Robust Scheduling Practices in the U.S. Airline Industry: Costs, Returns, and Inefficiencies.

Scott E. Atkinson* Kamalini Ramdas† Jonathan W. Williams‡

October 2014

Abstract

Airlines use robust scheduling to mitigate the impact of unforeseeable disruptions on profits. We examine how effectively three common practices – flexibility to swap aircraft, flexibility to reassign gates, and scheduled aircraft downtime – accomplish this goal. We first estimate a multiple-input multiple-outcome production frontier, which defines the attainable set of outcomes from given inputs. We then recover unobserved input costs, and calculate how expenditure on inputs affects outcomes and revenues. We find that the per-dollar return from expenditure on gates, or more effective management of existing gate capacity, is three times larger than the per-dollar returns from other inputs. Next, we use the estimated tradeoffs faced by carriers along the frontier to measure the value to carriers of reducing delays. Finally, we calculate the improvement in carriers' outcomes and profits if their operational inefficiencies are eliminated. On average, we estimate that operational inefficiencies cost carriers about $1.7 billion in revenue annually.

Keywords: Airline Performance, Robust Scheduling, Delays, Stochastic Production Frontiers, Directional Distance Functions

1 Introduction

The increasing frequency and severity of flight delays directly impact passengers and airline profitability. In 2007 alone, the costs of flight delays to the US economy were estimated to be over $41 billion (Schumer and Maloney 2008), with airlines incurring nearly half of these costs in the form of additional operating expenses. For this reason, airlines devote substantial resources to making their operations more robust to disruptions (Arguello et al. 1997, Lan et al. 2006, Rosenberger et al. 2002). Clausen et al. (2010) distinguish between two types of robustness. Recovery robustness is achieved by designing schedules that facilitate fast recovery in the event of a disruption, through actions such as rerouting aircraft and crews or reassigning gates. In contrast, absorption robustness is achieved by providing slack in aircraft schedules to protect against disruptions.

*Department of Economics, University of Georgia, atknsn@uga.edu.
†Management Science and Operations, London Business School, kramdas@london.edu.
‡Department of Economics, University of Georgia, jonwms@uga.edu.

To best deploy robust-scheduling practices, operations managers require information on the costs and returns associated with each form of robustness. The direct costs associated with increasing robustness are known by airlines and readily observable, but the full opportunity costs are far less clear (Smith and Johnson, 2006). For example, increasing flexibility to swap aircraft and crew by having more aircraft of the same type scheduled to depart from an airport in close proximity to one another can mitigate the effect of schedule disruptions. While this action incurs measurable direct costs, the airline must also account for the full opportunity cost of foregone revenues from not deploying aircraft to depart at other times of the day or from other airports in the airline's network. Measuring the returns from robust-scheduling practices is also complicated, since multiple outcomes that drive airlines' profits – loadfactor, cargo, and delays of various lengths – are impacted simultaneously. Given the crucial role of robust-scheduling practices in airlines' operations, the objective of this paper is to develop and implement an empirical methodology that measures their impact on profitability.

We contribute in several ways to the rapidly-growing empirical literature in operations management and economics that analyzes different metrics of airline performance and competition.[1] We first characterize efficient operating practices in the airline industry by estimating a multiple-input, multiple-outcome production frontier, which defines the attainable set of outcomes from a given level of robust-scheduling inputs. We use these estimates, together with detailed data on gate prices, to recover input costs that are unobserved. Next, using data on passenger fares and cargo prices, we estimate how expenditure on these inputs affects the attainable set of outcomes and revenues, which allows us to identify ways that carriers can most profitably target improvements. To our knowledge, we are the first to empirically measure the effectiveness of airline recovery robustness strategies, complementing the insightful OR research on this topic (e.g., see Bratu and Barnhart 2006, Lan et al. 2006, Jiang and Barnhart 2009).[2] We also use our estimates of the tradeoffs faced by carriers along the frontier, together with data on passenger fares, to identify the value of reducing delays of varying lengths. Finally, we calculate the potential improvements in firms' outcomes and profits if they eliminate inefficiencies in their operations.

Airline service can be viewed as a technology that converts inputs such as aircraft and gates into a variety of outcomes, eventually impacting revenues and costs (Cachon and Terwiesch 2011, Atkinson

[1]See Brueckner 2002, Mazzeo 2003, Mayer and Sinai 2003, Brueckner 2005, Rupp et al. 2005, Rupp and Holmes 2006, Shumsky 2006, Rupp 2009, Deshpande and Arikan 2011, Li and Netessine 2011, Li and Netessine 2014, Cannon et al. 2012, and Arikan et al. 2013.

[2]Rupp et al. (2005) study airlines' recovery operations following airport closures. In contrast we examine plans put in place in advance of future disruptions.

and Cornwell 1993, 1994). The multiple outcomes, of which some are desirable and may increase revenue (e.g., loadfactor), and others are undesirable and increase costs (e.g., delays), complicate characterizing efficient operating practices. To characterize efficient operating practices in the least-restrictive manner, we use an output-oriented directional distance function approach (Atkinson and Primont 2002, Atkinson and Dorfman 2005, Färe et al. 2005), which is a generalization of the well-established stochastic production frontier estimation method (Greene 2005, Lieberman and Dhawan 2005, Kumbhakar and Lovell 2000). The estimates of the output-oriented directional distance function characterize the attainable set of outcomes from any given level of inputs in a specific operating environment, i.e., a particular combination of airport, time of day, and aircraft type, for both low-cost and legacy carriers. We also include controls for sources of endogeneity and for other factors, both observed and unobserved, that impact a carrier's ability to convert inputs into outcomes. Subject to these controls, we model the impact of three robust-scheduling inputs (flexibility to swap aircraft, flexibility to reassign gates, and scheduled aircraft downtime) on two "good" outcomes (loadfactor and cargo on passenger aircraft) and three "bad" outcomes (frequencies of short, intermediate, and long delays). In Section 3, we provide a detailed discussion of the numerous advantages of this recently-developed econometric approach.

To understand how robust-scheduling practices impact profitability, it is necessary to know the costs associated with each form of robustness. From our estimates of the directional distance function, we can infer how a carrier can trade off between inputs while keeping outcomes constant.[3] If a carrier had sought to obtain its observed levels of outcomes at least cost, which requires the per-dollar return for each form of robustness to be equal, then our estimates along with data on an input price can be used to infer how the carrier perceives the costs of all other inputs.[4] Using unique data on the price of boarding gates, we calculate both the cost of flexibility to swap aircraft and crew, and scheduled aircraft downtime. By our estimates, the cost of downtime for low-cost carriers is nearly twice that of their legacy competitors, consistent with legacy carriers scheduling more downtime to increase connectivity within their networks. Also, the cost of flexibility for low-cost carriers is 75% less than that for legacy carriers, consistent with low-cost carriers operating more homogenous fleets on high-density point-to-point networks. Thus our cost estimates rationalize differences in carriers' scheduling patterns, providing validation for our econometric methodology.

Besides knowing the cost of each form of robustness, to gauge how these scheduling practices

[3]This is analogous to the marginal rate of technical substitution between inputs (i.e., slope of an isoquant) for a single-output production technology.

[4]This is similar to the calculation in Olivares et al. 2008.

impact profits, airlines must also understand how changing robustness can improve outcomes and increase revenues. For example, consider Delta's "operation clockwork" experiment in 2005 (Hart and Maynard 2005). By "flattening" the schedule of departures in its Atlanta hub, Delta hoped to increase revenues through a steadier flow of connections and lower the frequency of delays of all types, while simultaneously reducing expenditures on airport facilities and aircraft by eliminating spikes in departures and lessening aircraft downtime, respectively. Yet, the magnitude of the impact on revenues and costs was difficult to predict. Our directional distance function estimates are helpful here, as they enable us to provide a precise measure of the improvement in each outcome from each form of robustness. We find that each form of robustness increases passenger and cargo revenues and reduces delays, and the gains are substantially larger for low-cost carriers. Interestingly, the largest per-dollar returns are from expenditure on gates, or from smoothing departures across the day for a given number of gates to maintain a more consistent ratio of gates to departures, which is consistent with Delta's strategy of "flattening." Importantly, the literature to date has paid little attention to managing gate capacity. In fact, in a review of OR applications in air transport, Barnhart et al. (2003) note that gates are "not covered at all or are touched on peripherally." Our empirical findings provide strong evidence that improving recovery robustness by better managing gates is an important avenue for future modelling that will enhance airline profitability.

