (first period) and 34 minutes (second period) to taxi out to the runway, most of which was spent queuing.

Given that a large fraction of the flights in departure queues at any given moment were scheduled by Continental, the airline could have spread out its departures by removing flights from the peaks around 7.00 a.m. and 8.30 a.m., and into the times following the departure banks, at which many of the flights actually departed. All else equal this would have reduced overscheduling and queues, and brought gate departure times closer to the times at which aircraft actually took off from the runway. If Continental had taken this action, however, the lower expected delay at these attractive morning times might have invited entry from competing airlines and general aviation users, and this entry would have imposed congestion externalities on Continental's flights.[15]

[14] Delays are discussed in more detail in Section 5.

[15] Airport operations researchers and hub airlines have referred to the issue of de-peaking leading to competitor entry as the "backfill problem", see Hansen et al. (2001). A rival theory of peak scheduling might be that airlines schedule peaks of departures as "options" to depart aircraft according to their uncertain arrival time. This is not how airlines schedule departures, see Lan et al. (2006) and Ball et al.

The scarcity of runway capacity and the externalities generated by scheduling banks of flights in excess of airport capacity provide an opportunity for airlines to engage in behavior that raises rival costs (Salop and Scheffman, 1983, 1987). Flights scheduled during peaks that exceeds capacity can be profitable for the hub airline while not being profitable for the non-hub airline because the former benefits from additional demand from potential connections, economies of density and its status as a hub carrier.[16] To the extent that the hub airline and potential entrants would compete for passenger demand at peak times, the hub airline's practice of scheduling banks of flights to multiple destinations provides a competitive benefit similar to entry deterrence through product proliferation (Schmalensee, 1978) and the distortion of location choice (Bonanno, 1987; Borenstein and Netz, 1999).

Competing airlines may be deterred into scheduling at less congested times (at the expense of supplying passengers with less preferred travel times), providing service with lower frequency (lowering the quality of their service from this market, both direct and connecting through hubs), or dropping service to some markets altogether, helping to preserve the hub premium. If deterrence through strategic congestion does in fact occur, the marginal congestion to the hub airline is only the difference between congestion under deterrence and congestion from otherwise deterred commercial airlines and other airport users (such as general aviation, freight and charter flights). The fact that schedules are decided and entered into ticket reservation systems 6 to 18 months ahead of the flight date makes a strategy of scheduling such a bank of flights credible to potential entrants: as the flight date approaches, the hub airline sells tickets (both direct and connecting) on each flight scheduled in the bank. This creates an exit cost for removing flights from the bank, which is the condition identified by Judd (1985) for entry-deterring proliferation to be credible.

Previous empirical work on airport congestion has highlighted particular elements of the entry deterrence argument described above. Daniel (1995) first argued that an airline may schedule an inefficiently large number of flights during a peak period if an attempt to reduce congestion by lowering the number of flights would be undone by additional

(2007) on airline planning practices for robustness to delay. Additionally, morning departure delays at Newark were not due to late arriving aircraft: according to Bureau of Transportation Statistics On-Time Performance data for June 2007, only 13.5% of delayed flights that were scheduled to depart before 12.00 p.m. had a fraction of their delay attributed to late arriving aircraft.

[16]On economies of density see Caves, Christensen, and Tretheway (1984), Brueckner, Dyer, and Spiller (1992) and Brueckner and Spiller (1994). For the hub premium see the discussion in Borenstein and Rose (2011), and in particular Borenstein (1989, 1991); Evans and Kessides (1994); Berry, Carnall, and Spiller (2006) and Lederman (2007, 2008).

entry.[17] However, Daniel's empirical approach to testing this prediction (see also Daniel and Harback, 2008) rests on the assumption that both the hub and non-hub airlines value alternative schedule times in the same way. In particular, the fact that airport banks are largely composed of flights by the hub airline is directly rationalized as support for his prediction. Because air travel is a time-specific commodity that is sometimes supplied by combining inbound and outbound flights at hub airports, the value of scheduling a flight at a particular time may differ substantially between airlines. We should therefore expect banks of flights at hub airports to be composed mostly of flights by the hub airline, regardless of strategic considerations.

In contrast to Daniel, Mayer and Sinai (2003) propose an analytical framework in which banks of flights arise only due to connection benefits for the hub airline and there are no strategic interactions between airlines in scheduling. The authors estimate regressions of flight delay performance on airport characteristics and find that delays are i) associated with hub airports and largely borne by the hub airline, and ii) decrease only slightly with airport concentration. The paper attributes the association between delays and hub status to an efficient airline tradeoff between connection benefits and congestion.

Examples of over-scheduling at congested airports (see Figures 3 and 5) and the fact that strategic considerations are discussed by airport operations researchers and practitioners (Hansen et al., 2001) suggest the need for an empirical approach to explicitly test for entry deterrence. Empirical studies of entry deterrence are rare because an action taken to deter entry may be seen as one of the benefits of a competitive market.[18] In the current setting, this may include scheduling service at peak times, increasing the number of destinations served or the frequency of service. Serving passenger demand and creating potential connections provide non-strategic reasons for hub airlines to schedule peaks of flights, and could account for the suggestive patterns at Newark and other congested airports. The empirical strategy in this paper will account

[17] The airport congestion literature refers to this prediction as an airline not fully "internalizing" the congestion effects that its flights have on each other. The prediction stands in contrast to a standard "tragedy of the commons" outcome in which airlines are assumed to internalize the congestion they impose on themselves but not on rivalrous users of the runway, and inefficient congestion decreases with airport concentration as a larger share of the marginal damage from each scheduled flight accrues to the airline that schedules it. See Brueckner (2002, 2005) for statements of the standard result, and Brueckner and Van Dender (2008) for models with incomplete internalization.

[18] The empirical literature on entry deterrence includes Kadiyali (1996), Ellison and Ellison (2011), Dafny (2005), Lieberman (1987), Chevalier (1995a,b), Scott Morton (2000) and Kadiyali and Conlin (2006). Goolsbee and Syverson (2008) study how incumbent airline fares respond to the threat of entry from Southwest Airlines, either to deter or accommodate entry, and Prince and Simon (2012) apply the same identification strategy to delay performance.

for such determinants, both observable (through a model of potential connections), and unobservable (through time-of-day effects and an instrumental variables strategy for marginal congestion measures).

3 Data

Comprehensive data on commercial airline schedules is from Official Airline Guide North America and Worldwide edition flight planner CDs, for the months of January, February, April and June 2007. Unlike the Bureau of Transportation Statistics' On Time Performance data used in previous research in the congestion delay literature, the OAG data includes all scheduled commercial carriers. Accounting for all operations is essential to the empirical strategy in this paper. The schedule data includes international flights, codeshares and the "doing business as" names of regional airlines, which is required to correctly attribute flights by regional airlines to the mainline carriers that contract them. In the estimation I use schedules from Wednesdays in the middle of each of these four months.[19]

Engineering estimates of airport capacity are from the Airport Capacity Benchmark Report 2004 (ACBR), which includes engineering estimates of airport capacity for the 35 "busiest" airports in the United States, as well as projected benchmarks for the capacity expansions that were underway at the time the report was drafted. The sample of schedules that can be used in estimation is limited to airports for which we can construct measures of congestion, which limits the estimation sample to flight segments between the 35 airports included in the ACBR.[20] In addition to the restriction imposed by availability of capacity estimates, I drop segments to and from the three airports (Washington National, New York La Guardia and Chicago O'Hare) that were under slot control during the sample dates in 2007, as well as from Honolulu International Airport. Route maps are displayed in Figure 4 to describe the sample: Panel 4a shows the segments between 31 airports that are used in the estimation sample. Connection data is not limited to routes within this sample, however, and includes domestic, international, carrier and codeshare connections. Panel 4b shows the segments in the sample for American Airlines. Panel 4c shows American Airlines domestic routes connecting to any of the American Airlines segments in the sample, excluding codeshares (AA connections look sparse in the US northwest because the airline codeshares extensively with Alaska

[19]The dates in the estimation sample are January 17, February 14, April 18 and June 13, 2007.

[20]Flights between these airports account for 60.2% of the flights in the On-Time Performance database over the sample period, with 94.9% taking off or landing at one of these airports.

Airlines in this region). Panel 4d displays American Airlines international routes. Panels 4f and 4e show the segments in the estimation sample and connecting domestic segments for Southwest Airlines.

Data on intra-day time paths of airport benchmark capacities are constructed from 5 years of hourly data on visibility and cloud ceiling at each airport's meteorological station, obtained from a National Oceanic and Atmospheric Administration's National Climatic Data Center, as well as airport-specific mappings from meteorological conditions to throughput capacity rates defined in ACBR (2004). Residual capacity for commercial aviation is estimated from historical data on general aviation runway use, obtained from the FAA's Enhanced Traffic Management System database.

Minimum connection times are obtained from tables of International Air Transport Association (IATA) Minimum Connecting times, which are determined by airport stakeholders as the shortest time interval required to transfer a passenger and her luggage from one flight to a connecting flight. Minimum connecting times vary by airline, connection airport, and whether a connection is domestic to domestic, domestic to international or international to domestic, as well as by origin or final destination airport in a small number of cases. With some exceptions, IATA minimum connecting times are observed by all ticketing systems when building connecting tickets.

4 An empirical framework of flight schedules as discrete choices

An airline's decision to schedule flights into banks of arrivals and departures is the aggregate result of many schedule choices over individual flights. In this section I develop an empirical framework for modeling airline schedules as dependent variables, which I will use to estimate airline tradeoffs between attributes of alternative times as a discrete choice.

Flight schedules specify both a departure and an arrival time. I use the departure time to track the scheduling decision, as scheduled gate departure time is used to coordinate passengers to show up at the airport gate and is the time at which the pilot plans to make a claim for access to the runway for take-off. Given a scheduled departure time, the airline can model flight-time and congestion at origin and destination to produce a probabilistic forecast for the time of arrival, and use this forecast to issue a scheduled arrival time. I map scheduled departure times to arrival times for connections and for congestion to account for determinants of scheduled choice at destination.

The unit of observation is the individual flight, which is identified by an origin (o), a destination (d), a controlling airline (a), a calendar day (y) and a flight number (f). Flight number f indexes the set of flights for a given *oday*, in order of departure time. I model departure time as a location choice over a time interval: each alternative time has costs and benefits given passenger preferences, operational costs and other scheduled flights by the same airline and its competitors, and the relationship between payoffs for alternative times will be modeled explicitly and through an econometric error term. I define the time interval on a day y to start at 3.01 a.m. and end 24 hours later at 3.00 a.m. and I discretize time into 15-minute periods, for a maximum 96 alternative flight times within a day. To model the dependence between schedule choices on the same *oday*, I define the set of alternatives for a flight f to be each of the periods between the observed scheduled time for flights $f-1$ and $f+1$. If a flight's predecessor and successor are scheduled for periods t_{f-1} and t_{f+1}, respectively, there will be $J_f = t_{f+1} - t_{f-1} - 1$ alternative schedule times available to flight f. For the first (last) flight of the day, I define the choice alternatives to be the periods between the first (last) period of the day and the flight's successor (predecessor). This definition of a choice set as a range of alternatives between observed predecessor and successor will allow me to model the dependence of schedule choices within the day.

There are three ways that time-varying covariates enter alternative-specific payoffs given this choice structure. 1) *Absolute time-indexed* variables enter an alternative payoff purely as a function of the absolute time of day: for any flight f their value does not depend on schedule choices on other flights f'. An example is time-of-day effects, and Figure 6 provides a simple example of how 3 flights over a total of 10 periods are structured into individual discrete choices with an absolute time-indexed regressor. In this example, the value of regressor x_4 enters the first choice in the fourth alternative and the second choice in the second alternative. 2) *Alternative and time-indexed* variables are a function of both absolute time and schedule choice on neighboring schedules. Examples will include connection and congestion terms to be discussed in the next two sections. 3) *Alternative-invariant* regressors, as they are commonly referred to in the multinomial choice literature, are a function of the alternative identifier and variables that are common across alternatives. In the schedule choice problem, alternative identifiers do not map to any attributes of the alternative other than its proximity to other departures on the same *oday* schedule.

The main drivers of an airline's decision to schedule flights into banks of arrivals and departures are the benefits from flying at passengers' most preferred times and enabling connections, and the effect of congestion, which has an operational cost but may also provide a benefit if it deters competitor entry. The next two sections describe the measurement of these components and how they are included in the schedule choice framework.

A limitation of this approach is that scheduled time is treated as a dependent variable conditional on the airline's decision to fly on the segment and its frequency of service. The approach is therefore not directly applicable to estimating the determinants of where airlines fly (such as market size and competition, as studied by Berry (1992) and Ciliberto and Tamer, 2009), or the frequency of service.

5 An engineering model for the measurement of marginal congestion

In this section I describe an engineering model for the ex ante measures of delay that arise from the interaction of runway demand from airline schedules and uncertain airport capacity. Ex ante measures are relevant to choice because schedules are made 6 to 18 months ahead of the flight date. The interaction between scheduled demand and an airport's limited throughput capacity is modeled as a first-come, first-serve queue at the airport level. I take a particular time-path for an airport's runway capacity over the day (e.g. 8 departures per quarter hour under heavy fog at Newark Airport until 9 a.m. and 12 departures under optimal weather thereafter) and compute the delay that would accrue to each scheduled flight under that path. I use observed historical weather to construct empirical distributions of the time-paths of airport capacity, flexibly capturing the shape of capacity time-paths within the day across airports and seasons. Airline schedules are put through the queueing model to obtain the delays for each flight under each capacity time-path. The mean and variance of delay for each flight are then obtained by integrating delay over the empirical distribution of the capacity paths. Perturbations from the observed schedules are then used to compute the marginal effects of alternative schedule choices.

The first component of the congestion model are the empirical distributions of airport capacity paths. An airport's throughput capacity is determined by a large number of factors, the most important of which is the number and layout of its runways. Intra-and inter-day variation in capacity, however, depend mostly on weather: airport-specific rules on cloud ceiling and visibility constrain air traffic management procedures at each airport.[21] The FAA Airport Capacity Benchmark Report (2004)

[21] In particular, runway configurations and aircraft-separation requirements (i.e. constraints on how close

contains engineering estimates of capacity for the 35 busiest airports in the United States under three meteorological condition-based protocols: Visual flight rules, Marginal VFR, and Instrument flight rules. The ACBR defines its capacity benchmarks as "the maximum number of flights an airport can routinely handle in an hour, for the most commonly used runway configuration in each specified weather condition".[22] Runway capacity can be allocated to departures, arrivals, or a mixture of both, so air traffic controllers effectively face a budget set for arrivals and departures and may in principle choose different runway configurations over the day based on demand for take-offs and landings. This is not the case at most congested airports in the United States, where runways are dedicated to specific operations to avoid switching costs from mixed use, but there are exceptions (e.g. Cincinnati and Washington Dulles). I assume a constant runway allocation throughout the day, and I assign benchmark capacity to arrivals and departures based on the single pair of arrival and departure rates reported by air traffic controllers in the ABCR 2004, except at Cincinnati and Dulles where the ACBR 2004 reports rates under two configurations and I use the outer envelope of arrival and departure capacities.

A day's weather can be described by a time-path for meteorological conditions, i.e. a function $w : t \to \{VFR, mVFR, IFR\}$. To describe weather for a particular date I take the 31 day window around the date in the prior 5 years, yielding 155 calendar days of seasonally-comparable weather. For each airport, weather state and time-of-day I also estimate runway use from non-commercial aviation (details in Appendix A), and remove the capacity used by these users from the throughput capacities provided in ACBR 2004, yielding a set of residual capacity time-paths $C_s : t \to \mathbb{Z}_+$, indexed by s, and associated empirical probabilities $\hat{p}_{sy}$ for each state s and calendar day y.

I use the queuing model detailed in Appendix B to compute $D_t(s)$, the delay to a flight scheduled at t under capacity-path state s and the aggregate schedules of all commercial airlines. A flight's expected departure delay and variance of delay are obtained by integrating over the empirical distribution of the capacity-paths:

$$\mathbb{E}[D_t] = \sum_s D_t(s)\, \hat{p}_{sy} \tag{1}$$

$$\mathbb{V}[D_t] = \sum_s [D_t(s) - \mathbb{E}[D_t]]^2\, \hat{p}_{sy} \tag{2}$$

any two aircraft can be in both space and time, see De Neufville and Odoni, 2003).

[22] Additional weather conditions such as precipitation and wind do affect runway operations and throughput capacity, but there is currently no study that benchmarks their effect on capacity across multiple US airports.

I assume that the marginal congestion effect relevant to an airline's schedule choice is the sum of the incremental effects on all of the airline's delays, which is an assumption of linearity both in the cost to the airline from the duration of a flight's delay and in delays across the airline's flights.[23] Define $\mathcal{H}_{oy}$ as the set of all scheduled departures from airport o on day y, and $\mathcal{H}_{oay} \subset \mathcal{H}_{oy}$ as the subset of schedules owned (or controlled) by airline a, of which an element h is scheduled to depart at $t(h)$. The sum of the expected delays to airline a flights departing from o is then

$$\sum_{h \in \mathcal{H}_{oay}} \mathbb{E}\left[D_{t(h)} | \mathcal{H}_{oy}\right]$$

The incremental sum of expected departure delay to airline a from a flight $odayf$ scheduled at t is defined as

$$\mu_{odayft} \equiv \sum_{h \in \mathcal{H}_{oay}} \mathbb{E}\left[D_{t(h)} | \mathcal{H}_{oy}\right] - \sum_{h \in \mathcal{H}_{oay(-odaft)}} \mathbb{E}\left[D_{t(h)} | \mathcal{H}_{oy(-odayft)}\right] \tag{3}$$

and the incremental sum of the variances of delay is defined as

$$V_{odayft} \equiv \sum_{h \in \mathcal{H}_{oay}} \mathbb{V}\left[D_{t(h)} | \mathcal{H}_{oy}\right] - \sum_{h \in \mathcal{H}_{oay(-odayft)}} \mathbb{V}\left[D_{t(h)} | \mathcal{H}_{oy(-odayft)}\right] \tag{4}$$

Computing μ or V for a flight identified by $odayf$ with alternative schedule choices $j = 1, \ldots, J$ mapped to departure times $t_{j,f} = t_{1,f}, \ldots, t_{J,f}$ requires computing the incremental effect for each alternative relative to the aggregate schedule net of the flight at the time it is observed to be scheduled in the data. I obtain delay metrics for arrival at the destination airport d in an analogous manner, but map scheduled flights to arrival delay metrics using the scheduled departure time plus the computed expected departure delay $\mathbb{E}[D_t]$ and unimpeded flight time.