Aside from increasing expenditure on inputs, carriers may improve outcomes along any one dimension by accepting lower performance along other dimensions (Porter 1996, Lapre and Scudder 2004). Our estimates of the frontier reveal the tradeoffs faced by carriers, and how they perceive the value of each outcome when these estimates are combined with data on passenger fares. We infer that low-cost and legacy carriers are willing to sacrifice \$74.33 and \$68.45, respectively, in per-flight revenue for a 1% reduction in the probability of a flight being delayed over 15 minutes. For example, for Southwest Airlines in 2007 this amounted to about 0.8% of total revenues per flight, or \$78 million across their entire network (over 1.1 million flights).[5] Both types of carriers value a 1% reduction in the probability of a flight being delayed over 180 minutes at more than 4 times as much. This reinforces the findings of Ramdas et al. (2013), that long delays have a substantial impact on a carrier's market value, and rationalizes the strategies that some carriers, such as Southwest, have adopted to minimize their frequency at the cost of a greater frequency of shorter delays. The estimates of the value placed on delays by carriers can also guide regulatory policy, which seeks to balance the benefits to passengers of fewer delays against the costs to carriers

[5] Shumsky (1993) demonstrates the responsiveness of carriers to the On-Time Disclosure rule, which is consistent with our finding that carriers place significant value on improving delays.

of limiting the severity and frequency of delays.

Unlike carriers operating on the production frontier, inefficient carriers operating inside this frontier can increase profits with no additional expenditure on inputs. Our estimates measure how much an inefficient carrier must improve each outcome, for a given level of inputs, to reach the frontier. We find that even relatively small inefficiencies can result in substantial forgone revenues. For example, Delta achieved only 91% of the efficient level of loadfactor, conditional on its observed levels of inputs and other outcomes, resulting in over $500 million in lost passenger revenue in 2007. Our efficiency estimates can also explain the substantial improvement realized by some carriers both during and after our period of study. We find that much of Southwest's highly-acclaimed success hails from its strategic choice of operating environments that are highly concentrated and uncongested. Interestingly, once we control for the environments Southwest operated in during 1997-2009, we find that Southwest was relatively inefficient with respect to loadfactor. That is, the loadfactors achieved by Southwest in its highly-desirable operating environments were substantially lower than best practice, indicating substantial room for improvement along this dimension without changes to scheduling practices or worsening of other outcomes. Consistent with a reduction of this inefficiency, Southwest increased loadfactor across its network by over 10% and its revenues by over 50% from 2007 to 2013 with no increase in delays of any type.

The remainder of this paper is organized as follows. We describe our data in Section 2, and the details of the distance-function approach and structural econometric model in Section 3. We present our results in Section 4 and our conclusions in Section 5.

2 Data

We use five data sources from the Bureau of Transportation Statistics (BTS): the on-time performance database, B43 form on aircraft inventories, T100 segment database, P-1.2 Financial Schedule, and Origin and Destination Survey. Other data we use in our analysis include information on airlines' lease payments to airports from FAA Form 127, and a survey on carrier-specific access to airport facilities conducted jointly with North America's largest airport trade organization, the ACI-NA (Williams 2014). We discuss how each source of data is used below.

2.1 On-Time Performance Data

The BTS' on-time performance database allows us to compare carriers' operational performance in the same environment (i.e., airport, day of week, and departure time block combination) while operating similar equipment (i.e., aircraft type) at each point in time (i.e., year and month). Let $i = 1, ..., I$ represent a particular environment, i.e., a unique combination of airport (e.g., Atlanta Hartsfield or ATL), day of week (e.g., Monday), and scheduled departure time block (e.g., 10am-2pm EST), $j = 1, ..., J$ a carrier (e.g., Delta), $k = 1, ..., K$ an aircraft type (e.g., Boeing 737), and $t = 1, ..., T$ a year and month (e.g., January 2009). Our unit of analysis is the 4-tuple, (i, j, k, t). We group flights into one of 4 departure time blocks; midnight-10am (morning), 10am-2pm (midday), 2pm-6pm (afternoon), and 6pm-midnight (evening).[6] For aircraft types, we proxy using an aircraft's crew-rating.[7] To focus on domestic service, we eliminate from our sample all aircraft that appear to be involved in international service.[8] Using these data, we calculate three different bad outcomes, which are quantiles of the arrival delays distribution for each 4-tuple, (i, j, k, t):

$\mathbf{P}(> \mathbf{15})_{ijkt}$ is the proportion of flights flown in environment i by carrier j using equipment k in time period t that arrived 15 or more minutes late. $\mathbf{P}(> \mathbf{60})_{ijkt}$ and $\mathbf{P}(> \mathbf{180})_{ijkt}$ are similarly defined. The variable $\mathbf{P}(> \mathbf{15})_{ijkt}$ is identical to the DOT's measure of on-time performance, while the other measures capture the shape of the tail of the delays distribution.[9]

We use the detail of the on-time performance database to track each aircraft's routing, and calculate two input variables:

$\mathbf{Downtime}_{ijkt}$ is the difference between the average number of minutes between an aircraft's scheduled arrival and its next scheduled departure, and the minimum turn time (Lan et al. 2006). The average is taken over all departures in 4-tuple, (i, j, k, t), and the minimum turn time is over all departures in the 3-tuple, (i, k, t). Since scheduled downtime is by definition determined at the schedule planning stage, it is a measure of absorption robustness.

$\mathbf{Flex}_{ijkt}$ is the total number of aircraft of type k belonging to carrier j that are scheduled to depart from environment i in time period t. An increased number of aircraft of the same aircraft

[6] Our results are robust to variations either way in the length of the time blocks. We settled on this definition because of the inherent tradeoff in defining the length of these time blocks. Shorter windows may not accurately capture substitution opportunities, i.e., flexibility, since the cost of swapping an aircraft depends on the remaining sequence of flights with the same aircraft type over the day at the airport. Longer time blocks miss important variation over the day in utilization of airport facilities.

[7] In all but a few cases, a crew-rating corresponds to a single aircraft type, so we use the two terms interchangeably.

[8] To do this, we first merge, by tail number, the BTS on-time performance database with aircraft information in the BTS B43 form. We then remove aircraft equipped for international travel, i.e., flying over large water bodies. Next, we remove aircraft that appear to have flown an international segment, identified via gaps in the aircraft's routing. Finally, we remove obvious reporting errors such as aircraft that report an infeasible flight sequence.

[9] Cancellations are treated as delays of longer than 180 minutes and included in each of the measures above. We replicated the entire analysis using net arrival delays, calculated by subtracting the incoming arrival delay for each flight from the arrival delay for the next flight, prior to calculating quantiles of the delays distribution. We also used different quantiles to measure short, intermediate, and long delays. The results are qualitatively similar in each instance.

type give a carrier additional flexibility to substitute aircraft and crews when needed. Scheduled flexibility is our first measure of recovery robustness. It is conceptually similar to the "station purity" measure of Smith and Johnson (2006).[10]

In estimating a production frontier, one must carefully control for factors that may limit the attainable set of outcomes. In the airline industry, the operating environment varies widely even within an airport. For example, Atlanta, a large hub airport, is busy during certain periods when carriers concentrate flights to facilitate connections. During these periods, airport capacity is highly utilized and the attainable level of outcomes is lower for any level of inputs employed (Caulkins et al. 1993). In Section 3.2, we discuss how we control for time-invariant factors unique to each operating environment i, carrier j and aircraft type k, and also the impact of industry-wide shocks like weather or the large dampening effect on demand of 9/11, which may differ by region. In addition to these controls, we also include time-varying factors that impact production by including measures of market concentration and congestion:

HHI$_{it}$ is the Herfindahl-Hirschman index for scheduled departures in the departure time-block, day-of-week, and airport that characterize environment i in time period t. Carriers internalize the effect of delays when airport concentration, HHI_{it}, is high (Mayer and Sinai 2003, Brueckner 2002, and Brueckner 2005). Therefore we expect the frontier to be characterized by a lower level of delays in more concentrated environments.

Congestion$_{it}$ is the total number of flights scheduled to depart in the time-block, day-of-week, and airport that characterize environment i in time period t, divided by the number of gates at the airport.[11]

After constructing these variables, we limit our sample to 4-tuples (i, j, k, t) for which at least 10 flights were used to construct each of the variables described above. Essentially, we consider a carrier j to serve a particular environment i – characterized by a specific airport, departure time block and day of week – with aircraft type k in year-month t, if there were at least ten departures in (i, j, k, t).[12]

[10] Flights scheduled to depart at a similar time and using the same aircraft type may not always be able to swap aircraft or crew. Scheduled maintenance and limitations on crew hours make our measure an upper bound on flexibility. Our empirical approach, which exploits variation within an environment (i) to identify model parameters, removes the average effect of any systematic upward bias in flexibility within an environment.

[11] Since fixed effects for an operating environment are included in the analysis, and the number of runways at an airport varies so little, the results do not change if the number of flights is normalized by the number of runways.

[12] This captures over 85% of enplanements since the majority of them occur at large airports with frequent departures. Our results are insensitive to varying this cutoff in either direction.

2.2 T100 Data

Loadfactor has trended upwards over the past decade. Many carriers also carefully manage the number and size of passenger bags, to increase revenue from other cargo, e.g., mail. The BTS T100 domestic-segment database contains detailed information on passenger volumes, cargo, and available seats by airport of origin, carrier, aircraft type, month, and year. For each 4-tuple, (i, j, k, t), we use these data to calculate two good outcomes:

> $\mathbf{LoadFactor}_{ijkt}$ is the proportion of available seats filled by revenue-generating passengers, averaged over all departures in (i, j, k, t).
>
> $\mathbf{Cargo}_{ijkt}$ is the average cargo, i.e., sum of weights of freight and mail transported, in tens of thousands of pounds per scheduled departure, averaged over all departures in (i, j, k, t).

Note that $LoadFactor$ and $Cargo$ are obtained from the T100 dataset, which is not specific as to the day of week or departure time block in which each aircraft departs.

2.3 DB1B Survey and P-1.2 Financial Schedule

To study the revenue impact of varying robust-scheduling inputs, we collect data from two sources. The first source is the BTS' Airline Origin and Destination Survey (DB1B), which consists of a quarterly 10% sample of itineraries sold by domestic carriers. Each observation or itinerary in the data includes the origin and final destination of the passenger, identity of the carrier providing the service, fare paid, as well as detailed information on any connections made en route to the final destination.[13] The second source is the BTS' Schedule P-1.2 database, which contains quarterly information on sources of revenue and costs. For each 4-tuple, (i, j, k, t), we use these data to calculate prices for the two good outcomes:

> $\mathbf{p}_{ijt}^{pax}$ is the average price per mile paid by revenue-generating passengers, averaged over all departures in (i, j, k, t). This is calculated as the total fare paid divided by the distance flown, accounting for connections and whether the itinerary is roundtrip.
>
> $\mathbf{p}_{jt}^{cargo}$ is the average price paid to transport one ton of cargo one mile, averaged over all departures in (i, j, k, t). This is calculated by dividing total revenues from cargo by the product of tonnage and miles flown, or ton-miles.

Note that $\mathbf{p}_{ijt}^{pax}$ does not vary by aircraft type since this is not reported in the DB1B data, and $\mathbf{p}_{jt}^{cargo}$ varies only by carrier and time since cargo revenues are aggregated to the carrier level and

[13] The DB1B does not report revenue generated by other sources like checked bags, which are becoming an increasingly important source of revenue for carriers.

reported quarterly. This lack of detailed revenue information precludes its use in the regression analysis. If revenue were included, we would have to substantially change the unit of observation to a much more aggregate level, losing many of the insights we provide regarding tradeoffs and costs. Alternatively, we would introduce substantial amounts of measurement error by arbitrarily disaggregating revenues to the (i,j,k,t) level.

2.4 ACI-NA Airport Survey and FAA Form 127

Access to airport facilities has been shown to be critical to many market outcomes including delays, fares, and entry (Ciliberto and Williams 2010, Snider and Williams 2014). Our data on boarding gates come from a recent survey on access to airport facilities conducted jointly with the ACI-NA (Williams 2014), which reports information on carriers' access to boarding gates during 1997, 2001, 2007, 2008 and 2009. From these data, for each 3-tuple, (i, j, t) we construct two variables:

$\mathbf{Gates}_{ijt}$ is the total number of gates leased to carrier j at the airport pertaining to environment i in time period t. Note that within any i, there is no variation in $Gates_{ijt}$ across either day of week or departure time block, because gates are typically leased for at least a year at a time.

$\mathbf{GatesSchedDep}_{ijt}$ is $Gates_{ijt}$ divided by the number of scheduled departures pertaining to (i, j, t). This is our second measure of recovery robustness. Since we do not know the operational limitations of the gates leased to each carrier, such as ability to handle particular aircraft, $GatesSchedDep_{ijt}$ does not vary with k.

We also collect unique detailed data on costs associated with leasing airport facilities from the FAA's Form 127 for 2007-2009.[14] For each airport, we use these data to construct one additional variable:

$\mathbf{p}_{ijt}^{gate}$ is the total monthly fees paid by carrier j to airport i for gate leases and apron fees divided by the number of gates leased by the carrier at the airport.[15]

2.5 Final Sample and Descriptive Statistics

After merging the datasets described above by unique 4-tuples (i, j, k, t), we have 172,380 observations, drawn from 81 airports, all of which are in the top 150 airports by passenger enplanements in the years in which we have survey data on boarding gates: 1997, 2001, 2007, 2008, and 2009. The carriers in our final sample are American (AA), JetBlue (B6), Continental (CO), Delta (DL),

[14] Available at: http://cats.airports.faa.gov/Reports/reports.cfm

[15] The apron is the area of an airport where aircraft are parked, unloaded or loaded, refueled, and boarded.

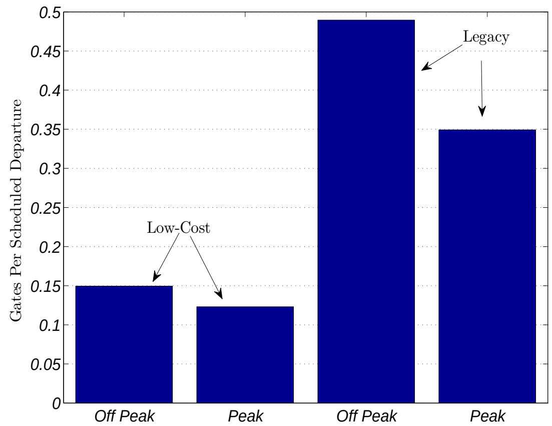


Figure 1: *Gates Per Scheduled Departure by Carrier Type and Time of Day*

Frontier (F9), AirTran (FL), Northwest (NW), United (UA), US Air (US), and Southwest (WN).[16] Tables 1a and 1b contain summary statistics, by carrier and year, for outcomes and inputs, respectively.

We find starker differences among the carriers in the levels of inputs than in outcomes. On average, low-cost carriers schedule half the downtime of their legacy competitors, with Southwest responsible for the majority of this difference. Other low-cost carriers, which have more network overlap with their legacy competitors and operate in the same large hubs (e.g., AirTran in Atlanta Hartsfield), tend to schedule downtimes more similar to those of legacy carriers. Regarding flexibility, low-cost carriers tend to schedule more aircraft of the same type to depart from an airport in the same time block and day of week, despite operating at airports that are smaller on average.

We find that low-cost carriers tend to use gates much more intensively, i.e., they have a lower ratio of gates to scheduled departures. Yet the carrier-specific averages reported in Table 1b mask important variation in utilization. Figure 1 gives the mean of $GatesSchedDep_{ijt}$ by carrier type and time of day. Peak is defined as the middle two 4-hour time blocks, while off peak includes the first and last time block. Legacy carriers use gates less intensively at all times, but experience a much greater change in gate utilization during peak hours.