Table 5 displays summary statistics of expected departure delays computed from the congestion model. I present airport-level statistics of the incremental sum of expected departure delays from the addition of a marginal flight in the first quarter of the hour, from 7 a.m. to 10 p.m. given all other flights as were actually scheduled. The first nine columns present statistics for the incremental congestion effect on all flights, regardless of airline ownership, and the last five columns present statistics for effects calculated at the airline level. Marginal effects of about 1 aircraft-hour of incremental delay are not uncommon and are near the mean for the two New York City airports in the sample. For some airports (Orlando, Pittsburgh, St. Louis) the model calculates no impact on

[23] Relaxing this assumption requires pre-committing to a non-linear specification or computing the congestion model for each trial value of a curvature parameter, which is computationally infeasible. Modeling the variance of delay is intended to partly capture non-linear marginal delay effects.

expected delays from adding a marginal flight, as scheduled runway demand is always at least one departure below the lowest benchmark airport capacity.

As validation for the congestion model, I find that expected delay and variance of delay of a scheduled flight predict both the amount of "buffer" schedule time that airlines add to schedules (Table 6, Column 3) and actual delays (Column 7), for the set of flights between the 31 airports in the main estimation sample. An additional minute of expected departure delay is associated with an additional 0.336 minutes in additional schedule buffer and 0.511 minutes of actual delay.[24] These results hold both when expanding the sample to departures to destinations outside the 31 airport sample (Columns 1 and 5 include departures to all domestic and international destinations) and when controlling for airport and time effects (Columns 4 and 8), so the coefficients on the congestion metrics are estimated from variation within time-of-day and across calendar days. Similar results hold for arrival congestion metrics, but the association between expected arrival delay and airline schedule buffer is weaker.

Figure 7 provides examples of the intra-day variation in the metrics produced by the congestion model for Newark Liberty International, a heavily congested airport. For every 15-minute period in the day, the figure plots the incremental sum of expected delay (eq. 3, plotted with solid lines) and variance of delay (eq. 4, dashed lines) that would accrue to all Continental Airlines flights scheduled to depart from Newark, using schedule and capacity data for January 17 and June 13, 2007. Congestion effects spike at 6.30 a.m. as airlines begin scheduling departures in excess of airport capacity, and incremental expected delay reaches a peak about two hours later.

As an example of the schedule data that underlies the spike in incremental congestion around 8.30 a.m. airlines scheduled 20 or more departures from 8.25 a.m. to 8.40 a.m. in each of the sample dates, well in excess of Newark's benchmark departure capacity of 12 departures per quarter hour under optimal weather.[25]

Data on actual delays from the Bureau of Transport Statistics' On-Time Performance database is consistent with substantial delays around the 8.30 a.m. peak in departures at Newark. For the sample of all weekday flights scheduled from 8.25 a.m. to 8.40 a.m. during the first half of 2007, the median delay against schedule was 2 minutes, the average 8 minutes, and 10% of flights were delayed 40 minutes or more. Scheduled arrival times include buffer for expected delay, so one approximate measure

[24]Mayer and Sinai (2003) find that airlines do not fully incorporate congestion into schedule buffer.

[25]Actual historical operations data from the FAA's Enhanced Traffic Management System (ETMS) validates the benchmark rate: at most 42 aircraft departed in either the 8 to 9 a.m. hour or the 9 to 10 a.m. hour in any of the four sample dates. The 99th percentile of actual departures per hour from Newark during the first half of 2007 was 49, one departure more than the benchmark rate.

of operational delay is the difference between actual flight time and its 10th percentile within origin-destination-airline.[26] Taking this percentile over the sample of all flights in the first half of 2007, the median delay was 28 minutes, the average 31 minutes and 10% of flights were delayed 55 minutes or more. An alternative metric of actual departure queues out of Newark is taxi-out times in excess of a frequently observed minimum. The median departure from Newark spent 36 minutes taxiing and the mean 39 minutes, compared to a 10th percentile taxi-out time of 14 minutes and a minimum of 2 minutes. Additionally, 3.5% of departures were cancelled.[27] Although departure delays around 8.30 a.m. appear to have been a repeated and predictable outcome at Newark, a single airline (Continental) accounted for 91.3% of the scheduled departures in the four sample dates.

6 A model of timed connections between flights

In the hub and spoke system used by major US airlines, connecting traffic is an important determinant of the value of a particular schedule time for a flight. A flight that departs shortly after an arrival "bank" generates value for an airline by enabling many potential connections. Further, the value will be higher if passengers on the arriving flights would demand continuing service on to the destination of this particular flight. Similarly, a flight that arrives shortly before a departure bank will gain value from connections, higher value if many of its passengers would continue on to the destinations served by the departing flights. In short, airlines create value for their network of flights by coordinating arrivals and departures to serve city pair markets that will be well-travelled by customers. The timing of this coordination is a key component of schedule choice. Some times will rule out potential connections, and too-short layovers may mean missed connections in the event of delay. Too-long layovers lower product quality by increasing total travel time for the customer. In this section, I construct measures of the value of connection from the perspective of a particular flight, holding constant other flights' schedules, including the timing of banks. These connection values are time-specific, allowing bank coordination to form an important element of my empirical model of

[26] This metric is used in Mayer and Sinai (2003), Rupp (2009) and Ater (2012).

[27] Actual delay on the 4 sample dates turned out worse than these statistics, as two of the four dates resulted in unfavorable weather. Median flight delay against schedule was 13 minutes (20 minutes on average), median flight time against the 10th percentile was 34 minutes (36 minutes on average) and the median taxi-out time was 41 minutes (45 on average), with 30.5% of flights cancelled, based on 61 flight records in the BTS database with a scheduled departure time between 8.25 a.m. and 8.40 a.m. on the four sample dates, relative to 81 records in the schedule data. The discrepancy is due to BTS data lacking international flights and smaller domestic carriers.

schedule choice.

Consider Figure 8a, which provides an example choice set with $j = 1, \ldots, 11$ alternative flight times for a flight f. A number of inbound flights whose passengers would become available for connection in each period j is depicted by the height of the blue bars in the diagram (and summarized by $x_{t(j)}$). The periods included in the choice set are defined by a predecessor on the same segment scheduled at $t^*(f-1)$ and a successor scheduled at $t^*(f+1)$, which in the example are 3 hours apart leaving 11 quarter-hour period alternatives.

Every alternative schedule implies a different set of potential connections: if flight f is scheduled to depart in period 1 (i.e. 15 minutes after the previous departure to the same destination) it creates an immediate connection to $x_{t(1)} = 2$ arrivals, but passengers arriving at times $t(2)$ through $t(11)$ would be unable to connect to this flight and would have to wait until $t^*(f+1)$ to connect through to destination d. If f is scheduled for period 2, the $x_{t(1)} = 2$ arrivals can still connect to f with a 15-minute longer layover, but passengers on flights arriving at $t(2)$ would now be able to connect to f, reducing the layover for connection on to d by 10 periods, or 2 hours and 30 minutes.

Let β_k be the function from a layover length k to the value for the airline of making a connection, so that β_0 is the value of making a connection that is as short as feasible, β_1 is a connection that is 15 minutes longer, and so on. A decision to shift a scheduled departure from period 1 to 2 would shift the connection value of the period 1 arrivals from $\beta_0 x_{t(1)}$ to $\beta_1 x_{t(1)}$, and the period 2 arrivals from $\beta_{10} x_{t(2)}$ to $\beta_0 x_{t(2)}$.

Appendix C generalizes this example to show how we can estimate an unrestricted layover discount function as a set of parameters β_k using the mapping from the set of inbound flights to layover-specific potential connection terms under each alternative.[28] The value of connections to the airline may depend on more attributes than the count of layover-discounted potential connections. In the empirical section I estimate a scale coefficient and an independent discount function for each of four types of connection: domestic-domestic, domestic-international, domestic-domestic codeshare and domestic-international codeshare. However, within each of these connection types a count of arrivals into the airport would assume that all inbound connections are equally valuable to an airline for onward connection, regardless of destination. In practice, airlines

[28]We can invert this statement from parameters to choices to make the revealed preference argument clearer: given a layover discount function β_k, we can compare the values of alternative schedule choices to the airline and all else equal derive the optimal schedule. The solution for the example in Figure 8a assuming an exponential layover discount with a single parameter $\beta \in [0, .35)$ is $j^*(f) = 7$, and for $\beta \in (.35, 1)$ the optimal choice is $j^*(f) = 8$. For $\beta = 1$ all choices are optimal, and for $\beta = 0$ the optimal choice is always the period with most arrivals, in this case period 7.

schedule some banks to have complementary geographic orientations. As a simple example of how this may affect schedule choice, consider Figure 8b where we now assume that arrivals are divided into two mutually exclusive sets: a set x_d arriving from the east, connection to which the airline only values if it allows passengers to continue traveling westbound, and a set $x_{d'}$ of flights arriving from the west for which the airline only values connections eastbound. Alternative-specific connection values now depend on the destination of the departure.[29]

More broadly, if the value of a potential connection to the airline depends of its value to passengers as a means of travel between the city-pair market, this will depend on its geographic characteristics: for a given city-pair market served through connection, a less direct connection is worth less to passengers due to increased travel time and is likelier to be dominated by alternative routings offered by the airline's competitors. I define the "directness" of a connection to be the ratio of the great circle distance between the city-pair endpoints and the distance over the connecting path, e.g. for an inbound flight from o' to o for connection on to d the directness is:

$$directness\left(o', o, d\right) = \frac{\text{gcd}(o', d)}{\text{gcd}(o', o) + \text{gcd}(o, d)}$$

and where gcd (a, b) is the great circle distance between airports a and b. Figure 9 is a single example from the schedule data that shows that the directness measure captures a first-order feature of real airline schedules. It shows all United Airlines domestic flights scheduled to arrive at Denver International Airport on June 13, 2007, each flight scaled by the directness of the route from its origin through Denver and on to either Boston or San Diego. The figure shows the period immediately after 8 a.m. to be suited for more direct connections to San Diego than Boston, which is due to a large number of inbound early-morning flights from the Midwest. Similarly, the period from 10 a.m. to 11 a.m is more suited for direct connections to Boston due to recently arrived early-morning flights from the West Coast.

I provide descriptive evidence of the relationship between connection directness and its value to the airline through passenger willingness to pay, as measured by ticket fares, in Table 2. For the subsample of two-segment tickets included in the BTS DB1B 10% sample of tickets, I regress log fares on distance flown as well as on directness, controlling for fixed effects at the level of the origin-destination city-pair market and quarter. The origin-destination market effects absorb variation in fares due to point-to-

[29]If two schedule choices indexed by $odayf$ and $od'ayf$ differed only in terms of these two profiles of arrivals, the optimal schedule for the former would continue to be $j^* \in \{7, 8\}$, but the optimal schedule for the latter would be $j^* = 6$.

point distance and other market characteristics, including the competitive environment on the city-pair, so the effect of distance flown or directness on fares is estimated purely from variation in connection paths within market and quarter. The results in Column 1 show that a routing with 10% greater distance flown is associated with 1.1% lower fares. Column 2 presents the results for directness, which are consistent with the results for distance, but not as readily interpretable. For ease of interpretation, Column 3 presents the results for directness binned into four intervals, and Figure 1 provides context for these intervals.[30] Column 3 indicates that routes following an approximately straight line path have fares 8.7% higher than routes with a path that is more than double the straight line length, and also shows the relationship to be monotone.[31] The value to an airline from an indirect (or circuitous) connection is further decreased by the fact that, among potential routes serving an origin-destination market, a less direct connection involves more flight time and fuel cost per passenger.

To estimate the schedule choice model, I define the inbound (at origin) and outbound (at destination) connection measures x_t for each particular segment *oday* and connection type (domestic or international, intraline or codeshare) to be the sum of the directness-scaled flights that become available for connection at time t.[32]

The connections model exploits the geographic structure of airline networks to generate variation in the value of specific times across the flights of the same airline, which helps identify layover terms. A restriction in the specification is that connection payoffs are additively linear in the directness of each potential connection. Given this restriction, airline choices for flights bound for different destinations help identify layover discount parameters from the same set of inbound flights. Consider the example choices in Figure 10: If, all else equal, we observe in Choice 1 that the airline prefers to make 2 angled connections with a one-hour layover over 3 angled connections with a two-hour layover, this puts an upward bound on the airline's tradeoff between our measure of angled connections and the additional layover from one to two hours. If we

[30]The first interval spans routings with a directness similar to Los Angeles - Chicago - Washington (LAX-ORD-IAD) or better. The second interval spans routings with an angle between that of LAX-ORD-IAD and LAX-ORD-Miami. The third interval spans routings with an angle between that of LAX-ORD-MIA and LAX-ORD-Dallas Fort-Worth, and the omitted category includes intervals with a routing less direct than LAX-ORD-DFW. Routings at such angles are rare for transcontinental distances, but not at the scale of regional markets.

[31]Although stark, these results do not establish a causal relationship. For example, there may be association between direct routings and other in-flight amenities.

[32]To rule out feasible but economically irrelevant connections, such as flying from a to b and connecting back to a or airports near a, I set a lower bound on the directness required for counting a connection: for each airport I take the set of all passengers flying two-segment connections through that airport in the DB1B ticket data for 2007, and set the lower bound at the 0.1 percentile of the directness measure observed on these connections.

observe in Choice 2 that the airline prefers 2 equally angled connections with a two-hour layover over a single, one-hour angled connection, this puts a lower bound on the same tradeoff. For the same set of inbound flights we observed in Choice 1 but with a rotated destination (Choice 3), the change in directness angle rules out an inbound connection from each alternative (as the angle becomes too acute and the connection too indirect for passengers to demand it), and slightly improves the tradeoff between connection angles and layover, sharpening the identification of the lower bound of this tradeoff.

7 Empirical strategy

The framework developed in Section 4 allows us to model as choice alternatives the times of the day at which an airline schedules a particular flight. To estimate the effect of strategic congestion motives on airline scheduling behavior, we are interested in quantifying the effect of marginal congestion on airline's schedule choices under conditions in which increasing congestion can deliver the strategic benefit of deterring competitor entry by reducing the value of that entry.

I classify the relationship of an airport to an airline as either a dominated hub (H), a non-dominated hub (N) or a spoke (S), based on the airline's share of departures from the airport. Table 3 details the classification for the airlines and airports in the estimation sample. I estimate congestion parameters for each of these airline-airport categories using variation in marginal congestion at each flight's origin and destination. I define the marginal incremental mean delay that an airline imposes on itself at its own hubs under each alternative schedule as $\mu_j^H \equiv \mathbb{I}_{o \in \mathcal{H}_a} \mu_{odayft(j)}^{dep} + \mathbb{I}_{d \in \mathcal{H}_a} \mu_{odayft(j)}^{arr}$, where $\mathcal{H}_a$ is the set of airline a's dominated hubs, and where congestion under each alternative j is mapped with an origin and destination specific function $t(j)$ to the time at which the schedule impacts congestion at the origin or destination. I similarly define μ_j^N and μ_j^S for non-dominated hub and spoke airports, and the three measures for incremental variance to an airline's delays: V_j^H, V_j^N, and V_j^S.

I model an alternative-specific payoff to the airline from scheduling flight n (shorthand for the index $odayf$ for origin-destination-airline-day-flight) at the alternative time j as

$$\begin{aligned}
\pi_{nj} = & \sum_{k\in\{H,N,S\}} \theta_{\mu}^{k}\mu_{nj}^{k} \\
& + \sum_{k\in\{H,N,S\}} \theta_{V}^{k}V_{nj}^{k} \\
& + \sum_{c}\sum_{k=0}^{K_c} \beta_{k}^{c} z_{nj,k}^{c} \\
& + T_{nt(j)}^{dep} + T_{nt_{o\to d}(j)}^{arr} + \alpha_{j}^{prev} + \alpha_{J_f - j+1}^{next} + \varepsilon_{nj}
\end{aligned} \tag{5}$$

The two congestion measures (incremental mean and incremental variance) help identify four parameters when the origin and destination of a flight are in different hub categories. When the origin and destination are in the same category, the marginal congestion measures identify the parameter from that category. Table 4 provides a breakdown of the estimation sample by hub category: 82% of flights in the sample are between airports classified under different categories.

The connection model is included through the summation of $\beta_k^c z_{j,k}^c$ terms, where c stands for one of four types of connection (domestic or international, mainline or codeshare). Each parameter β_k^c refers to a value to the airline's payoff from a single direct connection of type c with a layover of k quarter-hours in excess of an airport's minimum connection time, and is estimated from $z_{j,k}^c$, the number of directness-scaled connections of type c that the airline connects to with layover k if it schedules the flight in period j. I estimate $K_c = 16$ discounted connection parameters for domestic flights and for domestic codeshares, and $K_c = 20$ parameters for international flights and codeshares, which captures the time component of enabling connections in a flexible and unrestricted way.[33]

I include time-of-day effects for both arrival and departure, which absorb passenger preferences for arrivals and departures at particular times of day, as well as any other unmodeled determinants of airline scheduling that vary over the day but are constant over airlines and airports.

I model these time-of-day effects using 16-degree polynomials for both arrivals and

[33] By cutting off the number of parameters at K_c, the estimates of connection discount parameters are relative to an omitted category that assigns no scheduling payoff from connections with a layover that exceeds K_c (which including minimum connection times is about 5 hours for domestic connections, 6.5 hours for international connections). This does not assume that airline does not derive a payoff from connections with such long layovers: rather, the assumption is that beyond, e.g. 6.5 hours for an international connection, the scheduled time of the connecting flight does not enter the scheduling decision for flight $odayf$

departures for each of the 4 time-zones in my estimation sample. In addition, I allow for "on the hour" effects for both departures and arrivals, which are indicators for schedules in the 16^{th} to 30^{th}, 31^{st} to 45^{th} and 46^{th} to 60^{th} minute of each hour, common across time zones. The terms $T^{dep}_{t(j)}$ and $T^{arr}_{t(j)}$ therefore stand for 99 parameters each (96 polynomial terms, 3 quarter-hour effects), which absorb variation in scheduling payoffs that is common across airlines and airports in a time zone.

The terms α^{prev}_j and $\alpha^{next}_{J_f-j+1}$ are indicators for the distance of schedule alternative j from the nearest flights preceding and following it on the airline's same origin-destination segment. These indicators do not map to any time-specific attributes of the choice environment, but instead capture an airline's preference for spreading flights out over the day to provide differentiated travel time choices to passengers. I include 16 indicators terms each for distance to the predecessor and successor flight.

The term ε_j makes scheduling payoff random from the point of view of the econometrician. I will make precise assumptions on the error term below, but first consider how unobservables included in this error may pose a challenge to interpreting estimates of the parameters in eq. (5) as airline preferences over trade-offs such as creating connections, avoiding the cost of congestion, and obtaining a strategic benefit from congestion.