The difference in the average per-month price paid by legacy and low-cost carriers to lease a gate, p^{gates}, is almost entirely due to low-cost carriers disproportionately serving the least expensive airports, e.g., Phoenix and Dallas Love Field, while legacy carriers disproportionately serve the

[16] We drop Alaska (AS) because its regional nature and limited range of service match up poorly with the responses to the gates survey data, and America West (HP) because its merger with US Airways limits its data to the first two sample years, and it dramatically reduced its destinations served following a bankruptcy in the mid-1990s.

most expensive airports, e.g., San Francisco, Miami, Newark, and LaGuardia. Table 1b also shows that low-cost carriers, particularly JetBlue and Southwest, tend to concentrate their operations in less congested yet more concentrated airports, which gives them tremendous flexibility to shuffle departures across gates to minimize delays.

Tables 1a and 1b also provide summary statistics by year. Interestingly, with the exception of $LoadFactor$, there are no consistent trends in outcomes over time. $LoadFactor$ has generally increased from 1997 to 2007, and then leveled out, while delays of all types appear to have reached a peak in 2007 and subsided over the next two years. The descriptive statistics provide little evidence of any shift in the intensity with which different operational inputs have been used over time. However, airport concentration has risen substantially during this period.

Collectively, the carrier-specific summary statistics convey the importance of controlling for all the factors that affect firms' decisions about the levels of different operational inputs to employ in a particular environment. This motivates our fixed-effects econometric approach in Section 3.2.

3 Econometric Specification of Directional Distance Function

The productivity literature in economics has recently developed theoretical and econometric tools to characterize efficient operating practices in complex environments where firms convert multiple inputs ($Downtime$, $Flex$, and $GatesSchedDep$) into multiple good outcomes ($LoadFactor$ and $Cargo$) and bad outcomes ($P(>15)$, $P(>60)$, and $P(>180)$). Specifically, an output-oriented directional distance function characterizes the attainable set of outcomes for any given level of inputs.[17]

3.1 Theoretical Properties of Directional Distance Function

Consider an airline production technology by which firms combine multiple operational inputs, $\mathbf{x} = (x_1, \ldots, x_N) \in R_+^N$, to produce multiple good outcomes, $\mathbf{y} = (y_1, \ldots, y_M) \in R_+^M$, and bad outcomes, $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_L) \in R_+^L$. The firm's production technology, $P(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{y}})$, can be written as

$$P(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{y}}) = \{(x, y, \tilde{y}) : x \text{ can produce } y \text{ and } \tilde{y}\}, \tag{1}$$

where $P(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{y}})$ consists of all feasible input and outcome vectors.

[17] Note that the production frontier does not characterize the revenue potential associated with a set of inputs. If more disaggregated proprietary data on revenues were available, one could estimate an output-based directional revenue function to characterize the frontier firm that would be a function of the quantity of inputs, output prices, and revenues. The theory for this has been developed in Chambers et al. (2013). Unfortunately, estimation of this model would omit critical output quantities (such as delays) which are of prime importance in our analysis. If disaggregated proprietary data on revenues were available, one could also embed a directional distance function into a profit-maximization framework as in Atkinson et al. (2014).

Following Chambers et al. (1998) and Färe et al. (2005), we define the output-oriented directional distance function as

$$\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) = \beta^* = \sup_{\beta}\{\beta : (\mathbf{y} + \beta\mathbf{g}_y, \tilde{\mathbf{y}} - \beta\mathbf{g}_{\tilde{y}}) \in P(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}})\}. \tag{2}$$

The output-oriented directional distance function, $\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}})$ provides a measure of the increase in good outcomes and decrease in bad outcomes needed to move a carrier to the frontier of the production set, $P(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}})$, for a given level of inputs. Thus, a carrier's *distance* is the scalar multiple of the direction vector, $(\mathbf{g}_y = 1, -\mathbf{g}_{\widetilde{y}} = 1)$, required to reach the production frontier.

As discussed in Chambers et al. (1998), the output-oriented directional distance function, $\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}})$, defined in Equation 2 must satisfy the following properties to serve as a reasonable model of a firm's production technology:

Non-Negativity:

$$\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) \geq 0 \iff (\mathbf{y}, \tilde{\mathbf{y}}) \in P(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}). \tag{P1}$$

Property P1 requires that the output-oriented directional distance function be non-negative, with the most efficient firm having $\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) = 0$. We ensure this property is satisfied via a normalization of the fitted function.

Translation Property:

$$\overrightarrow{D}(\mathbf{x}, \mathbf{y} + \alpha\mathbf{g}_y, \widetilde{\mathbf{y}} - \alpha\mathbf{g}_{\widetilde{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) = \overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) - \alpha. \tag{P2}$$

Property P2 is true by definition of the output-oriented directional distance function, Equation 2. This property implies that increasing $\mathbf{y}$ and decreasing $\tilde{\mathbf{y}}$ by α, multiplied by their respective directions, $(\mathbf{g}_y = 1, -\mathbf{g}_{\widetilde{y}} = 1)$, holding inputs constant, will result in a decrease in the output directional distance by α, and therefore an increase in efficiency equal to α. We ensure that Property P2 is satisfied via the parametric restrictions imposed during estimation and described in Section 3.2.

g-Homogeneity of Degree Minus One:

$$\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \lambda\mathbf{g}_y, -\lambda\mathbf{g}_{\widetilde{y}}) = \lambda^{-1}\overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}), \lambda > 0. \tag{P3}$$

Property P3 also follows directly from Equation 2, since scaling each element of the direction vector, $(\mathbf{g}_y = 1, -\mathbf{g}_{\tilde{y}} = 1)$, by λ will scale the output directional distance by λ^{-1}. Multiplying

$\mathbf{g}_y$ and $\mathbf{g}_{\tilde{y}}$ by λ divides β^* by λ.[18] Property P3 is satisfied once P2 is imposed (via parametric restrictions) and the model is estimated as described in Section 3.2.[19]

Good Outcome Monotonicity:

$$\mathbf{y}' \geq \mathbf{y} \rightarrow \overrightarrow{D}(\mathbf{x}, \mathbf{y}', \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) \leq \overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}). \tag{P4a}$$

and

Bad Outcome Monotonicity

$$\widetilde{\mathbf{y}}' \geq \widetilde{\mathbf{y}} \rightarrow \overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}', \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}) \geq \overrightarrow{D}(\mathbf{x}, \mathbf{y}, \widetilde{\mathbf{y}}, \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}). \tag{P4b}$$

Properties P4a and P4b state that the output-oriented directional distance function is monotonically decreasing in good outcomes, while it is monotonically increasing in bad outcomes. That is, holding all else constant, increasing a good outcome (or decreasing a bad outcome) moves the directional distance function closer to zero, i.e., the carrier becomes more efficient. We test for Properties P4a and P4b in Section 4 and find that they are satisfied, so that we do not need to impose them during estimation.[20]

3.2 Econometric Methodology

The empirical specification of the directional distance function must capture the differences in the operating models of low-cost and legacy carriers, control for sources of endogeneity, and satisfy Properties P1-P4b. To accomplish this, we specify the directional distance function as

$$0 = \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta) + \varepsilon_{ijkt}, \tag{3}$$

where

$$\begin{aligned} & \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta) \\ = \quad & + \sum_{n=1}^{N} (\gamma_x^n + \phi_x^n d_{Leg})\, x_{ijkt}^n + \sum_{m=1}^{M} (\gamma_y^m + \phi_y^m d_{Leg})\, y_{ijkt}^m + \sum_{l=1}^{L} \left(\gamma_{\widetilde{y}}^l + \phi_{\widetilde{y}}^l d_{Leg}\right) \widetilde{y}_{ijkt}^l \\ & + (\gamma_w + \phi_w d_{Leg})\, \mathbf{w}_{it} + \left(\gamma_i^{envir} + \phi_i^{envir} d_{Leg}\right) d_i^{envir} \\ & + \eta_k^{aircraft} d_k^{aircraft} + \eta_t^{yr-mon} d_t^{yr-mon} + \eta_r^{reg-mon} d_r^{reg-mon} + \eta_j^{carrier} d_j^{carrier} \end{aligned} \tag{4}$$

[18] By normalizing all outputs to be unit-free prior to estimation, we make the distance function estimates unit free. Following estimation, we can then convert back to the original units for our analysis.