Congestion measures are partly driven by airline's aggregate schedules. When airlines schedule more flights within a period than the airport's runways can handle, the congestion model will predict expected delay. When scheduled flights exceed capacity for several periods, as can be the case during banks, delayed flights early in the bank will have an external effect on the flights that are scheduled to follow. A time that is favorable for the departure of one flight may be favorable for flights departing to other destinations, leading an airline to coordinate a bank of flights around that time. An unobservable that directly enters into the error term for one flight's alternative is therefore likely to enter congestion measures through the schedules of other flights, biasing estimates of congestion parameters to be positive even if airlines do not schedule to take advantage of the strategic benefits from congestion.

Time of day effects and the connection model are included to control in part for such unobservables. Time of day effects that are common across airlines and flights, such as passenger preferences for certain times of the day (and ruling out departures and arrivals in the middle of the night) are likely to lead to banks of flights, and through banks to self-imposed delays. Banks are also coordinated to make connections: a period with many arrivals will be suitable for an airline to schedule departures to many destinations. To the extent that such a determinant of scheduling is not observed and is shared across outbound departures, it will be correlated both with departure congestion and the error term. I use the model of connections, which generates variation across airlines in the value of specific times, to observe this common component jointly entering different departures of the same airline.

Congestion measures may still be endogenous even after controlling for a common unobservable with time effects and for connections. An additional unobserved component that varies within airport and enters the schedule alternatives for multiple flights would enter the delay measures generated by the congestion model as well as the individual schedules. Examples include airport-specific variation in passenger preferences over time of day, as well as measurement error in the connection model. Additionally, coordination of arrival and departure banks means that just as flights departing from an airport are scheduled to connect to inbound arrivals, the arrivals were also scheduled to connect to the outbound flights. This makes the connection measures endogenous in any particular schedule choice even after controlling for common components of scheduling, as schedules in one direction (into or out of) the airport depend on schedule choices in the opposite direction.

To deal with the endogeneity in the congestion measures I propose instrument that are specific to each of the determinants of runway congestion: capacity and scheduled demand, as well as the intersection of the two. I also propose instruments for domestic connections.

7.1 Weather instruments

A first set of instruments is based on the weather-driven capacity measures that are used as inputs to the queuing model. These are predictive of delay metrics by construction as well as plausibly exogenous. To capture the history of within-day variation in empirical capacity paths using a reduced set of instruments (recall the congestion model is based on 155 empirical capacity paths per sample date), I take the empirical cumulative distribution function of observed capacity at each hour of the day and invert it to obtain a set of percentiles of each airport's capacity: percentiles 1 through 15, followed by every other 5 percentiles through the 60th percentile of capacity.[34] To obtain measures of capacity shortfall that are relevant to each airport and airline, I construct two sets of

[34]Lower percentiles are over-represented since low capacity is the less common state for all airports and there is more intra-day variation at lower percentiles, whereas from the 65th percentile onwards all airports have optimal capacity at all times.

weather instruments:

$$w_q^{oyt} = \frac{Departures_{oy}}{C_{oyt,q}}$$

and

$$w_q^{oayt} = \frac{Departures_{oay}}{C_{oyt,q}} \quad \text{for } q = \{1, 2, \ldots, 15, 20, \ldots, 60\}$$

where $C_{oyt,q}$ is the q^{th} percentile of capacity at airport o on day y at time t and I scale either total departures on the date ($Departures_{oy}$), or the total number of departures by airline a ($Departures_{oay}$) by the capacity percentile $C_{oyt,q}$ to obtain two instruments of capacity shortfall, the former being relatively more predictive of total delay to an average flight and the latter more predictive of delay to the flights of a particular airline. Both departure aggregates are constant over an airport and day, so all within-day variation in the instrument is due to weather.

It is important to underscore that since both the schedules and the weather used to compute the congestion measures are ex ante the date of the actual flight, and not actual weather realized on the flight date, these instruments are not based on weather "surprises". Rather, they exploit the very plausible exogeneity between variation in weather patterns within the day (which predictably affects capacity, and therefore the probability of delay) and unobserved determinants of scheduling, after including the congestion model measures of marginal delay. Data from 31 airports and 4 dates in different seasons provides variation in intra-day weather patterns. If demand for predictably delayed flights is reduced because an airport's home passengers learn the times that are likeliest to result in weather delay, this would be part of the cost to an airline from scheduling flights at such times, and is part of the effect of delay on airline scheduling that we seek to estimate.

7.2 "Push-pull" instruments

Additional instruments of congestion measures can be obtained from any shifters of aggregate schedule demand on the runway that are excluded from the error term in schedule choice equation (5). For a particular origin airport and a particular flight departure, a shifter of the schedules of other flights departing from the same runway will be given by any time-varying effect that enters the scheduling decisions of those flights at their destinations. We can then use known flight times to map back the intra-day variation in arrival effects at each destination to the time at which they would shift demand on the departure runway. I define below a shifter based on this idea that I will refer to as a "pull" instrument, because it will predict take-off times based on variation at each flight's destination. The analogue for landing times I refer to as a "push" instrument.

The measure of arrival effects at a destination airport that I use for the pull instrument is the time-profile of an airline's arrivals at the airport, net of any arrivals from the origin airport for which the instrument is constructed.

For departures from an airport o to a destination d by airline a on day y, let o' be an index for all flight origin airports excluding o, and likewise for d'. I define a single "pull" on the runway at o coming from airport d' at time t as

$$Pull_{oay}^{d'}\left(t\right) \equiv \frac{F_{od'a}}{\sum\limits_{a,d'} F_{od'a}} Count_{od'ayt} \tag{6}$$

where F_{oda} is airline a's frequency of service from o to d, and

$$Count_{od'ayt} \equiv \sum_{o'} \mathbb{I}\left[\tau_{o\rightarrow d'}\left(t\right) \in \tau_{o'\rightarrow d'}\left(h^{*}_{o'd'ay}\right)\right] \tag{7}$$

is the count of arrivals at airport d' from all origins $o' \neq o$ at time $\tau_{o\rightarrow d'}\left(t\right)$, where $\tau_{o\rightarrow d'}\left(t\right)$ is the time at which a departure at time t from o would arrive at d'.[35] Note that $Pull_{oay}^{d'}\left(t\right)$ is defined only in terms of an airline's own arrivals. Under the null hypothesis that airlines do not impose congestion strategically, we could additionally predict demand on the runway from the pull effects on competitor departures. If airlines seek to congest times that are relatively attractive to competitors, however, a pull effect that predicts competitor demand on the runway would not be excluded from the error term in schedule choice equation (5).

I define the pull instrument on the aggregate schedules of airline a by summing $Pull_{oay}^{d'}\left(t\right)$ from every airport d', and rescaling in terms of all such pulls effects from the airport (from other airlines, and at other times):

$$w_{oday}^{pull}\left(t\right) \equiv \frac{\sum\limits_{d'} Pull_{oay}^{d'}\left(t\right)}{\sum\limits_{a,t}\sum\limits_{d'} Pull_{oay}^{d'}\left(t\right)} \tag{8}$$

Figure 11 provides an example of how eq. (8) is constructed for the shifter of aggregate departures out of Newark for a Continental flight bound for Atlanta. In the example data, which is from June 13, 2007, Continental had 7 scheduled arrivals into Albuquerque from airports other than Newark. The time for each of these arrivals

[35]In the preceding definition, $h^{*}_{o'd'ay}$ is the set of flights scheduled from o' to d' by airline a on day y, and $\tau_{b\rightarrow c}$ is a time-shift function that maps a scheduled departure time at b to a scheduled arrival time at c

enters the count measure at a time that is appropriately mapped back to departure (in this case, the minimum scheduled time it takes Continental to fly from Newark to Albuquerque), weighted by the number of Continental flights to Albuquerque from Newark (of which there is 1 a day) divided by all 464 domestic departures from Newark. The same procedure is repeated for Albany (4 arrivals not from Newark, and a frequency of 4 flights from Newark to Albany) and so on through all destinations that Continental serves from Newark, excluding Atlanta. The time profile of "pull" effects obtained from each of the destinations of an airline is summed over all such destinations, and re-weighted in terms of the sum of these time profiles over all airlines and times of day.

I define the equivalent "push" instrument for arrival demand on a runway as

$$\begin{aligned} Count_{o'day\tau} &\equiv \sum_{d'} \mathbb{I}\left[\tau^{-1}_{o'\to d'}\left(\tau_{o\to d'}(t)\right) \in h^{*}_{o'd'ay}\right] \\ Push^{o'}_{day}(t) &\equiv \frac{F_{o'da}}{\sum_{a,o'} F_{o'da}} Count_{o'dayt} \\ w^{push}_{oday}(t) &\equiv \frac{\sum_{o'} Push^{o'}_{day}(t)}{\sum_{a,t}\sum_{o'} Push^{o'}_{day}(t)} \end{aligned}$$

These instrument exploit the features of the airline's network structure that are fixed over the day (what airports the airline serves from a particular destination, and the service to those airports from other locations) to weight arrival times at other locations, for example at every other airport that Continental flies to from Newark. By counting scheduled arrival times at these airports, the instrument exploits any common arrival effects that enter both the schedule decisions of the counted arrivals and the Continental departures from Newark that are bound for the same airport. Pull effects can be thought of as arising partly from airport-specific periods of peak scheduling and partly by the regular operating hours of an airport: given the time-zone the airport is located in, and its distance to the departure airport we are constructing a "pull" instrument for, each potential destination airport will have an arrival demand that turns on and off with the start and end of the day. For example, the "pull" on West Coast airports from the East Coast shuts off in the early evening, as departures would arrive in the East Coast in the middle of the night when no other arrivals are scheduled, but turns back on around midnight to predict "red-eye" departures, as flight scheduled at this time arrive before the start of working day, along with other short-haul flights arriving from airports in the East Coast.

Flights out of o are removed from the count in eq. (6) because they are part of the aggregate departures from the airport that enter the delay measures and that the pull from arrival times is designed to instrument for. Excluding these flights removes origin-specific unobservables that concurrently enter schedule choices for departure to different destinations. For example, if a train arrives at Newark from New York City's Penn Station at 9 a.m., this could lead an airline to coordinate several departures at 10 a.m.. Excluding all flights out of Newark from the pull instruments for Newark removes any such common determinants of departure times that could violate the instrument's exclusion restriction.

Flights to d are removed from eq. (8) for two reasons. On a practical level, these are not departures we need to predict: departures from o to d as actually chosen by the airline are removed from aggregate schedules to calculate marginal congestion effects. More importantly, arrival times into d will be partly determined by any ot- and dt-specific unobservables that are not absorbed by time-zone-time-of-day effects, and inclusion in the instrument would directly violate the exclusion restriction.

7.3 Directed "push-pull" instruments

Connection terms can be endogenous as airlines jointly determine schedules that are inbound and outbound to an airport to enable connections. I exploit the idea of pull and push instruments to create instruments for a flight's domestic connection terms. To instrument the inbound connections of a flight departing from airport o and bound for d, I use excluded variation in the time profile of departures at the airports from which the airline has scheduled flights to o, mapping the departure times to arrival times as with the push instruments for runway arrival demand. Excluded in this case means dropping from every airport time-profile any departures to o. I weight the time-profile of "push" effects from each such airport for inbound connections to o by the directness angle between the airport, the connection airport o and the destination d. Connection angles therefore not only enter the measure of inbound connections, generating variation in connection terms across destinations, but also enter the instrument for these connection terms.

7.4 Estimation

Even if relevant and plausibly excluded instruments are available, dealing with endogeneity in a discrete choice setting is not straightforward. I use the control function approach (Villas-Boas and Winer, 1999, Train, 2009), which in my application with a large number of alternatives requires parametric assumptions on the error term. Rewrite eq.

(5) as

$$\pi_{nj} = V\left(\mu_{nj}, V_{nj}, z_{nj}, \gamma\right) + \varepsilon_{nj} \tag{9}$$

where γ stands for all the parameters in that equation. Write each endogenous explanatory variable as a function of instruments and an error term:

$$\begin{aligned}
\mu_{nj} &= W_\mu\left(w^\mu_{nj}, \delta^\mu\right) + \eta^\mu_{nj} \\
V_{nj} &= W_V\left(w^V_{nj}, \delta^V\right) + \eta^V_{nj} \\
z_{nj} &= W_z\left(w^z_{nj}, \delta^\mu\right) + \eta^z_{nj}
\end{aligned} \tag{10}$$

and there are a total of $i = 1, \ldots, N$ such first stage equations, one for each endogenous variable. Under standard conditional mean independence assumptions for the instruments, ε_{nj} and all η_{nj} are independent of w_{nj}, but correlated to each other, making μ_{nj}, V_{nj} and z_{nj} endogenous. I now make the parametric assumption that

$$\varepsilon_{nj} = \varepsilon^1_{nj} + \varepsilon^2_{nj} + \ldots + \varepsilon^N_{nj} + \varepsilon^{N+1}_{nj} \tag{11}$$

where ε^i_{nj} is jointly normal with η^i_{nj}, and ε^{N+1}_{nj} is an independent and identically distributed extreme value error. Consider the decomposition of ε^i_{nj} relative to its mean conditional on η^i_{nj} and the deviation from this mean:

$$\varepsilon^i_{nj} \equiv \mathbb{E}\left[\varepsilon^i_{nj} \mid \eta^i_{nj}\right] + \tilde{\varepsilon}^i_{nj}$$

Under the assumption of joint normality, we can write this as

$$\varepsilon^i_{nj} = \lambda^i \eta^i_{nj} + \tilde{\varepsilon}^i_{nj} \tag{12}$$

where λ^i is the population correlation coefficient between ε^i_{nj} and η^i_{nj}. We can substitute eq. (12) into the choice equation, which allows $\lambda^i \eta^i_{nj}$ to serve as a control function for the component of the error term ε^i_{nj} that induced endogeneity in explanatory variable i. Conditional on $\lambda^i \eta^i_{nj}$, the residual error term is no longer endogenous, allowing us to estimate the choice equation. After substituting for $\varepsilon^i_{nj} = 1, \ldots, N$ endogenous variables, the choice equation becomes

$$\pi_{nj} = V\left(\mu_{nj}, V_{nj}, z_{nj}, \gamma\right) + \sum_{i=1}^{N} \lambda^i \eta^i_{nj} + \sum_{i=1}^{N} \tilde{\varepsilon}^i_{nj} + \varepsilon^{N+1}_{nj} \tag{13}$$

I assume the $\tilde{\varepsilon}^i_{nj}$ errors have a common standard deviation, so the sum of residual terms $\sum_{i=1}^{N} \tilde{\varepsilon}^i_{nj}$ is a normal random variable with unknown variance, which I define as $\tilde{\varepsilon}_n \sim N\left(0, \sigma^2_{\tilde{\varepsilon}}\right)$. I carry out estimation with a limited information approach: first, I estimate first stages of the included endogenous variables on the instruments, and obtain the residuals from these regressions. These provide estimates of η^i_{nj} errors, which I include in the second stage as additional explanatory variables. I then estimate the choice model as a mixed conditional logit, as the error term is now the sum of $\varepsilon^{N+1}_{nj} \sim EV$ and $\tilde{\varepsilon}_n \sim N\left(0, \sigma^2_{\tilde{\varepsilon}}\right)$. Given the large alternative-space, I use sparse-grid integration following Heiss and Winschel (2008) to integrate out the multivariate normal error component.

8 Results

Table 8 presents the results from estimating the schedule choice model.[36] Column 1 presents estimates from a standard fixed-effects conditional logit, which does not include control functions to address endogeneity. Estimated marginal effects on the congestion terms are positive, consistent with an unobservable that is correlated both with the congestion measures and schedule choices. Specifications in Columns 2 and 3 are estimated using the weather and push-pull instruments for congestion terms, and directed-push pull instruments for domestic flight connections, separately including congestion terms for the mean or variance of delay, and Columns 4 and 5 include both jointly.

The marginal delay caused by scheduling an arrival or departure at a spoke airport is purely an operational cost to an airline, so absent a strategic motive at spokes (where airlines schedule few flights) we expect its effect on airline scheduling to be negative. In specifications (2) and (3) I find large effects of airline congestion avoidance at spoke airports, after instrumenting the congestion measures: airlines will reduce their probability of scheduling at a given time by an (imprecisely estimated) 1.17 percentage points if it increases the expected delay to one of their flights by 15 minutes, relative to an average baseline probability of scheduling in a given 15-minute period of 3.5% in the sample. The effect of avoiding variance to flights is smaller but more precisely estimated: airlines are 0.41 percentage points less likely to schedule a flight at a time

[36] The results are preliminary. Column 5 presents the estimation of the model with the error structure displayed in eq. (13), but does not present bootstrapped confidence intervals. Columns 2-4 do not integrate out a multivariate mean-zero error component, but present bootstrapped standard errors to account for generated residual terms from first stages. The point estimates presented are average marginal structural effects in percentage points, obtained by bootstrapping the model under its assumed error structure, and two sets of bootstrapped confidence intervals are presented, for 95% and 90% confidence.

when it increases the variance of delay to one of the airline's flights by 15 minutes. These results are consistent with the observation that spoke airlines tend to schedule their flights in the "valleys" between the banks of hub airlines.[37] Specification (4) jointly includes congestion measures for incremental mean delay and incremental variance of delay. Results for scheduling at spokes are similar although attenuated for each measure. Avoidance of incremental variance of delay is no longer statistically significant, and point estimates are substantially lower for incremental mean delay.

The main results in this paper are the estimates of the effect that causing incremental congestion has on an airline's decision to schedule at its own hubs. Instrumenting the congestion terms ensures that the marginal effects estimated and reported in Columns 2-4 come only from exogenous variation that shifts the consequences of scheduling in terms of delay, and not from an unobservable that is correlated with congestion measures and schedule choice (such a time-specific shock to the payoff of scheduling that is common across all of an airline's flights at an airport).

I find that airlines will schedule a flight to contribute to congestion at their own hubs, which is consistent with the strategic benefit (increasing the congestion costs to competitors to deter their entry into attractive times) outweighing the airline's own cost from the congestion. An alternative interpretation of this finding, which does not assign an adversarial intent to airlines, is as follows: conditional on determinants of schedule choice such as creating connections and time-of-day effects that make some times more attractive than others, hub airlines will schedule marginal flights at certain times, at the expense of self-imposed congestion on their other flights, because they expect the benefit from removing those flights to less congested times to be counteracted by competitor entry. For such an airline, entry offsets the reduction in congestion and potentially increases competition for passenger demand. I find strong effects for incremental mean delay on schedule choice at both dominated and non-dominated hubs: in the specification with both mean and variance terms, airlines are 1.25 percentage points likelier to schedule a flight if it leads to an additional 15 minutes of congestion to one of their own flights at a dominated hub, and 1.41 percentage points likelier at non-dominated hubs. The results are slightly lower in the specification in Column 2, but the effect remains stronger for non-dominant hubs.[38] The effect of incremental variance on schedule choice are relatively smaller and significant only in the specification in Column 3, but are consistent with the results for incremental mean delay.