[19] These parametric restrictions are analogous to imposing linear homogeneity when estimating a cost function (e.g., Equation 2 in Berndt and Wood 1975).

[20] Our empirical specification of the directional distance function ensures the production set is also concave.

and

$$\varepsilon_{ijkt} = \upsilon_{ijkt} - u_{ijkt}. \tag{5}$$

Following Färe et al. (2005), the left-hand side of Equation 4 is set equal to zero, representing frontier production (i.e., fitted *distance* of zero).

The functional form of Equation 4 permits differences in the operating models of low-cost and legacy carriers through the interaction of a legacy indicator, d_{Leg}, with the vector of good outcomes, $\mathbf{y}$, comprised of $LoadFactor_{ijt}$ and $Cargo_{ijt}$, the vector of bad outcomes, $\widetilde{\mathbf{y}}$, comprised of $P(>15)_{ijkt}$, $P(>60)_{ijkt}$, and $P(>180)_{ijkt}$, the vector of inputs, $\mathbf{x}$, comprised of $Flex_{ijkt}$, $Downtime_{ijkt}$, and $GatesSchedDep_{ijt}$, and a vector of time-varying factors that impact a carrier's ability to convert inputs to outcomes, $\mathbf{w}$, comprised of $Congestion_{it}$ and HHI_{it}.

To capture as many sources of endogeneity as possible, including unobserved determinants of carriers' input choices and outcomes, as well as unobserved inputs, by including fixed effects for an aircraft's crew-rating ($d_k^{aircraft}$), each year-month (d_t^{yr-mon}), each region-month ($d_r^{reg-mon}$), each carrier ($d_r^{carrier}$), and each environment for each carrier type (interaction of d_{Leg} and d_i^{envir} indicators). The aircraft crew-rating fixed effects control for larger aircraft requiring greater downtime, as well as unobserved inputs like the number of flight attendants and pilots, which are mandated by the FAA. The year-month fixed effects control for any temporal industry-wide shocks that impact productivity, like the events of September 11^{th} and its lingering effects (e.g., additional security procedures and diminished demand). The region-month fixed effects control for differential weather patterns across regions. The carrier fixed effects control for time-invariant carrier-specific factors and the average effect of time-varying carrier-specific factors, such as wage differentials and carrier size. The carrier-type and environment-specific fixed effects control for time-invariant unobservables specific to environment i for low-cost and legacy carriers, respectively. Collectively, in addition to removing sources of endogeneity, these controls allow the frontier to be characterized by lower levels of good outcomes and higher levels of bad outcomes from the same level of inputs, in more difficult environments.

Following Atkinson and Primont (2002) and Färe et al. (2005), we decompose the error term, ε_{ijkt}, into a one-sided component, $u_{ijkt} > 0$, that captures unmeasured time-varying differences in carriers' managerial ability or operational difficulties unique to specific carriers, and a two-sided mean-zero idiosyncratic component, υ_{ijkt}, that allows for random and transitory shocks impacting carriers' ability to transform inputs to outcomes, such as poor weather. We make no parametric distributional assumptions for either error.

Like Färe et al. (2005), our parameter estimates are chosen to satisfy

$$\{\widehat{\gamma},\widehat{\phi},\widehat{\eta}\} = \underset{\{\gamma,\phi,\eta\}}{\arg\min} \sum_{i,j,k,t} \left(0 - \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta)\right)^2,$$

subject to

i) $\overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta) \geq 0,$ for all (i,j,k,t)

ii) $\sum_{m=1}^{M} \gamma_y^m g_y - \sum_{l=1}^{L} \gamma_{\widetilde{y}}^l g_{\widetilde{y}} = -1$

iii) $\sum_{m=1}^{M} \left(\gamma_y^m + \phi_y^m\right) g_y - \sum_{l=1}^{L} \left(\gamma_{\widetilde{y}}^l + \phi_{\widetilde{y}}^l\right) g_{\widetilde{y}} = -1$

iv) $\frac{\partial \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta)}{\partial \widetilde{\mathbf{y}}_m} \geq 0.$

v) $\frac{\partial \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \gamma, \phi, \eta)}{\partial \mathbf{y}_m} \leq 0$

The estimated parameters, $\{\widehat{\gamma},\widehat{\phi},\widehat{\eta}\}$, make the carriers look as efficient (i.e., fitted *distance* as close to zero) as possible while satisfying properties **i**)-**v**).[21] Constraint **i**) ensures Property P1, non-negativity, is satisfied. We ensure that **i**) holds via a normalization of the estimated distance function, increasing all fitted distances by the absolute value of the most negative fitted distance. Constraints **ii**) and **iii**) ensure that the translation property and g-homogeneity are satisfied for legacy and low-cost carriers, respectively. These constraints are imposed parametrically during estimation. Constraints **iv**) and **v**) ensure Properties P4a and P4b, monotonicity, hold. We do not impose **iv**) or **iv**), since we find the unconstrained solution satisfies monotonicity.[22]

The stochastic directional distance approach is very similar to well-studied non-stochastic linear-programming solutions, but also different in important ways. Specifically, if our objective function were an absolute rather than quadratic loss function, the problem would reduce to a linear program similar to the non-stochastic optimization specification of Färe et al. (2005). For our purposes, a stochastic approach is preferable for two main reasons. First, the non-stochastic problem does not allow for random shocks to productivity, the υ_{ijkt} term in Equation 5, which is important for our problem. Airlines face a number of events outside their control (e.g., 9/11, airport-specific operations, severe weather, etc.), that should not be reflected in any calculation of productivity. Thus,

[21] Rather than estimate over 4,000 unique coefficients corresponding to the carrier-type and environment specific fixed effects, we follow an algebraically equivalent procedure and difference the distance function prior to estimating the remaining parameters.

[22] Before estimating the model, we normalize inputs and outputs to make the fitted distance function unit free. We then convert back to the original units to ease interpretation of the results.

our calculation of technical inefficiency is based solely on u_{ijkt}. Second, we are particularly interested in comparing the statistical significance of the returns to various robust-scheduling practices, which requires that we employ a stochastic model.

3.3 Visualization of Directional Distance Function Approach

Below, we use a simple example to illustrate the mechanics of the directional distance function approach and how it characterizes a frontier corresponding to efficient production. Assume that there is one good outcome, y, and one bad outcome, $\tilde{y}$, and that our data consist of carriers (of the same type) that employ the same amount of an input, x. In this case, the econometric objective function is

$$\{\widehat{\varphi_0},\widehat{\varphi_1},\widehat{\varphi_2}\} = \underset{\varphi_0,\varphi_1,\varphi_2}{\arg\min} \sum_r \left(0 - \varphi_0 - \varphi_1 \widetilde{y}_r - \varphi_2 y_r\right)^2, \tag{6}$$

subject to

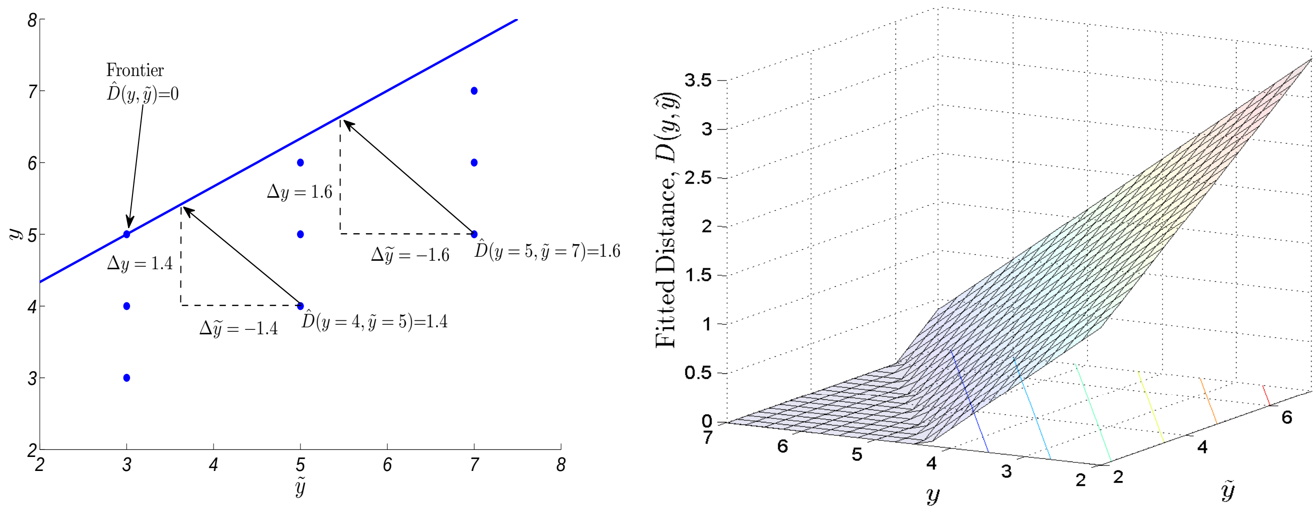
i) $\varphi_0 + \varphi_1 \widetilde{y}_r + \varphi_2 y_r \geq 0$, for all r

ii-iii) $\varphi_2 - \varphi_1 = -1$

iv-v) $\varphi_1 \geq 0$ and $\varphi_2 \leq 0$

Constraint i) ensures non-negativity, ii) and iii) reduce to one constraint (as there is only one carrier type), which ensures that the translation property and g-homogeneity are satisfied. The monotonicity restrictions are iv)-v). The objective is to identify parameter values that make the carriers look as efficient as possible – i.e., minimize fitted distance – while satisfying the constraints that impose properties necessary to have a well-defined production technology. The most efficient carriers have a fitted distance of zero.

The example data, nine ordered pairs indexed by r, are plotted in Figure 2(a). Figure 2(b) plots the fitted distance function, $\widehat{\varphi_0} - \widehat{\varphi_1}\widetilde{y} - \widehat{\varphi_2} y$ (normalized to be non-zero) through the sample points, along with the level sets of the plane. Note that the fitted distance is strictly increasing (decreasing) in the bad (good) outcome (i.e., monotonicity), and all fitted distances are strictly positive, while satisfying $\varphi_2 - \varphi_1 = -1$. The level sets of the distance function are useful, because the level set corresponding to a fitted distance of zero characterizes the tradeoffs between outcomes faced by an efficient carrier. This tradeoffs can also be seen as a traditional production frontier, plotted in Figure 2(a), by re-arranging the fitted distance function, as $y = -\frac{\widehat{\varphi_0}}{\widehat{\varphi_2}} - \frac{\widehat{\varphi_1}}{\widehat{\varphi_2}}\widetilde{y}$. The tradeoff is then the slope



(a) Frontier Interpretation

(b) Distance Function Interpretation

Figure 2: Visualization of Directional Distance Function Approach

of this frontier, $-\frac{\widehat{\varphi_1}}{\widehat{\varphi_2}}$, which would also be obtained by implicitly differentiating the fitted distance function directly. From the definition of the directional distance function, Equation 2, the fitted distances can be visualized as the scalar multiple of the direction vector, $(g_y = 1, -g_{\tilde{y}} = 1)$, required to reach the production frontier by moving horizontally and vertically as defined in Equation 2 and indicated in Figure 2(a) with the dotted lines.

One can also easily visualize our measures of technical efficiency, which we discuss in Section 4.5, using Figure 2(a). For example, our measure of technical efficiency for the good outcome is the maximum additional amount of the good outcome that is feasible given the levels of other outcomes and inputs, i.e., the vertical distance from one of the points in Figure 2(a) to the frontier. The measure of technical efficiency for the bad outcome is the horizontal distance to the frontier.

4 Results

Figure 3 provides a guide for the discussion below. We first discuss the coefficient estimates of the directional distance function in Section 4.1. In Section 4.2 we describe how these estimates, together with gate-price data, enable us to obtain the per-unit cost of the other inputs. Next, in Section 4.3, we demonstrate how to use the coefficient estimates from Section 4.1 and cost estimates from Section 4.2 to obtain the per-unit and per-dollar returns for each outcome from each input. We then combine these estimates of returns with data on ticket and cargo prices,

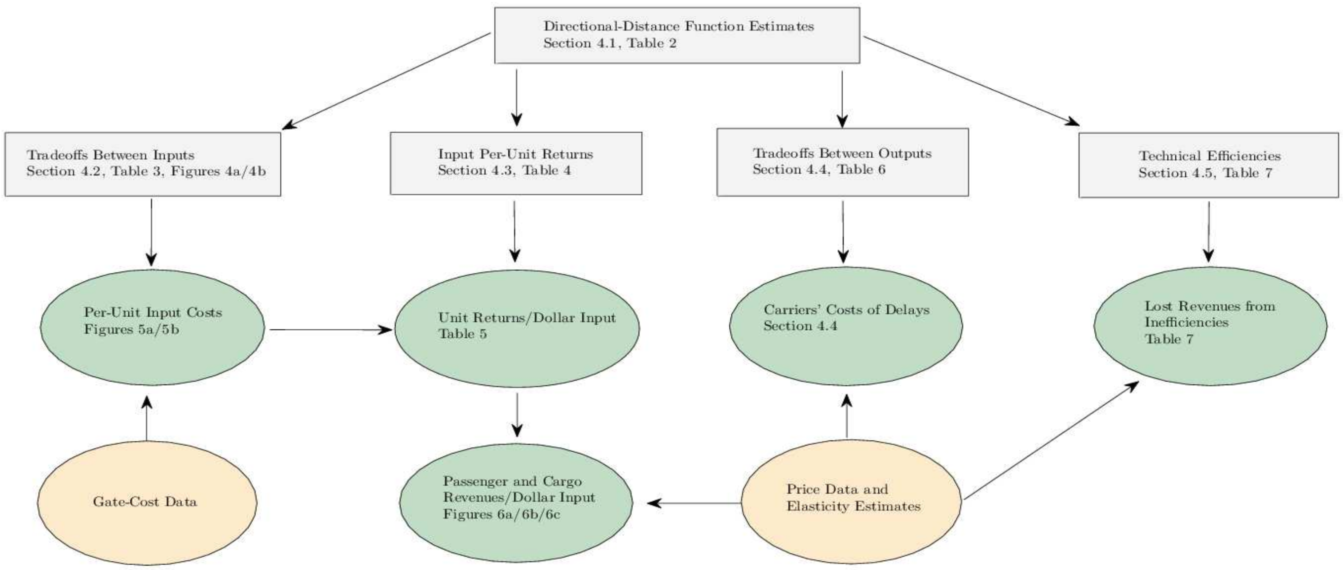


Figure 3: *Overview of Results Discussion*

and estimates of passenger and cargo own-price demand elasticities, to estimate the passenger and cargo revenues earned per dollar spent on each robust-scheduling input. In section 4.4 we use the coefficient estimates from Section 4.1 to analyze tradeoffs between outcomes, and then combine these estimated tradeoffs with data on ticket prices to uncover the value to carriers from reducing delays of different lengths. In Section 4.5 we estimate carriers' technical efficiencies and use these to identify avenues for operational improvement and predict the increase in profitability from these improvements.

4.1 Estimates of Directional Distance Function Coefficients

Table 2 contains the estimates of Equation 4.[23] Columns 1 and 2 of Table 2 report the coefficient estimates for low-cost carriers and the interaction of each variable with the legacy-carrier indicator, d_{Leg}, respectively. All coefficients in Table 2 are statistically significant, except for two interactions with the legacy-carrier indicator. This lack of significance for the congestion interaction is not surprising, as congestion should not affect carriers differentially. Thus the grouping of carriers into low-cost and legacy types separates out important sources of systematic heterogeneity among the carriers while using a computationally feasible number of parameters.

The coefficient estimates in Table 2 have a straightforward interpretation, as discussed in Section 3.2. For example, the positive coefficient on *GatesSchedDep* implies that holding constant other

[23] The t-statistics are calculated using asymptotic heteroscedasticity-robust White standard errors, allowing for arbitrary correlation within each origin/day-of-week/time-of-day combination by clustering on environment i.

inputs and outcomes, an increase in *GatesSchedDep* corresponds to a greater distance from the frontier. In other words, a carrier is regarded as less efficient if it achieves the same outcomes while using gates less intensively. The coefficients for other inputs can be interpreted similarly. For good (bad) outcomes, one expects a negative (positive) coefficient such that an increase in the outcome holding all other outcomes and inputs constant results in a lesser (greater) distance from the frontier. Our estimates yield the expected sign for each outcome. Thus, Properties P4a and P4b are satisfied despite not being imposed during estimation.