[37] See Figures 2 and 3 for examples of how hub and spoke airlines scheduled their morning flights at Denver and Newark.

[38] Although the difference is not-statistically significant in either specification, this result is consistent with non-monotonicity effects in entry-deterrence, see Ellison and Ellison (2011).

Column 5 presents point estimates for the full estimation strategy described in Section 7.4, as applied to the specification in Column 4. Point estimates of the average marginal structural effect are similar to those in Column 4 for connection terms (discussed below) and incremental variance of delay, and somewhat larger for incremental mean delay.

Estimates of the marginal effect of an added potential connection are closely estimated across specifications. All else equal, a time that has one additional immediate connection is 0.87 percentage points likelier to be chosen by an airline (columns 2 and 4), whereas an added connection with an extra 1 hour of layover (about 2 hours of layover as experienced by the passenger) increases the probability of scheduling by 0.32 percentage points. These results indicate that an airline's tradeoff between making a domestic 2 hr connection and delaying a single one of its flights by an expected 15 minutes are of approximately the same size.[39]

Estimates from the remaining terms in the connection model are presented in Figure 13, which shows all connection model coefficients normalized in terms of domestic connections with the minimum connection time, i.e. the coefficient β_0 for an airline's connections to its own flights. Domestic connection terms are estimated including control function terms from first stage regressions with directed push-pull instruments, whereas the remaining three categories are included without control functions. The value of domestic connections to an airline are discounted rapidly during the first hour of excess layover (since minimum connection times are usually 45 to 75 minutes, this decay occurs during what would be the second hour of layover as experienced by passengers), but domestic connections appear to retain value to the airline through the 3rd hour of excess layover, which approximately corresponds to the rule that sets a maximum of 4 hours for domestic connections on ticket reservation systems.

The estimates of discount terms for international flights suggest that airlines may build in layover in excess of minimum connecting time to allow for passengers to be delayed in border crossing: minimum connection time coefficients start low but climb quickly during the first hour of additional layover. Unlike domestic connections, which decay in value completely by the fourth hour of layover, international connections are discounted more slowly and retain value to the airline beyond the 4 hours of

[39]The reported marginal effect calculation assumes the marginal connection is only available in one period for immediate connection, and not thereafter. Actual flights that are not not connected to immediately are available for connection with a greater layover. Connections with a longer layover have a discounted value to the airline, which increases the payoffs from scheduling under non-immediate alternatives. The reported effect therefore overstates the effect of adding a connection would have on schedule choice, which is properly accounted for in estimation.

domestic connections. A single connection to an international flight with a layover of as much as 6 hours is valued by airlines about as much as 2 domestic connections with a 2 hour layover. Creating potential connections to domestic codeshare partners are estimated with high precision to play almost no role in scheduling, whereas codeshares to international partners have a long lasting but imprecisely estimated effect on the domestic airline.

Unobserved variation in the payoff from scheduling at certain times that is common across airlines and airports is captured by estimated time-of-day effects. Figure 12 plots the point estimates of the polynomials for arrival and departure effects for two selected time zones.[40] In both time zones there is a strong arrival effect before the start and a departure effect at the end of the work day (all days in the estimation sample are Wednesdays). Morning arrival effects in the Eastern time zone start earlier than in the west, as most arrivals in the east are from other eastern airports, whereas a large number of the morning arrivals in the west are also from the east. Note that there are no large departure effects in the Pacific time zone to account for red-eye flights. The model does not need such effects, which would be counter to the fact for West Coast departures to other West Coast destinations. Red-eyes are explained in the model by the large early-morning arrival effects in the Eastern time zone. To the extent that time of day effects are driven by passenger preferences and demand, as the estimated patterns around the workday suggest, these help account for the fact that spoke airlines do schedule some flights during periods of high runway demand, in which congestion costs are largest, and that hub airlines have incentives to schedule blocks of flights that deter the marginal entrant from those times.

Table 7 presents linear specifications that approximate the schedule choice model. An observation in every specification of table is a 15-minute period on an origin-destination segment served by an airline. The dependent variable is whether the airline schedules a flight in those 15 minutes, and its average over the segment is determined by the airline's frequency of service. The effect of the frequency on the baseline probability of scheduling in any given 15 minutes is absorbed by the segment fixed-effect. The role of instruments in controlling for endogeneity is apparent from comparing Column 1, which reports the linear probability model estimated by OLS, and Columns 2, 3 and 4 which present 2SLS analogues of the control-function specifications in Table 8: mean congestion effects become smaller at dominated hubs and non-dominated hubs,

[40] Eastern standard time and pacific standard time, the results for other time-zones are similar. Coefficients for time-of-day effects are re-scaled as the conditional probability that an airline would select each time of the day to schedule a single flight.

and the magnitude of estimated congestion avoidance at spokes becomes economically large (a 15-minute increase in mean delay leads to a decrease of 1 percentage point in the probability of scheduling), although this effect is imprecisely estimated. Column 5 presents a robustness check on Column 4 using only the weather instruments: the results are qualitatively similar although estimated less precisely. The pattern of discount coefficients remains similar to that for the conditional choice model.

These linear specifications are not ideally suited to study schedule choice because they misspecify the interdependence between an airline's flights on the same segment, generally by not selecting from a set of relevant alternative and therefore failing to exploit information from the alternatives not chosen, and in particular by failing to track the proximity to other flights on the same segment in order to correctly attribute the layovers that arise from airline schedule choices. However, qualitative results on the congestion parameters are comparable and provide support for the instrumental variables strategy.

8.1 Evidence from response to competitor value for times

We can use push-pull metrics for a second test of strategic congestion externalities. Airlines may be more likely to self-impose congestion at their hubs at times that are more valuable to competitors. We can use the push-pull instrument for an airline's competitors as a measure of such times, which is driven by the structure of their hub-spoke networks and the destinations they serve from the airport. Table 9 presents preliminary coefficients from interacting the sum of push-pull instruments for all competitors with an airline's congestion measures. The strongest result from this table is that the interaction breaks up the positive incremental mean congestion effect reported in Table 8 into times at which the hub airline's competitors do not value access to the runway, and finds that the airline avoids congestion at its hub, and times that are valuable to competitors, when the airline schedules additional flights if they increase congestion. The result in the scheduling model specifications hold at dominant hubs but not non-dominant hubs. Similar results are reported for the linear specification in Columns 6-8 of Table 7, although in the linear specification the results hold only for non-dominated hubs. It is important to note that these results do not arise from variation in the unobserved value of specific times of day that is common to both the hub airline and its competitors, as such variation is absorbed by time-of-day effects.

9 Conclusions

This paper computes novel measures of congestion externalities at airports and develops an empirical approach to studying airline schedule choice that allows us to assess the effect of these externalities on airline choices. The focus on airline choices allows for a straightforward way to identify the tradeoffs airlines face in scheduling peaks of flights, and to test for the presence of strategic incentives. Such strategic incentives can compound the efficiency costs from multiple users congesting a finite resource.

Traditional approaches to addressing this issue in the airline industry include slot control and congestion pricing. There is very broad support for congestion pricing among economists: the statement "*in general, using more congestion charges in crowded transportation networks - such as higher tolls during peak travel times in cities, and peak fees for airplane takeoff and landing slots - and using the proceeds to lower other taxes would make citizens on average better off*" polled the highest degree of consensus among all questions in a poll of leading economists (IGM Forum, 2012), with 92% agreeing and none disagreeing. To date, discussion of congestion pricing has been hampered by the complex nature of stakeholder interests and multiple policy objectives, but not least of all because it may not be apparent that airport runways are run like a commons - and as this paper shows, one that is beset by strategic incentives.

References

Aguirregabiria, V., and C.-Y. Ho (2010): "A dynamic game of airline network competition: Hub-and-spoke networks and entry deterrence," *International Journal of Industrial Organization*, 28(4), 377–382.

Ater, I. (2012): "Internalization of congestion at US hub airports," *Journal of Urban Economics*, 72(2-3), 196–209.

Ball, M., C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, and B. Zou (2010): "Total Delay Impact Study," *The National Center of Excellence for Aviation Operations Research (NEXTOR)*.

Ball, M., C. Barnhart, G. Nemhauser, and A. Odoni (2007): "Air Transportation: Irregular Operations and Control," in *Handbooks in Operations Research and Management Science*, ed. by C. Barnhart, and G. Laporte, vol. 14, pp. 1–67. Elsevier.

Barnhart, C. (2009): "Airline Schedule Optimization," in *The Global Airline Industry*, ed. by P. Belobaba, A. Odoni, and C. Barnhart, pp. 183–212. John Wiley & Sons, Ltd.

Belobaba, P. (2009): "The airline planning process," in *The Global Airline Industry*, ed. by P. Belobaba, A. Odoni, and C. Barnhart, pp. 153–181. John Wiley & Sons, Ltd.

Berry, S., M. Carnall, and P. T. Spiller (2006): "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity," in *Advances in Airline Economics, Volume I, Competition and Antitrust*, ed. by D. Lee, pp. 141–162. Elsevier.

Berry, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, pp. 889–917.

Bonanno, G. (1987): "Location Choice, Product Proliferation and Entry Deterrence," *The Review of Economic Studies*, 54(1), 37–45.

Borenstein, S. (1989): "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry," *The RAND Journal of Economics*, 20(3), 344–365.

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

Borenstein, S., and J. Netz (1999): "Why do all the flights leave at 8 am?: Competition and departure-time differentiation in airline markets," *International Journal of Industrial Organization*, 17(5), 611–640.

Borenstein, S., and N. L. Rose (1995): "Competition and Price Dispersion in the U.S. Airline Industry," Working Paper 3785, National Bureau of Economic Research.

——— (2011): "How airline markets work... or do they? Regulatory reform in the airline industry," in *Economic Regulation and Its Reform: What Have We Learned?* University of Chicago Press.

Brueckner, J. K. (2002): "Airport congestion when carriers have market power," *American Economic Review*, pp. 1357–1375.

——— (2005): "Internalization of airport congestion: A network analysis," *International Journal of Industrial Organization*, 23(7), 599–614.

Brueckner, J. K., N. J. Dyer, and P. T. Spiller (1992): "Fare Determination in Airline Hub-and-Spoke Networks," *The RAND Journal of Economics*, 23(3), 309–333.

Brueckner, J. K., and P. T. Spiller (1994): "Economies of Traffic Density in the Deregulated Airline Industry," *Journal of Law and Economics*, 37(2), 379–415.

Brueckner, J. K., and K. Van Dender (2008): "Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate," *Journal of Urban Economics*, 64(2), 288–295.

Caves, D. W., L. R. Christensen, and M. W. Tretheway (1984): "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *The RAND Journal of Economics*, 15(4), 471–489.

Chen, Z., AND T. W. Ross (2000): “Strategic Alliances, Shared Facilities, and Entry Deterrence,” *The RAND Journal of Economics*, 31(2), 326–344.

Chevalier, J. A. (1995a): “Capital Structure and Product-Market Competition: Empirical Evidence from the Supermarket Industry,” *The American Economic Review*, 85(3), 415–435.

——— (1995b): “Do LBO Supermarkets Charge More? An Empirical Analysis of the Effects of LBOs on Supermarket Pricing,” *The Journal of Finance*, 50(4), 1095–1112.

Ciliberto, F., AND E. Tamer (2009): “Market structure and multiple equilibria in airline markets,” *Econometrica*, 77(6), 1791–1828.

Ciliberto, F., AND J. W. Williams (2010): “Limited Access to Airport Facilities and Market Power in the Airline Industry,” *Journal of Law and Economics*, 53(3), 467–495.

Dafny, L. S. (2005): “Games hospitals play: Entry deterrence in hospital procedure markets,” *Journal of Economics & Management Strategy*, 14(3), 513–542.

Daniel, J. I. (1995): “Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues,” *Econometrica*, pp. 327–370.

Daniel, J. I., AND K. T. Harback (2008): “(When) Do hub airlines internalize their self-imposed congestion delays?,” *Journal of Urban Economics*, 63(2), 583–612.

De Neufville, R., AND A. R. Odoni (2003): *Aiport Systems: planning, design, and management*. McGraw-Hill, London.

Ellison, G., AND S. F. Ellison (2011): “Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration,” *American Economic Journal: Microeconomics*, 3(1), 1–36.

Evans, W. N., AND I. N. Kessides (1994): “Living by the "Golden Rule": Multimarket Contact in the U.S. Airline Industry,” *The Quarterly Journal of Economics*, 109(2), 341–66.

Forbes, S. J. (2008): “The effect of air traffic delays on airline prices,” *International Journal of Industrial Organization*, 26(5), 1218–1232.

Goolsbee, A., AND C. Syverson (2008): “How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines,” *The Quarterly Journal of Economics*, 123(4), 1611–1633.

Gopalan, R., AND K. Talluri (1998): “Mathematical models in airline schedule planning: A survey,” *Annals of Operations Research*, 76, 155–185.

Hansen, M. M., G. D. Gosling, J.-D. Margulici, AND W.-B. Wei (2001): “Influence of Capacity Constraints on Airline Fleet Mix,” *NEXTOR Research Report UCB-ITS-RR-2001-6*.

Heiss, F., AND V. Winschel (2008): “Likelihood approximation by numerical integration on sparse grids,” *Journal of Econometrics*, 144(1), 62–80.

Hendricks, K., M. Piccione, AND G. Tan (1997): “Entry and Exit in Hub-Spoke Networks,” *The RAND Journal of Economics*, 28(2), 291–303.

Jacobs, T. L., L. A. Garrow, M. Lohatepanont, F. S. Koppelman, G. M. Coldren, AND H. Purnomo (2012): “Airline Planning and Schedule Development,” in *Quantitative Problem Solving Methods in the Airline Industry*, ed. by C. Barnhart, and B. Smith, vol. 169, pp. 35–99. Springer US, Boston, MA.

Judd, K. L. (1985): “Credible Spatial Preemption,” *The RAND Journal of Economics*, 16(2), 153–166.

Kadiyali, V. (1996): “Entry, Its Deterrence, and Its Accommodation: A Study of the U. S. Photographic Film Industry,” *The RAND Journal of Economics*, 27(3), 452–478.

Lan, S., J.-P. Clarke, AND C. Barnhart (2006): “Planning for Robust Airline Operations: Optimizing Aircraft Routings and Flight Departure Times to Minimize Passenger Disruptions,” *Transportation Science*, 40(1), 15–28.

Lederman, M. (2007): "Do Enhancements to Loyalty Programs Affect Demand? The Impact of International Frequent Flyer Partnerships on Domestic Airline Demand," *The RAND Journal of Economics*, 38(4), 1134–1158.

——— (2008): "Are Frequent-Flyer Programs a Cause of the "Hub Premium"?," *Journal of Economics & Management Strategy*, 17(1), 35–66.

Lieberman, M. B. (1987): "Postentry Investment and Market Structure in the Chemical Processing Industries," *The RAND Journal of Economics*, 18(4), 533–549.

Mayer, C., AND T. Sinai (2003): "Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil," *American Economic Review*, 93(4), 1194–1215.

Odoni, A. R. (2009): "The International Institutional and Regulatory Environment," in *The Global Airline Industry*, ed. by P. Belobaba, A. Odoni, and C. Barnhart, pp. 19–46. John Wiley & Sons, Ltd.

Prince, J. T., AND D. H. Simon (2012): "Do Incumbents Improve Service Quality in Response to Entry? Evidence from Airlines' On-Time Performance," Working Paper 2010-15, Indiana University, Kelley School of Business, Department of Business Economics and Public Policy.

Rupp, N. G. (2009): "Do carriers internalize congestion costs? Empirical evidence on the internalization question," *Journal of Urban Economics*, 65(1), 24–37.

Salop, S. C., AND D. T. Scheffman (1983): "Raising Rivals' Costs," *The American Economic Review*, 73(2), 267–271.

——— (1987): "Cost-Raising Strategies," *The Journal of Industrial Economics*, 36(1), 19–34.

Schmalensee, R. (1978): "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry," *The Bell Journal of Economics*, 9(2), 305–327.

Scott Morton, F. M. (2000): "Barriers to entry, brand advertising, and generic entry in the US pharmaceutical industry," *International Journal of Industrial Organization*, 18(7), 1085–1104.

Snider, C. (2009): "Predatory incentives and predation policy: the American Airlines case," *Manuscript. Department of Economics. UCLA.*

Snider, C., AND J. Williams (2011): "Barriers to Entry in the Airline Industry: A Regression Discontinuity Approach," *Working paper*.

Train, K. (2009): *Discrete choice methods with simulation*. Cambridge University Press, 2nd edn.

Vickrey, W. S. (1969): "Congestion Theory and Transport Investment," *The American Economic Review*, 59(2), 251–260.

Villas-Boas, J. M., AND R. S. Winer (1999): "Endogeneity in Brand Choice Models," *Management Science*, 45(10), 1324–1338.

Tables and figures

Table 1: Notation and definitions

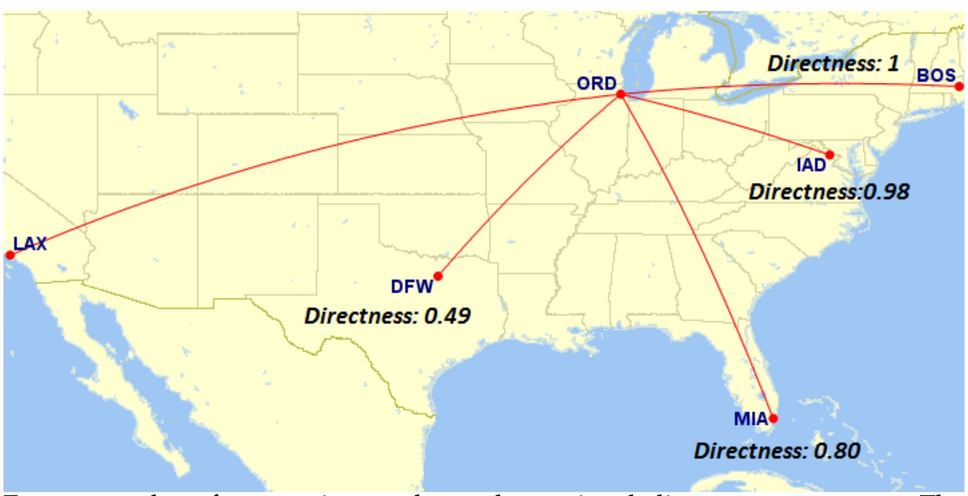
Variable/symbol	belongs to	Description
t	$\{1, ..., T\}$	Period or subinterval of the day.
h	$\{0, 1\}^T$	An airline schedule. Indexed by $oday$.
o	$\{ACA, ..., ZSA\}$	Origin airport
d	$\{ACA, ..., ZSA\}$	Destination airport
a	$\{AA, ..., WN\}$	Airline
y		Day
f	$\{1, ..., 32\}$	Flight number within a day's frequency. Indexed by $oday$.
F	$\{1, ..., 32\}$	Frequency of the schedule. Indexed by $oday$.
μ_{odayft}	$\mathbb{R}_+$	Incremental mean delay to airline a from the marginal flight $odayft$.
V_{odayft}	$\mathbb{R}$	Incremental variance of delay to airline a from the marginal flight $odayft$

Table 2: Airline fares and directness of connection.
DB1B sample of airline tickets for 2007.

Dependent variable: log fare	(1)	(2)	(3)
Log distance flown	-.110		
Directness		.135	
Directness in (0.98,1]			.087
Directness in (0.80,0.98]			.079
Directness in (0.50,0.80]			.058
O-D market-quarter effects	Yes	Yes	Yes
Observations	7.4 M	7.4 M	7.4 M
R^2	0.29	0.29	0.30

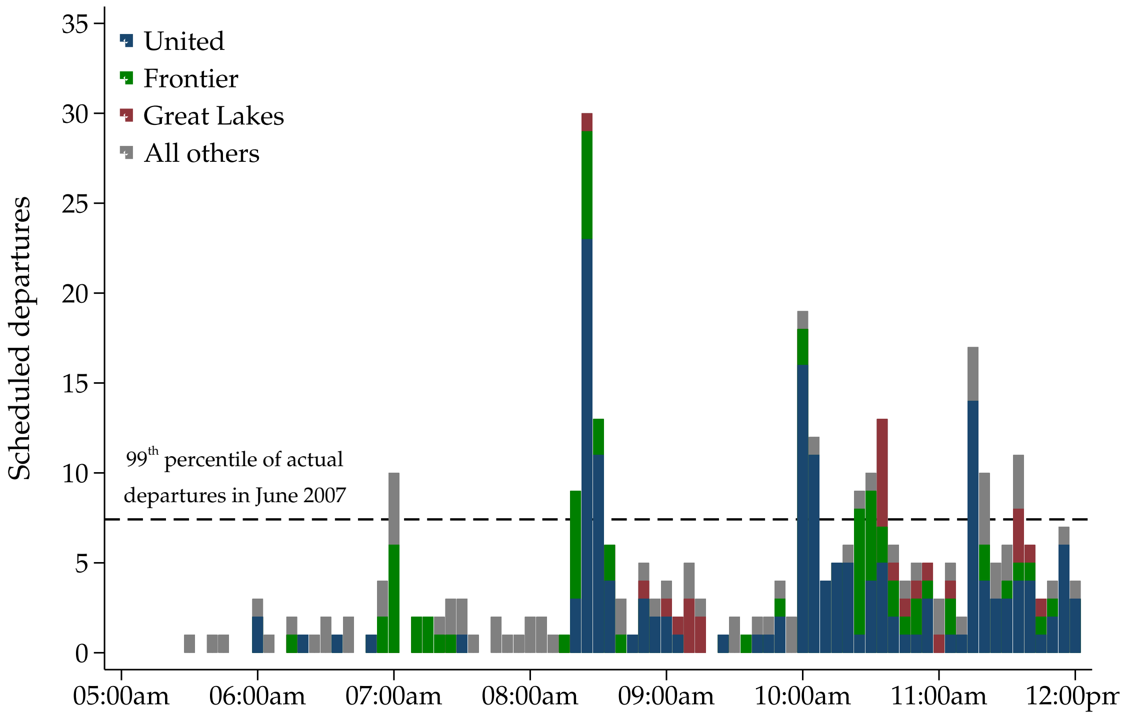
Sample consists of all two-segment tickets in the DB1B 10% sample of US domestic airline tickets for 2007, dropping fares of $25 or less and city-pair markets with less than 10 sampled passengers per quarter. An observation is a single passenger fare. All coefficients significant at the 1% level with standard errors clustered at the origin-destination market-quarter level. All specifications include origin-destination market-quarter effects. Omitted category in Column 3 are two-segment flights with directness in [.33,0.50], where .33 is the lowest directness measure observed in the sample.

Figure 1: Four examples of directness measures.
Starting at Los Angeles and connecting through Chicago.



Four examples of connection angles and associated directness measures. The airports LAX, ORD and BOS are on the same great circle, for a directness of 1. Distance for LAX-IAD is 0.981 of the LAX-ORD-IAD distance. Distance for LAX-MIA is 0.796 of the LAX-ORD-MIA distance. Distance for LAX-DFW is 0.485 of the LAX-ORD-DFW distance. Plotted with Great Circle Mapper, available at gcmap.com

Figure 2: Scheduled departures at Denver International Airport (DEN) on the morning of June 13, 2007



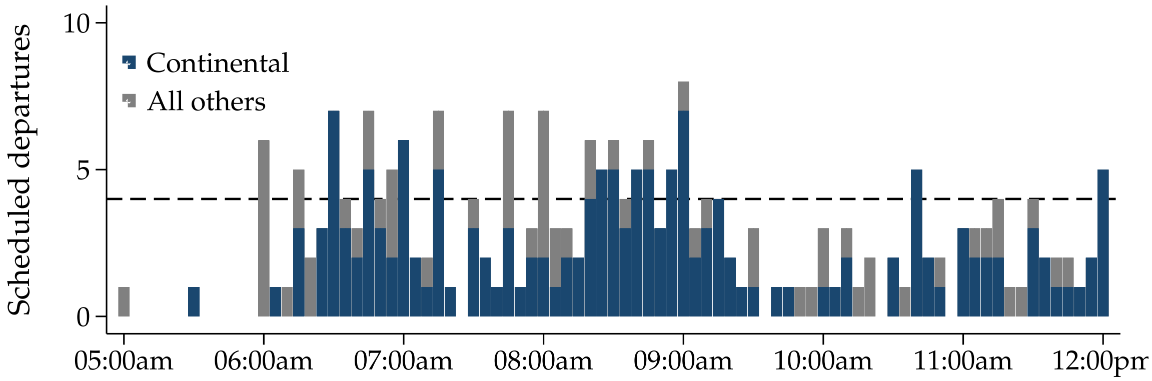
Scheduled commercial departures include domestic and international departures by all carriers, binned into 5 minute intervals. There were also 8 departures by freight and general aviation, not plotted. United Airlines, Frontier Airlines and Great Lakes Airlines are all hub operators at Denver. Great Lakes is a regional airline specializing in service to small markets in the Midwest and Mountain west, often subsidized by the Essential Air Service program. The black dashed line is the 99th percentile of *hourly* departures from Denver from 5 a.m. to 12 p.m. in June 2007 (89 departures), divided by 12 (from FAA Enhanced Traffic Management System data).

Table 3: Classification of airline-airport pairs

Airport	Dominated hub	Non-dominated hub
Atlanta	Delta	AirTran
Boston		American, Jet Blue, Delta, US Airways
Baltimore		AirTran, US Airways, Southwest
Cleveland	Continental	
Charlotte	US Airways	
Cincinnati	Delta	
Denver		Frontier, United
Dallas-Fort Worth	American	
Detroit	Northwest	
New York City Newark	Continental	
Fort Lauderdale		Jet Blue, Continental, Delta, Spirit, US Airways, Southwest
Washington Dulles	United	
Houston Bush	Continental	
New York City JFK		American, Jet Blue, Delta
Las Vegas		US Airways, Southwest
Los Angeles		American, United, Southwest
Orlando		Delta, AirTran, Southwest
Chicago Midway	Southwest	AirTran
Memphis	Northwest	
Miami	American	Continental
Minneapolis	Northwest	
Portland		Alaska, United, Southwest
Philadelphia	US Airways	Southwest
Phoenix		US Airways, Southwest
Pittsburgh	US Airways	
San Diego		American, United, Southwest
Seattle	Alaska	United, Southwest
San Francisco	United	American
Salt Lake City	Delta	Southwest
St. Louis		American, Southwest
Tampa		Continental, Delta, US Airways, Southwest

Notes: Airports are classified according to the distribution of airline shares of airport departures (domestic and international, by all airlines and to all destinations from an airport). An airline share of departures above the departure-weighted median classifies an airport as a dominated hub. Airline shares below the departure-weighted 20th percentile classify the airport as a spoke. Airports in between are classified as non-dominated hubs. Table 4 describes how these definitions classify the origin-destination segments in the estimation sample.

Figure 3: Scheduled departures at Newark Liberty International Airport (EWR) on the morning of June 13, 2007



Scheduled commercial departures include domestic and international departures by all carriers, binned into 5 minute intervals. There were also 12 departures by freight and general aviation, not plotted. Continental Airlines was the only hub airline at Newark. The black dashed line is the 99th percentile of *hourly* departures from Newark from 5 a.m. to 12 p.m. in June 2007 (48 departures), divided by 12 (from FAA Enhanced Traffic Management System data).

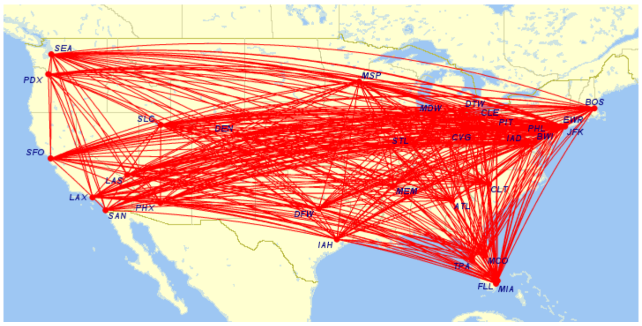
Table 4: Flights in estimation sample by hub classification

Origin	Destination			
	Dominated hub	Non-dominated hub	Spoke	Total
Dominated hub	896	2,006	6,116	9,018
Non-dominated hub	1,998	3,728	3,096	8,822
Spoke	6,104	3,096	562	9,762
Total	8,998	8,830	9,774	27,602

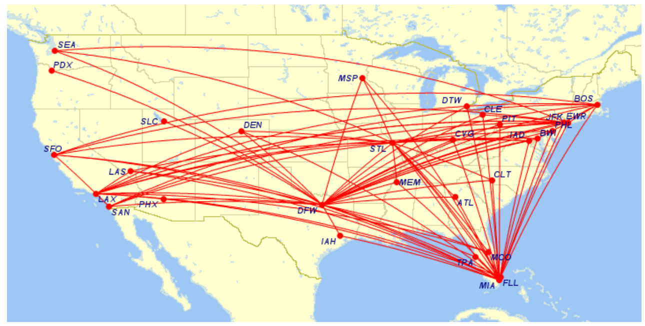
Notes: Flights in estimation sample: flights scheduled between 31 OEP airports on January 17, Feburary 14, April 18 and June 13, 2007. Airline-airport hub classification defined in Table 3.

Figure 4: Routes maps for January 17, February 14, April 18 and June 13, 2007

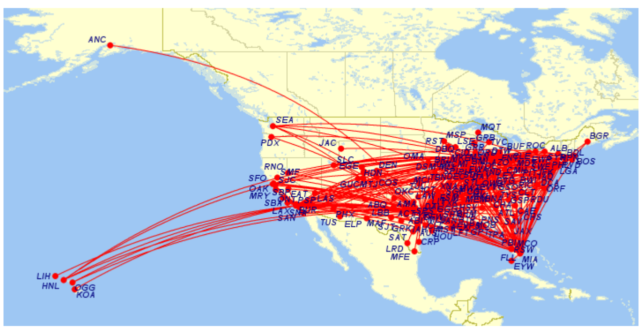
(a) All domestic routes in estimation sample



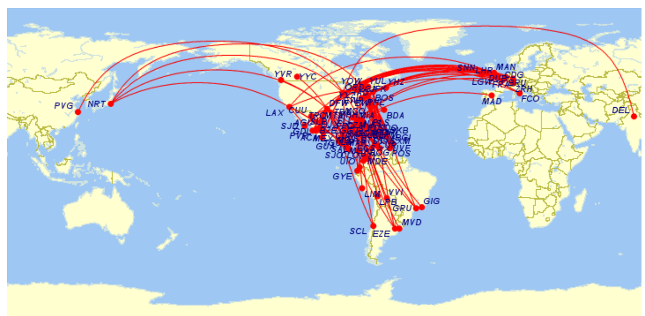
(b) American Airlines routes in estimation sample



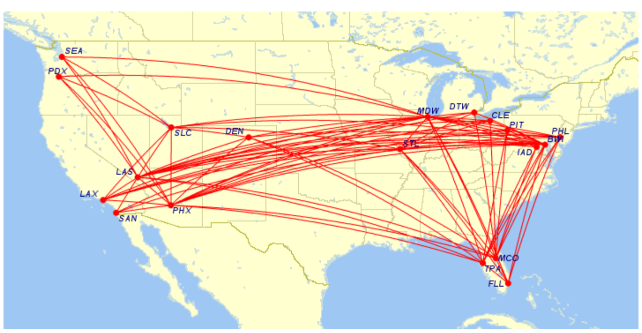
(c) American Airlines domestic routes connecting to routes in estimation sample



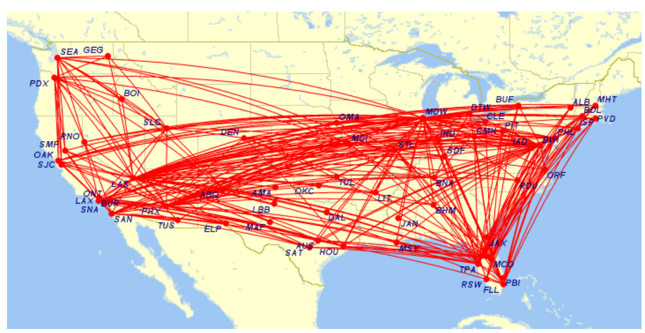
(d) American Airlines international routes connecting to routes in estimation sample



(e) Southwest Airlines routes in estimation sample



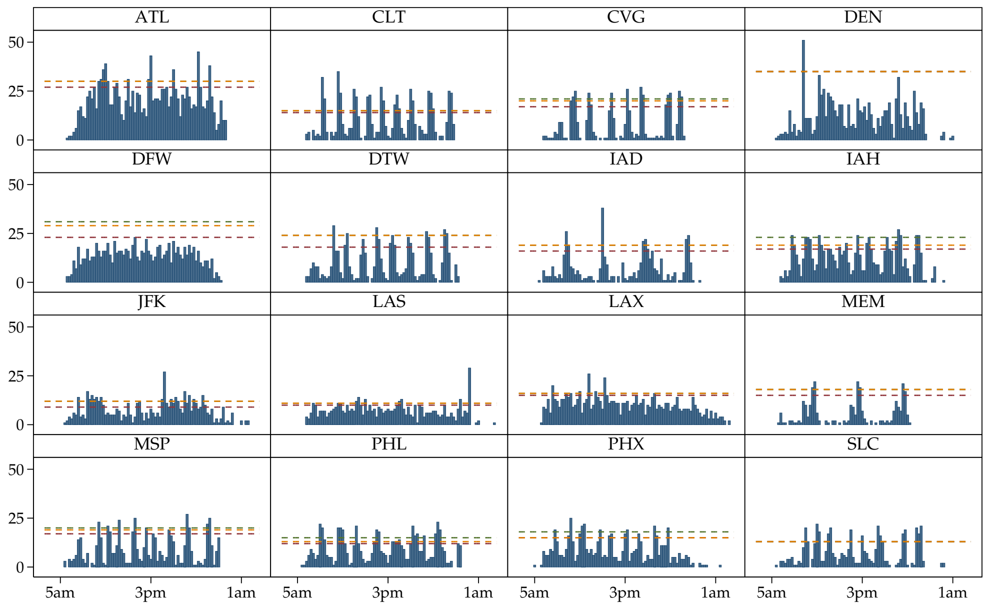
(f) Southwest Airlines domestic routes connecting to routes in estimation sample



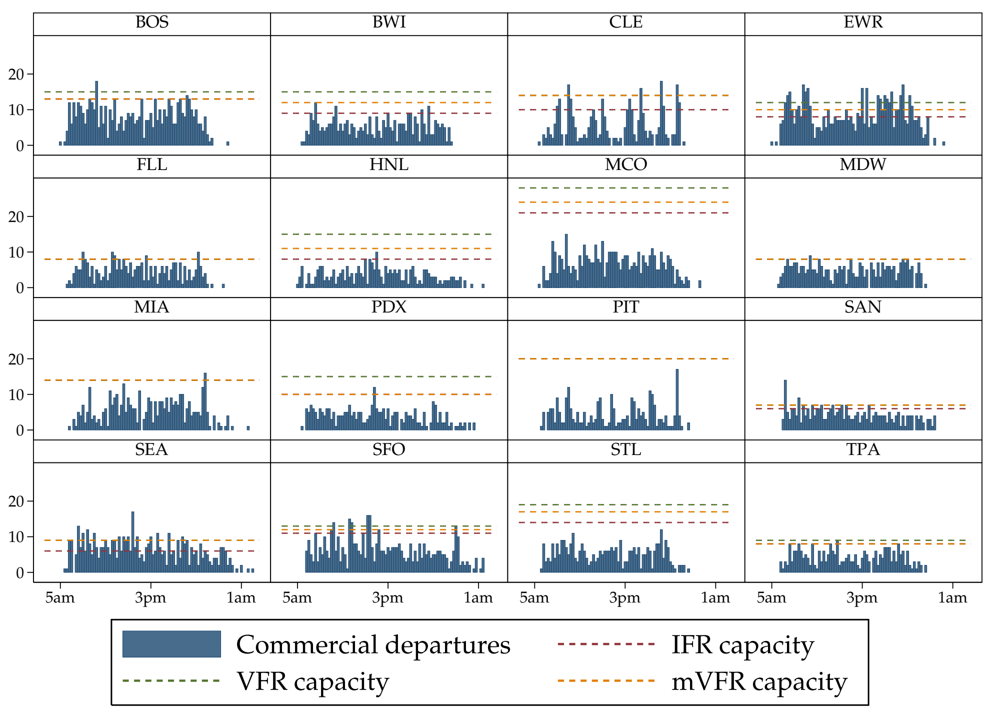
Routes included if active on January 17, February 14, April 18 or June 13, 2007. Plotted with Great Circle Mapper, available at gcmap.com

Figure 5: Scheduled departures and benchmark capacities. June 13, 2007

(a) Airports with more than 20 departures scheduled per quarter hour



(b) Airports with less than 20 departures scheduled per quarter hour



Scheduled commercial departures, including domestic and international departures by all carriers. General aviation not included. Capacity benchmarks from ACBR 2004, allocated to departure according to ATC reported rates in the benchmark report. VFR capacities in green, mVFR in yellow and IFR in red. mVFR (yellow) is overlayed over other flight rules if capacities are equal.