The positive sign on HHI is consistent with prior work (see Mayer and Sinai 2003), which suggests that the frequency of delays should be inversely related to concentration, since carriers have a greater incentive to internalize the effect of their actions in more-concentrated environments. Thus the minimum attainable level of delays is lower in more concentrated environments.

4.2 Estimates of Input Costs

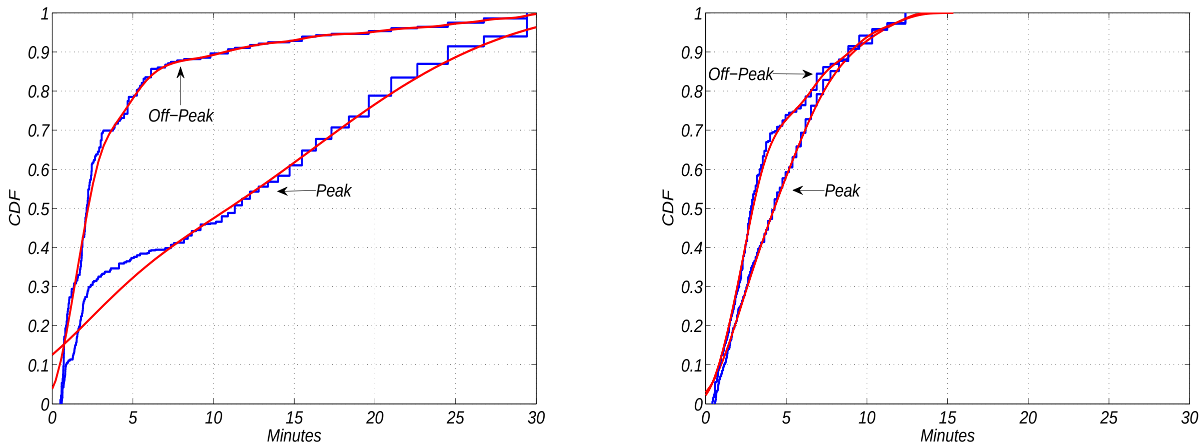
The use of robust-scheduling inputs generates both direct and indirect costs. For example, an additional minute of scheduled downtime directly increases costs due to lost revenue, but indirectly reduces costs due to increased *connectivity*. Since legacy carriers operate hub-and-spoke networks, we expect this benefit to substantially lower their costs of downtime. Similarly, altering aircraft routings to increase flexibility by ensuring that more aircraft of the same type depart at similar times can affect connection opportunities and the desirability of flight departure times.

To estimate the costs of robust-scheduling inputs, we first estimate the tradeoffs between these inputs, which we combine with gate price data to infer the costs of the other inputs. Since a frontier firm has a fitted distance of zero, $\overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \widehat{\gamma}, \widehat{\phi}, \widehat{\eta}) = 0$, by the implicit function theorem, the rate at which such a carrier can trade off between inputs x^n and $x^{n'}$, while holding outcomes constant, is

$$\frac{dx^n}{dx^{n'}} = -\left(\frac{\frac{\partial \overrightarrow{D}}{\partial x^{n'}}}{\frac{\partial \overrightarrow{D}}{\partial x^n}}\right). \tag{7}$$

Table 3 contains our estimates of these tradeoffs, which are themselves of interest. For example, we find that, at the margin, a legacy carrier can reduce scheduled downtime on each flight by about 3.13 minutes and keep outcomes constant if one more flight of the same aircraft type is scheduled to depart in the same time block (i.e., if flexibility is increased by one unit), in contrast to a reduction of only 1.06 minutes for low-cost carriers.

Similarly, Figures 4(a) and 4(b) report the *cdfs* of the number of minutes by which scheduled



(a) Legacy

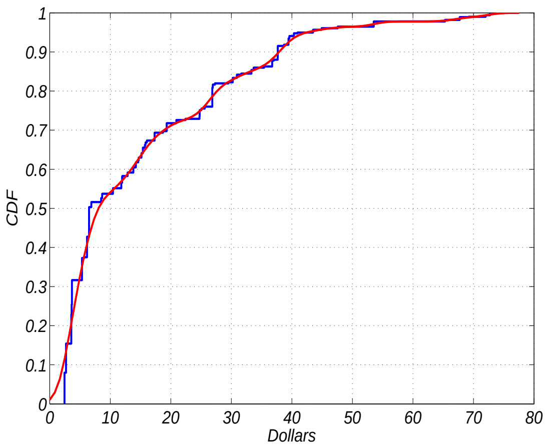
(b) Low Cost

Figure 4: CDF of Per-Flight Downtime (Minutes) Equivalent to Additional Gate

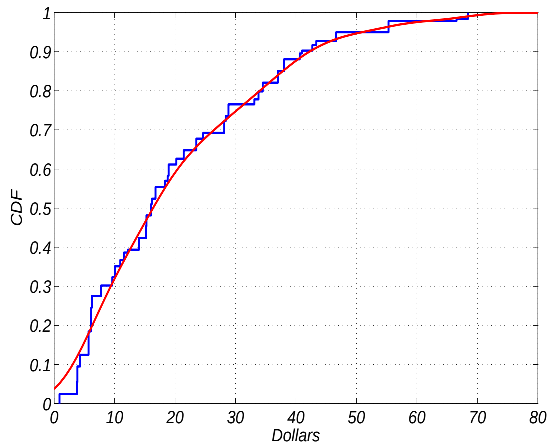
downtime on each flight must be increased to keep outcomes constant if a carrier had access to one less gate, during peak and off-peak hours, for legacy and low-cost carriers respectively. These distributions arise for each carrier type because of how the return to a gate varies with each environment, which we characterize as a combination of airport, time-of-day and day-of-week.[24] In environments where a carrier is using gates intensively, the loss of a gate can be offset only by a large increase in downtime. Notice that these results are consistent with Figure 1. Legacy carriers have a greater change in gate utilization during peak than low-cost carriers, resulting in a larger difference in how much downtime must be added during peak and off-peak hours to offset losing a gate. We perform a similar calculation to recover the increase in flexibility needed to offset losing a gate and find that low-cost carriers require a much larger increase in flexibility to hold outputs fixed.

Next, we combine the estimates in Table 3 with gate-price data to infer the costs associated with each robust-scheduling input. Knowing the tradeoffs faced by a carrier seeking to achieve a given level of outcomes at least cost, along with the price of one input, one can infer the costs of all other inputs. Specifically, to minimize the cost of obtaining any set of outcomes, the marginal productivity per dollar for each pair of inputs, x^n and $x^{n'}$, must be equal to the ratio of their

[24]Note that our estimates relate the fitted distance to GatesShedDep, not Gates. Thus the relationship between the fitted distance and Gates is a nonlinear function of SchedDep, $\frac{\partial \vec{D}}{\partial Gates} = \left(\frac{\partial \vec{D}}{\partial GatesSchedDep}\right)\left(\frac{1}{SchedDep}\right)$



(a) Legacy



(b) Low Cost

Figure 5: CDF of Per-Minute Downtime Costs by Carrier Type

respective prices, p^n and $p^{n'}$,

$$\frac{dx^n}{dx^{n'}} = -\left(\frac{\frac{\partial \overrightarrow{D}^-}{\partial x^{n'}}}{\frac{\partial \overrightarrow{D}^-}{\partial x^n}}\right) = \frac{p^n}{p^{n'}}. \tag{8}$$

For example, our estimates provide the additional minutes of scheduled downtime required to keep outcomes constant if a carrier operating on the frontier had one less gate. Since we know the actual cost of a gate at each airport, $GatePrice$, we can use Equation 8 to infer how a carrier perceives the per-minute cost of downtime. Figures 5(a) and 5(b) give the distribution of the calculated per-minute cost of downtime for legacy and low-cost carriers, respectively. These results rationalize low-cost carriers scheduling approximately half as much downtime as legacy carriers, as the median per-minute cost of downtime for low-cost carriers is \$17.83 versus only \$9.32 for legacy carriers. Direct measures of these operating costs would understate this difference, as the higher-fare legacy carriers may actually forgo more revenue from downtime. Instead, our cost estimates also incorporate indirect opportunity costs, which include the benefits of increased network connectivity, that are greater for legacy carriers operating hub-and-spoke networks.[25]