Figure 6: Data structure example

Observed schedules

t	chosen	x_t
1	0	x_1
2	1	x_2
3	0	x_3
4	0	x_4
5	1	x_5
6	0	x_6
7	0	x_7
8	1	x_8
9	0	x_9
10	0	x_{10}

Discrete choice for flight 1

j	$t_{(j,f=1)}$	chosen	x_j
1	1	0	x_1
2	2	1	x_2
3	3	0	x_3
4	4	0	x_4

Discrete choice for flight 2

j	$t_{(j,f=2)}$	chosen	x_j
1	3	0	x_3
2	4	0	x_4
3	5	1	x_5
4	6	0	x_6
5	7	0	x_7

Discrete choice for flight 3

j	$t_{(j,f=3)}$	chosen	x_j
1	6	0	x_6
2	7	0	x_7
3	8	1	x_8
4	9	0	x_9
5	10	0	x_{10}

Panel on the left shows the schedule data, as observed by the researcher, for an example with 10 periods and 3 scheduled flights. Column header t refers to an absolute index of time. Panels on the right show how the data is structured into three discrete choice problems, where column header j indicates the alternative and $t_{(j,f)}$ is the mapping from j to t that is specific to flight f.

Table 5: Summary statistics for expected departure delay (in hours and minutes)

	Total incremental expected delay									Airline incremental expected delay				
	Obs	Mean	Sd	Min	Max	7.00 am	8.00 am	9.00 am	10.00 am	Obs	Mean	Sd	Min	Max
Atlanta	240	0:06	0:16	0:00	1:36	0:00	0:00	0:00	0:44	2,160	0:01	0:05	0:00	1:12
Boston	240	0:01	0:03	0:00	0:15	0:09	0:04	0:14	0:00	2,400	0:00	0:00	0:00	0:06
Baltimore	240	0:00	0:00	0:00	0:04	0:06	0:00	0:03	0:00	2,160	0:00	0:00	0:00	0:02
Cleveland	240	0:02	0:06	0:00	0:32	0:00	0:02	0:13	0:00	1,920	0:00	0:02	0:00	0:30
Charlotte	240	0:21	0:22	0:00	1:17	0:00	0:41	0:00	0:46	1,920	0:10	0:12	0:00	1:11
Cincinnati	240	0:04	0:08	0:00	0:33	0:00	0:00	0:06	0:00	1,680	0:01	0:03	0:00	0:32
Denver	240	0:00	0:01	0:00	0:15	0:00	0:00	0:00	0:00	2,880	0:00	0:00	0:00	0:12
Dallas-Fort Worth	240	0:00	0:00	0:00	0:01	0:00	0:00	0:00	0:00	2,640	0:00	0:00	0:00	0:01
Detroit	240	0:03	0:07	0:00	0:32	0:00	0:00	0:09	0:00	2,640	0:00	0:02	0:00	0:31
New York City Newark	240	1:14	0:45	0:05	3:50	2:07	2:17	2:19	1:19	2,520	0:16	0:18	0:02	2:54
Fort Lauderdale	240	0:13	0:23	0:00	1:45	0:41	0:49	0:00	0:00	2,880	0:02	0:04	0:00	0:29
Washington Dulles	240	0:09	0:20	0:00	1:33	0:00	0:00	0:17	0:00	2,460	0:02	0:06	0:00	1:12
Houston Bush	240	0:05	0:09	0:00	0:48	0:00	0:00	0:24	0:00	1,680	0:01	0:03	0:00	0:46
New York City JFK	240	0:49	1:09	0:00	5:16	0:16	2:27	1:27	0:27	1,920	0:11	0:16	0:00	2:05
Las Vegas	240	0:16	0:28	0:00	2:30	0:11	0:00	0:00	0:22	3,600	0:03	0:06	0:00	1:07
Los Angeles	240	0:08	0:18	0:00	1:43	0:38	0:08	0:12	0:07	3,480	0:01	0:03	0:00	0:28
Orlando	240	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	3,240	0:00	0:00	0:00	0:00
Chicago Midway	240	0:06	0:12	0:00	1:15	0:26	0:03	0:03	0:00	1,800	0:01	0:04	0:00	0:55
Memphis	240	0:02	0:08	0:00	0:47	0:00	0:00	0:00	0:01	1,740	0:00	0:03	0:00	0:44
Miami	240	0:00	0:02	0:00	0:15	0:00	0:00	0:00	0:00	1,920	0:00	0:00	0:00	0:14
Minneapolis	240	0:03	0:08	0:00	0:45	0:06	0:00	0:00	0:00	2,160	0:00	0:03	0:00	0:43
Portland	240	0:00	0:01	0:00	0:12	0:05	0:02	0:00	0:00	2,640	0:00	0:00	0:00	0:07
Philadelphia	240	0:18	0:23	0:00	1:30	0:00	0:31	0:00	0:56	2,400	0:05	0:08	0:00	1:14
Phoenix	240	0:04	0:10	0:00	1:00	0:00	0:00	0:15	0:37	3,300	0:00	0:02	0:00	0:34
Pittsburgh	240	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	2,400	0:00	0:00	0:00	0:00
San Diego	240	0:04	0:10	0:00	1:21	0:59	0:16	0:02	0:12	2,940	0:01	0:02	0:00	0:31
Seattle	240	0:31	0:33	0:00	2:46	1:26	2:07	1:07	0:54	2,940	0:06	0:09	0:00	1:32
San Francisco	240	0:02	0:08	0:00	0:44	0:01	0:00	0:12	0:00	2,760	0:00	0:02	0:00	0:28
Salt Lake City	240	0:10	0:17	0:00	1:14	0:00	0:00	0:07	0:49	2,280	0:03	0:07	0:00	1:08
St. Louis	240	0:00	0:00	0:00	0:00	0:00	0:00	0:00	0:00	2,220	0:00	0:00	0:00	0:00
Tampa	240	0:01	0:05	0:00	0:32	0:05	0:08	0:00	0:00	2,640	0:00	0:01	0:00	0:11
Total	7,440	0:09	0:25	0:00	5:16	0:14	0:18	0:14	0:14	76,320	0:02	0:07	0:00	2:54

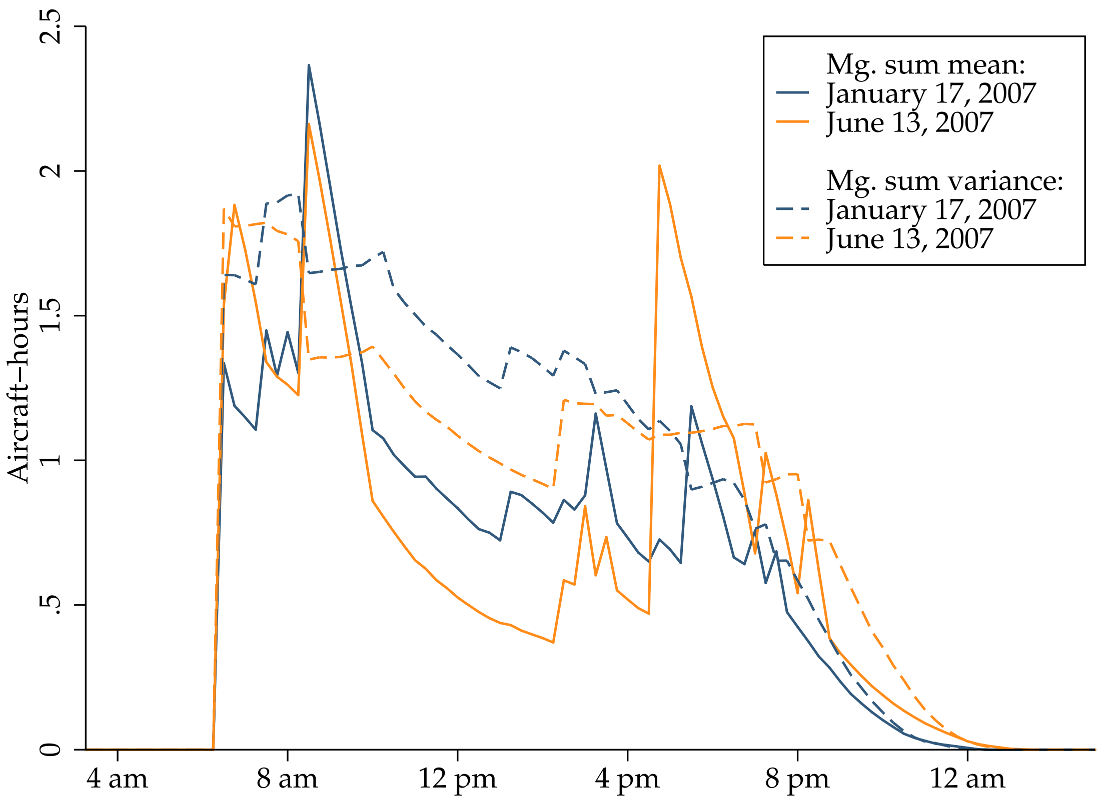
The first five columns present summary statistics for the sum of the incremental expected departure delay to all flights scheduled to depart from each airport, as calculated from the congestion model for the 15-minute period starting at the hours between 7.00 am and 10.00 pm, for the four sample dates (Jan. 17, Feb. 14, Apr. 18 and Jun. 13, 2007). For each airport there are 60 quarter-hours and 4 sample dates for a total 240 observations. The next four columns present means over the four sample dates when conditioning on a scheduled time of departure. The last five columns present summary statistics for the sum of the incremental expected delay to the departures of each airline, for the same 7.00 am to 10.00 pm subset of the hours of the day and sample dates. Observations vary between airports depending on the number of sample airlines operating departures from each airport.

Table 6: Output measures from congestion model predict airline schedule "buffer" and actual delays

	Dependent Variable: Extra schedule "buffer"				Dependent Variable: Flight delay against schedule			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At departure airport:								
Expected delay	0.262***		0.336***	0.232***	0.383***		0.511***	0.202*
	[0.009]		[0.015]	[0.015]	[0.066]		[0.095]	[0.119]
Variance of delay	0.159***		0.241***	0.103***	1.545***		1.419***	1.081***
	[0.020]		[0.033]	[0.036]	[0.109]		[0.158]	[0.223]
At arrival airport:								
Expected delay		0.000	0.002	0.055***		0.338***	0.360***	0.395***
		[0.008]	[0.009]	[0.010]		[0.061]	[0.082]	[0.100]
Variance of delay		0.199***	0.202***	0.170***		0.571***	0.502***	0.149
		[0.013]	[0.017]	[0.017]		[0.094]	[0.122]	[0.137]
Airport effects	*N*	*N*	*N*	*Y*	*N*	*N*	*N*	*Y*
Time-effects	*N*	*N*	*N*	*Y*	*N*	*N*	*N*	*Y*
Observations	63,643	63,600	27,646	27,646	43,516	43,315	22,917	22,917
R^2	0.02	0.01	0.06	0.35	0.01	0.01	0.01	0.08

Observations are individual flights scheduled on January 17, February 14, April 18 and June 13, 2007. For each dependent variable, the sample is restricted to departures from the 31 sample airports in the first column, arrivals to the 31 sample airports in the second column, and flights between the 31 sample airports in the third and fourth columns. An observation in the first four columns is a scheduled flight and the dependent variable is the difference between a flight's scheduled block time and the minimum block time scheduled for the same origin-destination-airline within the same month. An observation in the second set of columns is an actual flight included in the BTS On-Time-Performance database, and the dependent variable is the flight delay relative to the scheduled arrival time. Controls are predicted delay metrics from the congestion model: the expected delay and variance of delay to each observation from departure, arrival, or both. Full details on the construction of these are in the text and Data Appendix. Airport-effects are specific to arrival and departure, and time-effects are quarter-hour dummies for both departure times and arrival times. Robust standard errors are reported in brackets (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

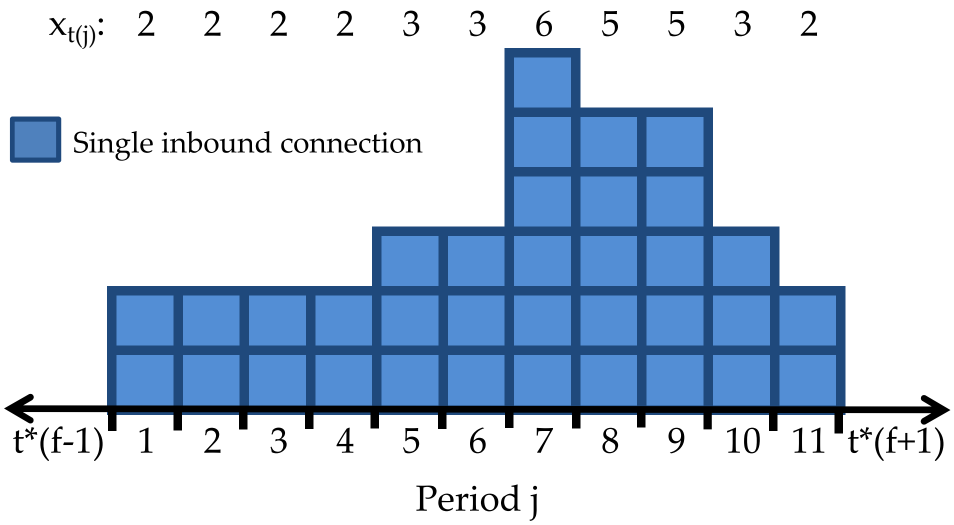
Figure 7: Marginal expected departure delay effects
at Newark Liberty International Airport



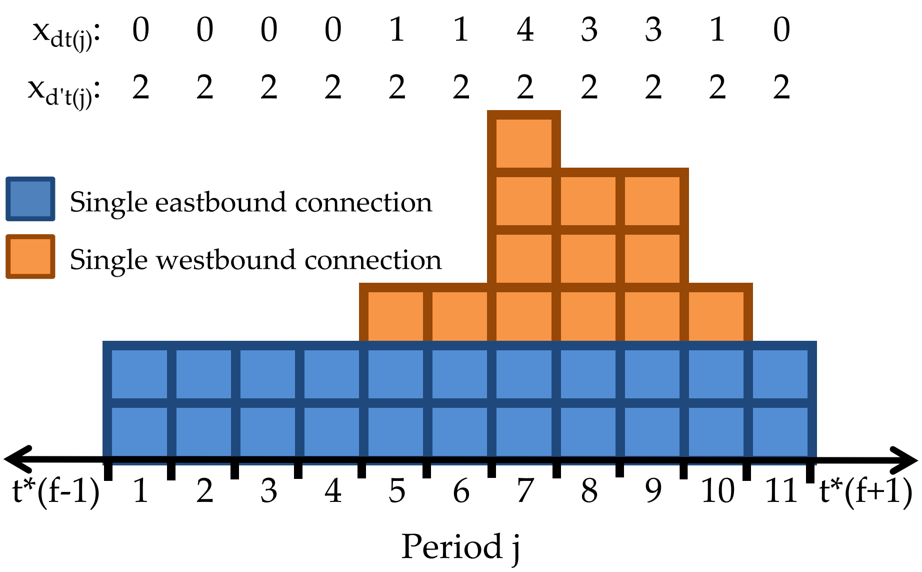
Solid lines are the sum of the incremental expected departure delay to all flights scheduled by Continental Airlines to depart from Newark Liberty International Airport (EWR), measured in aircraft-hours, given the addition of a single flight at each time of the day. Dashed lines are the sum of the incremental variance of departure delay for the same set of flights. Computed using the congestion model described in Section 5 for the sample dates of January 17 and June 13, 2007.

Figure 8: Inbound arrivals for connection model example

(a) Scheduling time alternatives and inbound connections for a flight *odayf*

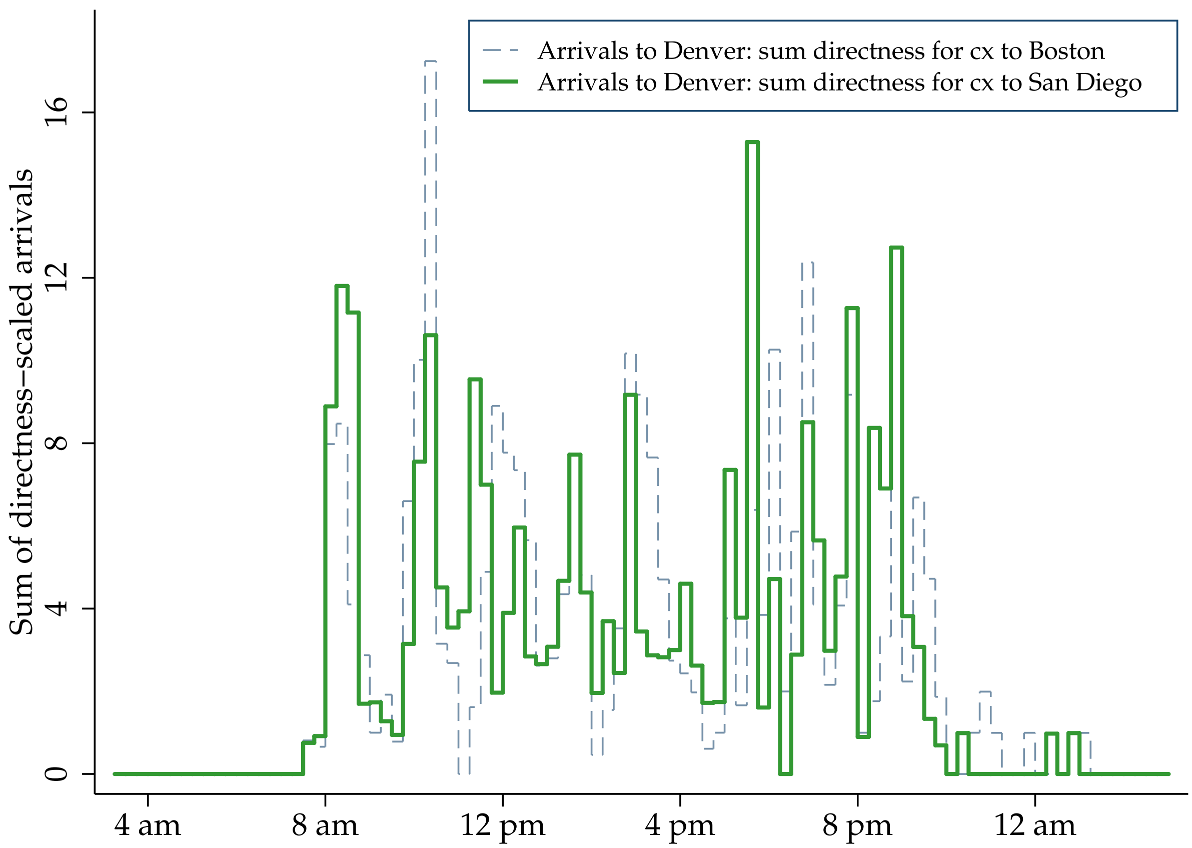


(b) Scheduling time alternatives and inbound connections for two flights to destinations d and d'



Note: A previous departure on the same origin-destination-airline is scheduled at time $t^*(f-1)$ and a successor is scheduled at time $t^*(f+1)$ in both figures. The successor and predecessor are scheduled 3 hours apart, leaving $j = 1, ..., 11$ quarter-hour alternative periods for the scheduling of flight f. $x_{t(j)}$ is the number of inbound flights that become available for connection in period t to a flight departing from airport o. In figure (b), $x_{dt(j)}$ and $x_{d't(j)}$ count mutually-exclusive subsets of the arrivals, available for connection to the destinations d and d', respectively.

Figure 9: United Airlines arrivals at Denver, June 13, 2007, scaled by directness-angle for connection to Boston or San Diego.



Note: United Airlines domestic arrivals at Denver on June 13, 2007, scaled by directness angle measure between origin of arrival, Denver, and Boston or San Diego.

Figure 10: Example airline choices over attributes of connections

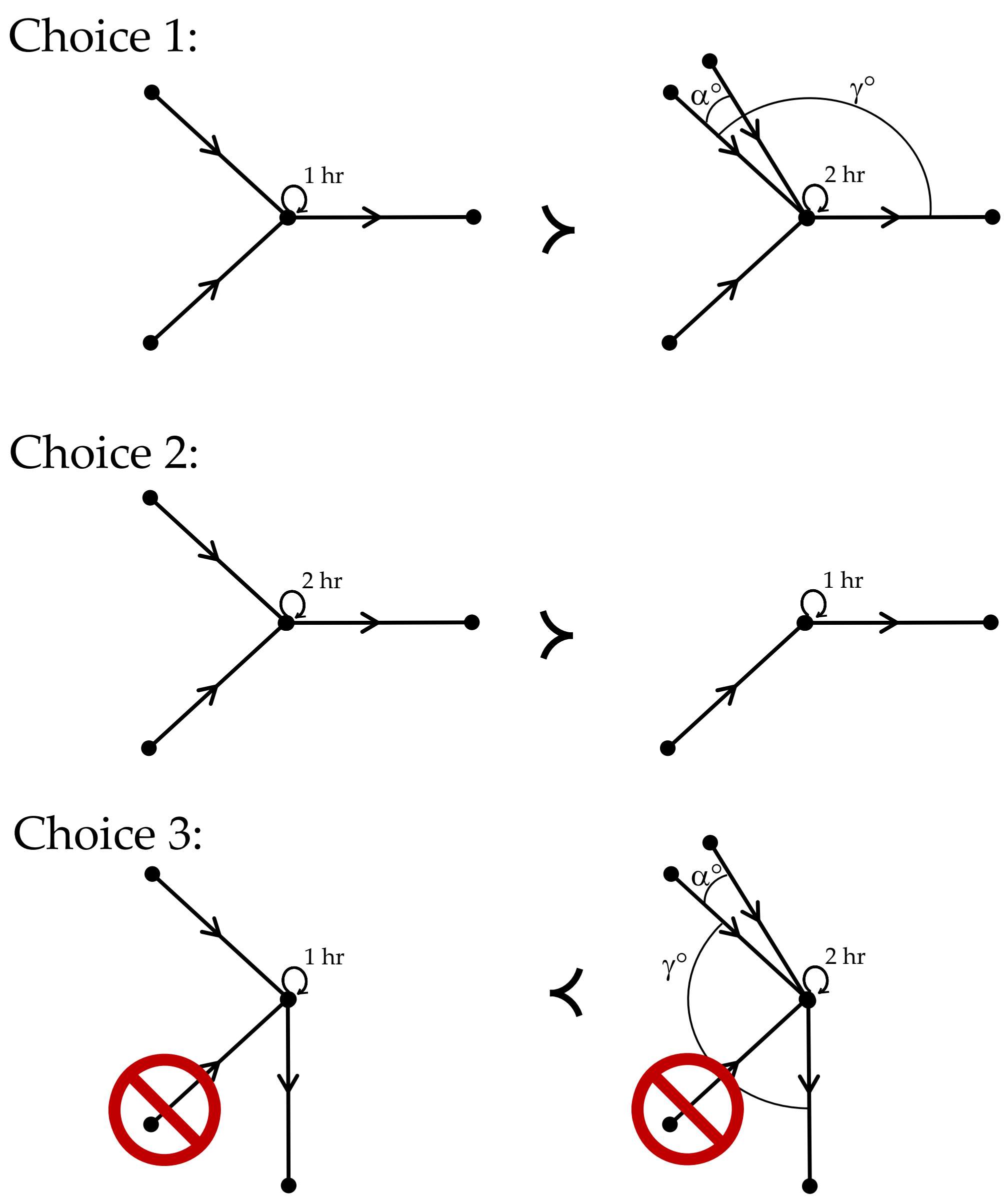


Figure 11: Example construction of the pull instrument for departure congestion measures

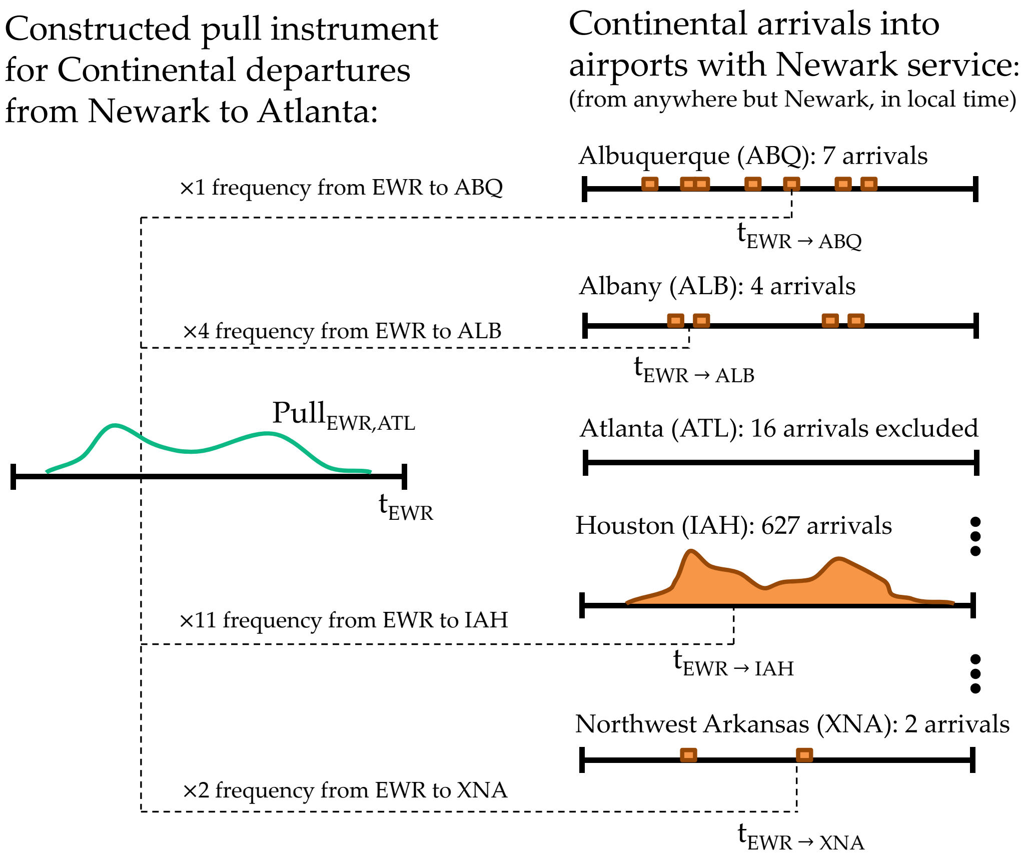
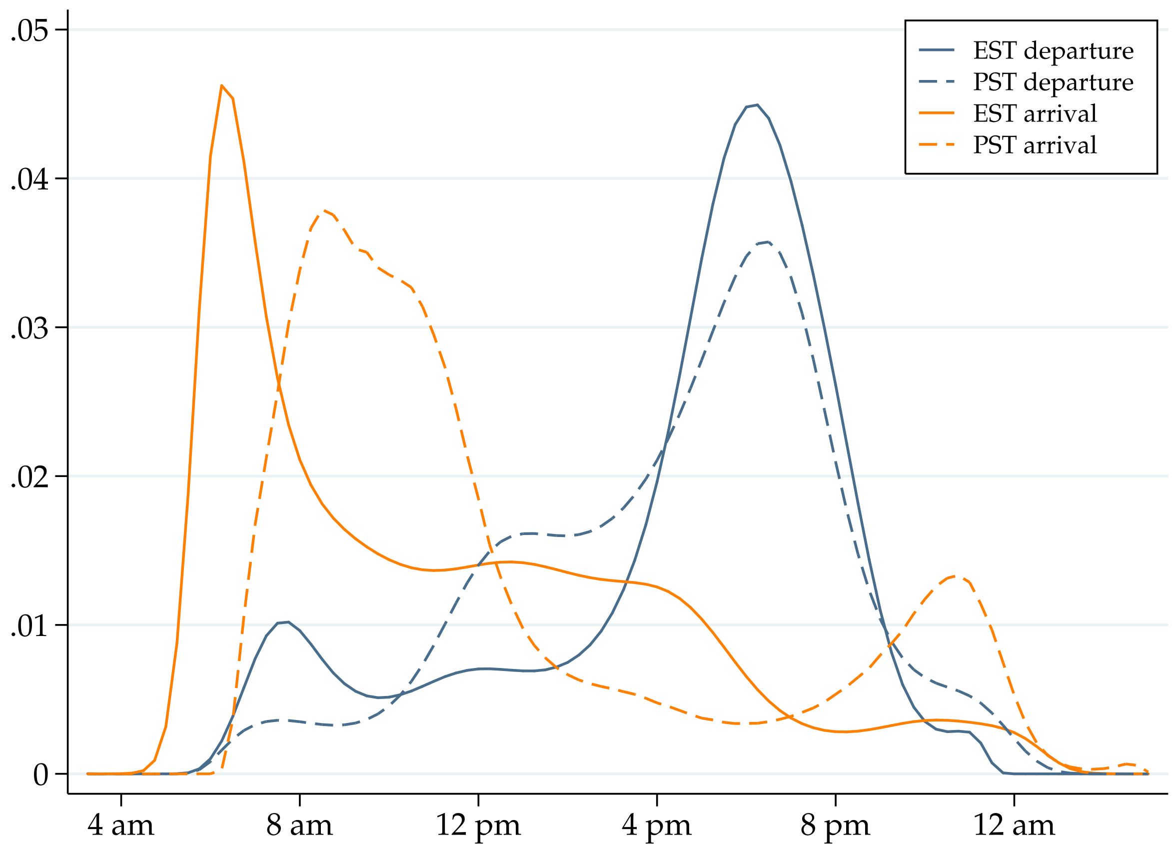


Figure 12: Schedule choice model results: time-of-day effects



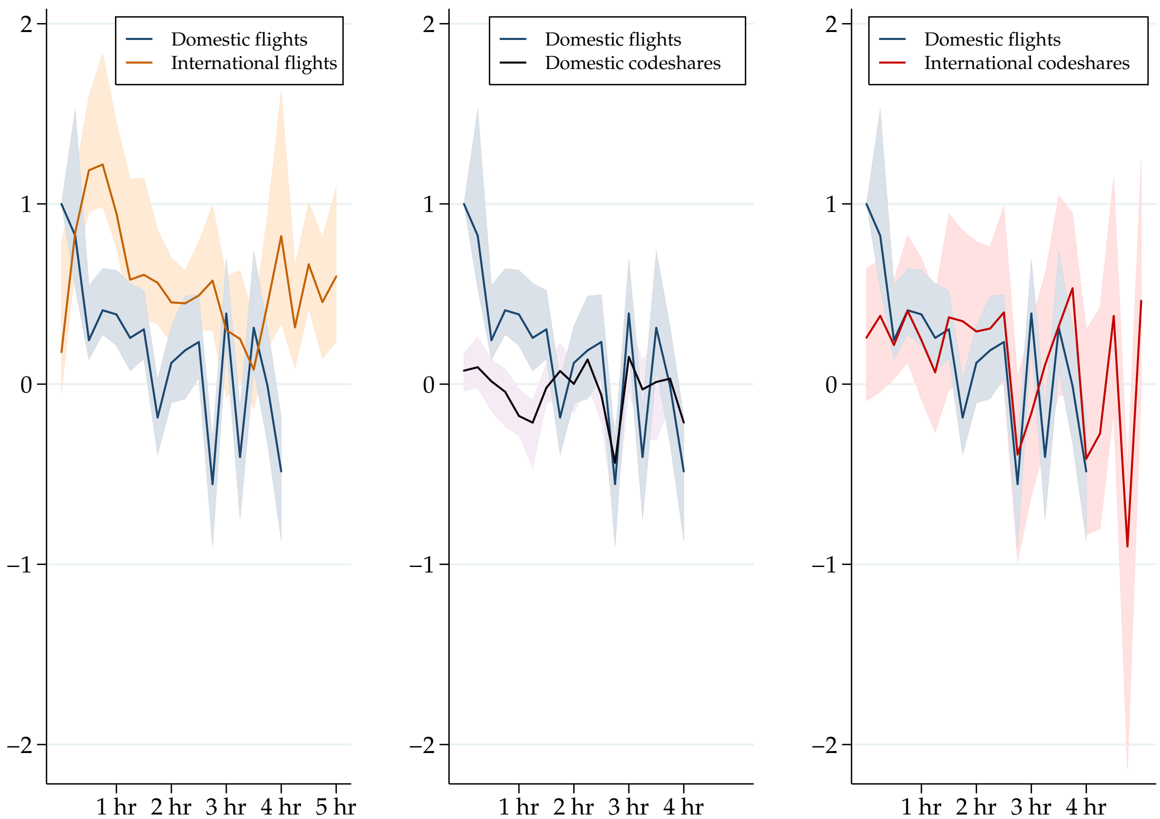
Estimated conditional probabilities of arrival (orange) and departure (blue), calculated from time-of-day effect polynomials for the Eastern Standard Time-zone (solid) and Pacific Standard Time-zone (dashed).

Table 7: Linear probability model results

	Dependent variable: Schedule time choice (in %)							
	(1) LPM OLS	(2) LPM 2SLS	(3) LPM 2SLS	(4) LPM 2SLS	(5) LPM 2SLS	(6) LPM 2SLS	(7) LPM 2SLS	(8) LPM 2SLS
Incremental mean delay								
Dominant hub	1.681***	0.920***		1.127***	0.894*	1.048		2.971**
	[0.118]	[0.122]		[0.293]	[0.375]	[0.678]		[1.003]
Non-dominant hub	0.562***	0.545***		0.495**	0.577	-1.048		-2.534**
	[0.047]	[0.101]		[0.187]	[0.310]	[0.601]		[0.817]
Spoke	0.099	-0.041		-1.029	-1.324	2.241		8.480*
	[0.125]	[0.429]		[0.795]	[1.049]	[1.843]		[3.756]
Incremental variance of delay								
Dominant hub	-0.190***		0.156***	-0.046	0.012		-0.107	-0.567*
	[0.021]		[0.024]	[0.057]	[0.076]		[0.166]	[0.244]
Non-dominant hub	0.014		0.351***	0.049	0.160		-0.076	1.348**
	[0.031]		[0.073]	[0.136]	[0.232]		[0.324]	[0.451]
Spoke	0.049		0.081	0.368	0.099		-0.015	-1.366*
	[0.037]		[0.128]	[0.237]	[0.314]		[0.304]	[0.627]
Incr. M. Delay × Competitor push-pull								
Dominant hub						-0.089		-1.436
						[0.499]		[0.751]
Non-dominant hub						1.403**		2.820***
						[0.529]		[0.734]
Spoke						-1.537		-6.574**
						[1.226]		[2.487]
Incr. V. Delay × Competitor push-pull								
Dominant hub							0.192	0.395*
							[0.121]	[0.181]
Non-dominant hub							0.405	-1.324**
							[0.307]	[0.465]
Spoke							0.069	1.221**
							[0.219]	[0.450]
Dom. connections: immediate	0.241***	0.244***	0.248***	0.243***	0.244***	0.245***	0.250***	0.245***
	[0.019]	[0.020]	[0.020]	[0.020]	[0.020]	[0.020]	[0.020]	[0.020]
Dom. connections: extra 1 hour	0.060***	0.066***	0.071***	0.064***	0.066***	0.066***	0.071***	0.060***
	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]
Int. connections: immediate	-0.026	-0.034	-0.019	-0.032	-0.037	-0.022	-0.016	-0.017
	[0.073]	[0.073]	[0.073]	[0.073]	[0.074]	[0.073]	[0.073]	[0.073]
Int. connections: extra 1 hour	0.303***	0.294***	0.302***	0.295***	0.290***	0.299***	0.305***	0.292***
	[0.071]	[0.071]	[0.071]	[0.071]	[0.072]	[0.071]	[0.071]	[0.072]
Observations	600384	600384	600384	600384	600384	600384	600384	600384
R^2	0.055	0.054	0.053	0.055	0.054	0.052	0.052	0.049
Weather instruments	N	Y	Y	Y	Y	Y	Y	Y
Push-pull instruments	N	Y	Y	Y	N	Y	Y	Y
Connection terms (72)	Y	Y	Y	Y	Y	Y	Y	Y
Time-zone time polynomials (2x4x16)	Y	Y	Y	Y	Y	Y	Y	Y
Quarter-hour effects (6)	Y	Y	Y	Y	Y	Y	Y	Y
Fixed-effects (o-d-a-y)	Y	Y	Y	Y	Y	Y	Y	Y

In all specifications: Time of day effects (16 degree polynomials) at both origin and destination, for four time zones. Standard errors clustered at the origin-destination-airline-day level. Sample: All flights between busiest 35 US airports, minus DCA, LGA, ORD, HNL, for Wed Jan 17, Wed Feb 14, Wed Apr 18 and Wed Jun 13, 2007.

Figure 13: Schedule choice model estimates: connection parameters by duration of layover in excess of minimum connection time



Discounted connection coefficients for four types of connection: domestic flights (plotted for comparison in all panels), international flights (left panel), domestic codeshares (center panel) and international codeshares (right panel). Are coefficients are normalized in terms of a domestic connection with minimum layover, plotted in blue in every figure. 95% confidence intervals for the ratio of coefficients are bootstrapped under specification (4) of Table 8.

Table 8: Schedule choice model results: average marginal structural effects

	Dependent variable: Schedule time choice				
	C. Logit (1)	Control Function* (mean only) (2)	Control Function* (var only) (3)	Control Function* (both) (4)	Control Function (both) (5)
Incremental mean delay					
Dominant hub	0.562^{***}	0.900^{***}		1.240^{***}	1.470
	(0.457 , 0.664)	(0.558 , 1.272)		(0.560 , 2.004)	(∗)
	[0.464 , 0.645]	[0.593 , 1.191]		[0.652 , 1.917]	(∗)
Non-dominant hub	0.787^{***}	1.070^{***}		1.409^{***}	1.471
	(0.652 , 1.020)	(0.650 , 1.513)		(0.891 , 1.928]	(∗)
	[0.664 , 0.998]	[0.653 , 1.490]		[0.965 , 1.886]	(∗)
Spoke	0.148	-1.174		-0.277	-0.368
	(-0.122 , 0.575)	(-1.718 , 0.030)		(-1.812 , 1.886]	(∗)
	[-0.109 , 0.445]	[-1.678 , 0.028]		[-1.578 , 1.741]	(∗)
Incremental variance of delay					
Dominant hub	-0.017		0.098^{*}	-0.087	-0.116
	(-0.099 , 0.037)		(-0.029 , 0.171)	(-0.289 , 0.054)	(∗)
	[-0.086 , 0.034]		[0.028 , 0.159]	[-0.242 , 0.026]	(∗)
Non-dominant hub	0.361^{***}		0.233	-0.080	-0.089
	(0.183 , 0.484)		(-0.086 , 0.508)	(-0.438 , 0.307)	(∗)
	[0.188 , 0.447]		[-0.070 , 0.467]	[-0.404 , 0.171]	(∗)
Spoke	0.041		-0.409^{***}	-0.308	-0.292
	(-0.151 , 0.153)		(-0.781 , -0.058)	(-0.897 , 0.319)	(∗)
	[-0.126 , 0.146]		[-0.597 , -0.061]	[-0.761 , 0.244]	(∗)
Dom. connections: immediate	0.562^{***}	0.871^{***}	0.896^{***}	0.869^{***}	1.042
	(0.540 , 0.590)	(0.577 , 1.037)	(0.583 , 1.063)	(0.560 , 1.037)	(∗)
	[0.542 , 0.580]	[0.587 , 1.036]	[0.632 , 1.058]	[0.593 , 1.026]	(∗)
Dom. connections: extra 1 hour	0.323^{***}	0.327^{***}	0.334^{***}	0.324^{***}	0.335
	(0.301 , 0.349)	(0.205 , 0.465)	(0.210 , 0.467)	(0.199 , 0.463)	(∗)
	[0.303 , 0.348]	[0.232 , 0.456]	[0.241 , 0.466]	[0.232 , 0.451]	(∗)
Int. connections: immediate	0.615^{***}	0.154^{*}	0.197^{**}	0.147	0.158
	(0.492 , 0.745)	(-0.039 , 0.431)	(0.002 , 0.469)	(-0.043 , 0.427)	(∗)
	[0.492 , 0.724]	[0.010 , 0.399]	[0.145 , 0.148]	[-0.005 , 0.395]	(∗)
Int. connections: extra 1 hour	0.840^{***}	0.804^{***}	0.799^{***}	0.822^{***}	0.808
	(0.747 , 0.990)	(0.677 , 0.969)	(0.690 , 0.946)	(0.708 , 0.985)	(∗)
	[0.753 , 0.957]	[0.705 , 0.961]	[0.693 , 0.944]	[0.721 , 0.973]	(∗)
Error component $\sigma_{\tilde{\varepsilon}}$		(∗)	(∗)	(∗)	0.159
					(∗)
Weather instruments	N	Y	Y	Y	Y
Push-pull instruments	N	Y	Y	Y	Y
Connection terms (72)	Y	Y	Y	Y	Y
Time-zone time polynomials (2x4x16)	Y	Y	Y	Y	Y
Quarter-hour effects (2x3)	Y	Y	Y	Y	Y
Pseudo R^2	0.27	0.27	0.27	0.27	
Log lik.	-63295	-63284	-63281	-63281	-49557

Bootstrapped confidence intervals in parentheses (95%) and brackets (90%). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In all specifications: Time of day effects (16 degree polynomials) at both origin and destination, for four time zones. Sample: All flights between busiest 35 US airports, minus DCA, LGA, ORD, HNL, for Wed Jan 17, Wed Feb 14, Wed Apr 18 and Wed Jun 13, 2007.

Table 9: Schedule choice model estimates. Interaction with competitor push-pull measure.

	Dependent variable: Schedule time choice	
	Control Function* (both) (1)	Control Function* (mean only) (2)
Incremental mean delay		
Dominant hub	-1.033***	-0.575**
	[0.258]	[0.175]
Non-dominant hub	0.572**	0.452***
	[0.213]	[0.137]
Spoke	-2.831*	-1.041
	[1.420]	[0.785]
Incremental variance of delay		
Dominant hub	0.149*	
	[0.068]	
Non-dominant hub	-0.037	
	[0.133]	
Spoke	0.503	
	[0.286]	
Incr. M. delay $\times$ Competitor push-pull		
Dominant hub	1.082***	0.597***
	[.210]	[0.133]
Non-dominant hub	-0.253	-0.223
	[0.199]	[0.123]
Spoke	1.996*	0.488
	[0.974]	[0.555]
Incr. V. delay $\times$ Competitor push-pull		
Dominant hub	-0.146**	
	[0.053]	
Non-dominant hub	-0.014	
	[0.130]	
Spoke	-0.440*	
	[0.209]	
Weather instruments	Y	Y
Push-pull instruments	Y	Y
Connection terms (72)	Y	Y
Time-zone time polynomials (2x4x16)	Y	Y
Quarter-hour effects (2x3)	Y	Y
Fixed-effects (o-d-a-y)	Y	Y
Log lik.	-63205	-63220

Standard errors *not* bootstrapped, in brackets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
In both specifications: Time of day effects (16 degree polynomials) at both origin and destination, for four time zones. Sample: All flights between busiest 35 US airports, minus DCA, LGA, ORD, HNL, for Wed Jan 17, Wed Feb 14, Wed Apr 18 and Wed Jun 13, 2007.

Appendix A Estimates of non-commercial runway use

Scheduled commercial airlines are only one of the users of public airports, which are also used by general aviation, military, freight and non-scheduled commercial aviation. The FAA's Enhanced Traffic Management System (ETMS) records arrivals and departures by hour, which I model as subtractions from capacity (since the ETMS data records the time at which the operation takes place, rather than the user's intended time of arrival or departure). From the perspective of a scheduling airline, a non-commercial user that takes-off or lands at time t effectively reduces residual capacity by one flight at t, regardless of the time at which that user started queuing to claim a runway slot at t.

For each airport and a sample of all hours on Tuesdays, Wednesdays and Thursdays from 2002 through 2006 I regress hourly non-commercial departures (or arrivals) on hour effects, month effects, weather state effects (VFR, marginal VFR or IFR weather) and interactions of month and weather, month and hour, and weather and hour. From these regressions I obtain state-dependent predictions on the number of non-commercial users per airport and hour of the day under each type of weather, for the volume of non-commercial use that prevailed from 2002 to 2006. I scale these up by the airport-specific ratio of non-commercial operations in 2007 to the average for 2002-2006. I round the hourly predictions and assign them to higher frequency quarter hours using the quadratic interpolation procedure of Denton (1971).

Appendix B Queuing model

I model queues (and associated delays) as arising from the interaction of scheduled claims for access to the runway and a state-dependent time-path of capacity. I describe the queueing model in terms of departures, but employ the same model for arrivals after mapping scheduled departure times to expected arrival times. For a given airport and day, divide the day into quarter-hour periods indexed by $t = 1, \ldots, T$. Let H_t represent the number of flights scheduled to depart in period t and C_t the departure capacity over period t under a particular time-path for capacity, where I omit the capacity-path state index s.

Aircraft with a scheduled departure time at t are assumed to claim access to the runway at t, and since access is granted under a first-come, first-serve (FCFS) priority these will have priority over aircraft scheduled at $t' > t$. I assume that aircraft scheduled within the same period are served in random order.[41]

Assume that there are no queued flights awaiting departure at the start of the day, i.e. at $t = 1$. If $H_1 \leq C_1$, then all H_1 flights making a claim to depart in period 1 will

[41] As a result of requesting runway access in random order due to idiosyncratic factors (e.g. passenger boarding time) realized on the flight date. Air traffic controllers sequence both arriving and departing aircraft for access to a runway according to a first-come, first-serve (FCFS) queue discipline (De Neufville and Odoni, 2003). However, the FAA carries out system-wide air traffic flow management measures that can also affect delay: for example, the FAA forecasts arrival demand in real time and issues "ground holds", delaying flights destined for a heavily congested airport at their point of origin. Since ground holds do not preserve the exact order in which aircraft would have reached air traffic controllers for access to a runway, they will cause deviations from delays under a strict FCFS discipline, particularly for arrivals.

depart on time. If $H_1 > C_1$, only C_1 flights will be able to depart on time and $H_1 - C_1$ flights will be placed in a queue and delayed at least one period. The "serve-in-random-order" assumption for concurrent flights implies that all flights scheduled for t have the same ex-ante probability of departure at t' conditional on not having departed by t', for any $t' \geq t$. Additionally, FCFS priority discipline means that all flights in the queue from a period t have priority for departure over flights scheduled for period $t' > t$, so the probability that a flight scheduled for period 1 will depart in period 1 is $q_{11} = \frac{C_1}{\max[H_1, C_1]}$, the probability that it will depart in period 2 is $q_{12} = (1 - q_{11}) \frac{C_2}{\max[H_1 - C_1, C_2]}$, for period 3 is $q_{13} = (1 - q_{11})(1 - q_{12}) \frac{C_3}{\max[H_1 - C_1 - C_2, C_3]}$ and so on through q_{1T}, where I assume departure capacity in period T is unbounded.[42] Consider a period t, which is described by H_t, C_t and a queue set Q_t which contains all the as-of-yet undeparted flights scheduled for periods prior to t. Total demand on the runway at t is $H_t + |Q_t|$. The probability that a flight scheduled at t departs on time is given by $q_{tt} = \frac{\max[C_t - |Q_t|, 0]}{\max[H_t, C_t - |Q_t|]}$. The probability that it departs at $t' > t$ is given by

$$q_{tt'} = \frac{\overbrace{\max[C_{t'} - \max[\varrho_{tt'}, 0], 0]}^{\text{period } t' \text{ capacity available to period } t \text{ claims}}}{\max[\underbrace{H_t - \max[-\varrho_{tt'}, 0]}_{\text{claims at } t \text{ not departed by } t'}, C_{t'} - \max[\varrho_{tt'}, 0]]} \prod_{\tau=t}^{t'-1} (1 - q_{t\tau}).$$

where $\varrho_{tt'} \equiv |Q_t| - \sum_{\tau=t}^{t'-1} C_\tau$ is the residual Q_t queue at time t'.

We can use $q_{tt'}$ to calculate the departure delay for a flight scheduled at period t:

$$D_t = \sum_{\tau=t+1}^{T} (\tau - t)\, q_{t\tau}$$

where D_t is conditional on a particular time-path for airport capacity.

Appendix C Connection model: mapping schedule decisions to a set of enabled connections

Consider an origin-destination-airline-day schedule composed of $f = 1, \ldots, F$ flights. Let x_{ot} be a measure of inbound arrivals into the origin airport that become available to connect to flight $odayf$ in period t, where I suppress indices for destination, airline and calendar day. The measure of inbound arrivals could be a count of arrivals or the sum directness measure described in the text, for any particular type of connection, e.g. domestic or international, intraline or codeshare.[43] Scheduling a flight f between the

[42] In the empirical application I define $t = 1$ to be the period from 3.01 a.m. to 3.15 a.m. in the early morning of y and $t = T$ to be the period from 2.46 a.m. to 3 a.m. on the night of y. By this time and at every airport there have always been several hours of surplus runway capacity, so the assumptions on queue clearing at T are inconsequential.

[43] For example, x_{ot} could stand for the number of inbound international flights scheduled to arrive at some period τ, defined by $\tau \equiv t - MCT_{oa,i \to d}$, where $MCT_{oa,i \to d}$ is the minimum connection time for

scheduled times for neighboring flights $f-1$ and $f+1$ defines a set of $j = 1, \dots, J_f$ schedule choice alternatives. Scheduling the departure at any period j creates $x_{ot(j)}$ potential connections that have no layover in excess of the minimum connection time. Scheduling at any $j > 1$ enables a connection to $x_{t(j-1)}$ (i.e. the inbound arrivals that become available for connection one period before the j departure), but the connection has a layover of one period in excess of the minimum connection time. Scheduling at any $j < J_f$ will not be able to connect to the arrivals at $x_{t(J_f)}$ by definition, but these arrivals can make connections to the departure of the successor flight $f+1$ with a one-period layover. Let β_k represent a discount factor for the value to the airline of a connection with a layover of k periods in excess of the minimum connection time, and let π_{oj} represent the sum of discounted connections at the flight's origin.

Assumption 1. ***(Exclusion of dominated connections)*** *A feasible connection to flight $odayf$ is excluded from all alternative-specific payoffs in the $odayf$ choice problem if there exists a flight $odayf'$ such that the same connection can be made through f' with a shorter layover than through f .*

Under A.1 the vector $\boldsymbol{\pi}_o \in \mathbb{R}^{J_f}$ for the alternative-specific sum of discounted inbound-connection payoffs is a linear transformation of the vector of available connections $\mathbf{x}_o \in \mathbb{R}^{J_f}$, given by a matrix of discount factors $\boldsymbol{\beta}_o$:

$$\begin{bmatrix} \pi_{o1} \\ \pi_{o2} \\ \pi_{o3} \\ \vdots \\ \pi_{oJ_f-1} \\ \pi_{oJ_f} \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_{J_f-1} & \beta_{J_f-2} & \cdots & \beta_2 & \beta_1 \\ \beta_1 & \beta_0 & \beta_{J_f-2} & \cdots & \beta_2 & \beta_1 \\ \beta_2 & \beta_1 & \beta_0 & \cdots & \beta_2 & \beta_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{J_f-2} & \beta_{J_f-3} & \beta_{J_f-4} & \cdots & \beta_0 & \beta_1 \\ \beta_{J_f-1} & \beta_{J_f-2} & \beta_{J_f-3} & \cdots & \beta_1 & \beta_0 \end{bmatrix} \cdot \begin{bmatrix} x_{ot(1)} \\ x_{ot(2)} \\ x_{ot(3)} \\ \vdots \\ x_{ot(J_f-1)} \\ x_{ot(J_f)} \end{bmatrix} \tag{14}$$

The above discussion accounts for the elements in the diagonal, subdiagonal and last column of matrix $\boldsymbol{\beta}$, the remaining elements following the same logic. A similar argument holds at destination for a vector of payoffs $\boldsymbol{\pi}_d = \boldsymbol{\beta}_d \mathbf{x}_d$, after appropriately indexing scheduled departure times to connection times with outbound flights at the destination airport d:

$$\begin{bmatrix} \pi_{d1} \\ \pi_{d2} \\ \pi_{d3} \\ \vdots \\ \pi_{dJ_f-1} \\ \pi_{dJ_f} \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_{J_f-2} & \beta_{J_f-1} \\ \beta_1 & \beta_0 & \beta_{J_f-2} & \cdots & \beta_{J_f-3} & \beta_{J_f-2} \\ \beta_1 & \beta_2 & \beta_0 & \cdots & \beta_{J_f-3} & \beta_{J_f-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_0 & \beta_1 \\ \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_{J_f-1} & \beta_0 \end{bmatrix} \cdot \begin{bmatrix} x_{dt(1)} \\ x_{dt(2)} \\ x_{dt(3)} \\ \vdots \\ x_{dt(J_f-1)} \\ x_{dt(J_f)} \end{bmatrix} \tag{15}$$

From these payoff matrices we obtain alternative-specific regressors that are a function of the entire vector of connections entering schedule $odayf$ (i.e. $\mathbf{x}_o$ and $\mathbf{x}_d$) and that interact with the individual terms of the discount function β_k. The sum of the connection terms multiplied by discount factor β_k that enter alternative j are:

international to domestic connections specific to this airport and airline.

$$
\begin{aligned}
z_{j,k} = {} & \overbrace{x_{ot(j-k)} \cdot \mathbf{1}\left[j-k \geqslant 1\right]}^{\text{inbound cx. made by } f} + \overbrace{x_{ot\left(J_f+1-k\right)} \cdot \mathbf{1}\left[j+k \leqslant J_f\right]}^{\text{inbound cx. left for } f+1} \\
& + \underbrace{x_{dt(j+k)} \cdot \mathbf{1}\left[j+k \leqslant J_f\right]}_{\text{outbound cx. shortened by } f} + \underbrace{x_{dt(k)} \cdot \mathbf{1}\left[j-k \geqslant 1\right]}_{\text{outbound cx. left for } f-1}
\end{aligned} \tag{16}
$$

for $k = 1, ..., J_f - 1$ and $j = 1, ..., J_f$. The $z_{j,k}$ terms are alternative-varying regressors that allow us to estimate the β_k discount parameters by shifting the measure of the potential connections enabled with a layover of any length under each schedule alternative.

The assumption of exclusion of dominated connections is not essential but is a parsimonious way to reduce the set of connections to track. Allowing the airline to value dominated connections requires a second set of terms in equation (16), as well as a second set of discount parameters. I make a further assumption to reduce the accounting of connections for the first ($f = 1$) and last ($f = F$) flight of the day.

Assumption 2. ***(No 3 a.m. layovers)*** *Connections that would require a passenger to wait for a connection at 3 a.m. local time do not generate a connection payoff.*

Under A.2 I obtain analogues of the $\boldsymbol{\beta}_o$, $\boldsymbol{\beta}_d$ matrices: The upper triangle of $\boldsymbol{\beta}_o$ is zeroed-out for the last flight of the day, as flights inbound to o that cannot connect to $odayF$ will generate no connection value. The lower triangle of $\boldsymbol{\beta}_d$ is zeroed-out for the same reason on the first flight of the day. On the last flight of the day, the matrix $\boldsymbol{\beta}_d$ is not square: departures on $oday$ may arrive at destination on $oda\,(y+1)$, and connect to departures in the morning of $y+1$ so $\boldsymbol{\beta}_d \in \mathbb{R}^{J_F \times (J_F + t^*(oda(y+1)1))}$ where $t^*\,(oda\,(y+1)\,1)$ is the scheduled time for the first successor to $odayF$ on the next calendar day $y+1$. Assumption A.2 requires zeroing-out the rectangular submatrix of $\boldsymbol{\beta}_d$ above and to the right of element (J_F, J_F+1). Analogues of equation (16) are used for the first and last flight of the day.