Performing a similar calculation using Equation 8 for flexibility, we find that this input is much more costly for legacy carriers. The median cost to a legacy carrier of increasing flexibility by

[25]The magnitude of our cost estimates for downtime are also sensible. JetBlue's 2009 annual report estimates that an A-320 operating at the average load-factor, 79.7%, for the average number of hours per day, 11.5, would earn approximately \$64,544.73 in revenue per day. This is approximately \$93.54 in revenue for each minute that the aircraft operates. Our estimate of \$17.83 for the per-minute cost of downtime, which reflects the per-minute reduction in profits due to the aircraft being idle, implies that the margin from operating the aircraft is a plausible 19% (17.83/93.54).

scheduling one additional flight of a given aircraft type to depart in a particular time block is over \$900, versus a median of just under \$200 for low-cost carriers. This differential is consistent with low-cost carriers focusing on high-density routes and operating more homogenous fleets, as well as having a less complex scheduling problem than legacy carriers, which must maintain a high degree of connectivity throughout their network, further increasing the opportunity cost of flexibility. The lower cost of flexibility enjoyed by low-cost carriers does come with greater risk, such as increased vulnerability to problems that might arise with a particular aircraft or engine type. These include design defects, mechanical problems, contractual performance by the manufacturers, or public perception regarding the safety of the aircraft.

Note that the input costs we identify are derived based on the assumption that airlines achieve the observed level of outcomes in the least-cost way, given their operating model. Since our estimates of the costs of robust-scheduling inputs can explain actual differences in the scheduling patterns of carriers, we conclude that carriers appear to be selecting the least-cost means of producing outcomes.

4.3 Estimates of Returns to Inputs

We first use our estimates of the directional distance function coefficients to identify the unit return for each outcome per unit of each input. Specifically, the rate at which a good outcome, y^m, changes in response to an input, x^n, for an efficient carrier is given by the implicit function theorem,

$$\frac{dy^m}{dx^n} = -\left(\frac{\frac{\partial \overrightarrow{D}^-}{\partial x^n}}{\frac{\partial \overrightarrow{D}^-}{\partial y^m}} \right). \tag{9}$$

We use a similar calculation for each bad outcome, $\widetilde{y}^l$. For each outcome, Table 4 reports estimates of the mean return per unit of input, with associated t-statistics. Note that the estimates all have the expected signs – inputs increase good outcomes and decrease bad outcomes – and are statistically significant. We find that per-unit returns for most of the inputs are greater for low-cost carriers. This is consistent with low-cost carriers employing more homogenous fleets, operating on tighter schedules, and using boarding gates more intensely. Interestingly, we find that for every outcome, the per-unit returns to gates per scheduled departure are significantly greater than for the other two inputs.

Next, we examine whether this pattern persists after controlling for the cost of each input. We estimate the improvement in each of the outcomes for each dollar spent on an input, by combining the above per-unit returns with the input prices recovered in Section 4.2. Table 5 reports the average improvement in outcomes per $1,000 spent on the our three inputs for both carrier types. Formally, this is accomplished by dividing Equation 9 by the per-unit price of the input (x^n), p^n, and multiplying by 1,000.

Consider an investment in gates by a low-cost carrier of $1,000 per flight.[26] From the bottom row of Table 5, this investment results in an improvement of 6.87% (1.23/17.9) in the proportion of flights 15 or more minutes late,[27] which is substantially greater than the 4.30% (0.77/17.9) reduction from flexibility or 1.96% (0.35/17.9) reduction from downtime, and this difference is statistically significant. For the low-cost carrier investing in gates, in addition to improved on-time performance, the attainable levels of cargo and loadfactor also increase, and the frequency of long delays is reduced. For legacy carriers, a $1,000 investment in gates yields the greatest relative improvement, although this effect is about one-fourth that achieved by low-cost carriers. This result is consistent with legacy carriers' heterogenous fleets, longer downtimes, and less intensive use of gates. Thus Table 5 reveals that the most cost-effective way for a carrier to target improvement along any outcome dimension is by better managing gates for greater recovery robustness. Yet gates management has received almost no attention in the OR scheduling literature (Barnhart et al. 2003).

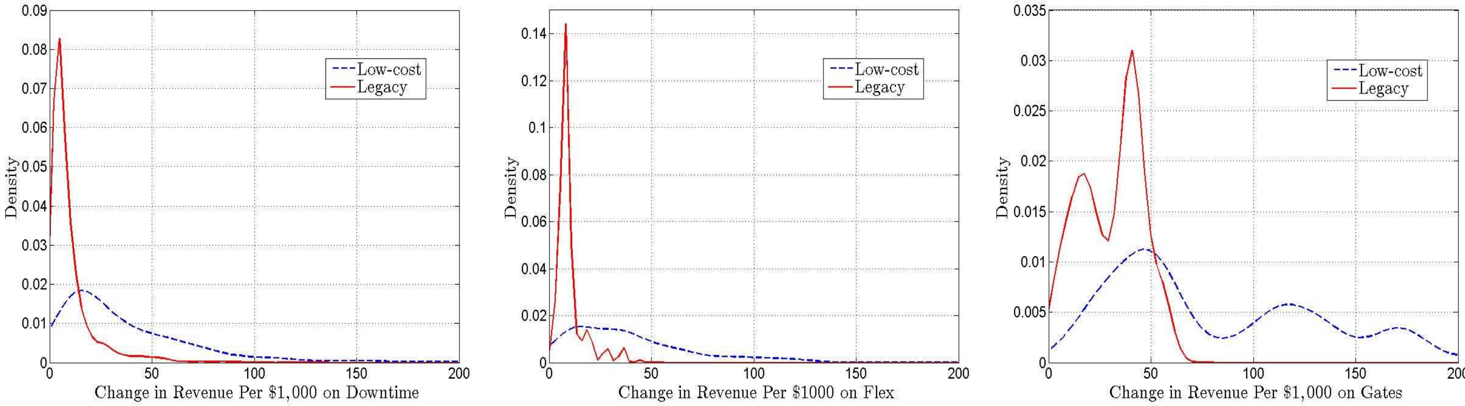
From a public-policy standpoint, the substantial differences in returns to gates between carrier types suggest that the greatest improvements in consumer welfare from reduced delays can be realized by assigning additional gate capacity to low-cost carriers. This result complements a growing literature on the role that airports have in determining equilibrium outcomes and passenger welfare through gate-leasing practices (e.g., Ciliberto and Williams 2010, Snider and Williams 2014).

The average returns reported in Table 5 mask substantial heterogeneity and do not reveal the revenues that result from changes to scheduling inputs. To demonstrate this, Figures 6(a)-6(c) report, by carrier type, the smoothed probability distribution functions of the increase in passenger and cargo revenues per flight, per $1,000 spent on each input. We calculate this additional revenue accounting for the negative tradeoff between loadfactor and ticket prices, and between cargo volume and its price. Specifically, for passenger revenues, we use route-specific price-elasticities of demand from the survey of empirical estimates in IATA (2008), which enables us to account for variation in the mix of routes (long-haul vs. short-haul) served by each carrier. We draw our estimates of the own-price elasticity for cargo from the survey of empirical estimates in Hellerman (2006).[28]

[26] To be comparable across the different inputs, the size of the investments must be normalized this way, since gates impact all flights, flexibility a subset of flights, and downtime a single flight.

[27] From Table 1a, 17.9% is the average frequency of flights over 15 minutes late for low-cost carriers.

[28] The median estimate of passenger's own-price elasticity of demand ranges from -1.1 for long-haul routes to -1.5



(a) Downtime (b) Flex (c) Gates

Figure 6: PDFs of Total Passenger and Cargo Revenues Per $1,000 Spent on Inputs

With these elasticity estimates, the rate of change in total revenues from passengers and cargo from increasing an input, x_n, is given by

$$\sum_{m=1}^{M} \frac{d[p^m y^m]}{dx^n} = \sum_{m=1}^{M} \frac{dy^m}{dx^n}\left[p^m + \frac{dp^m}{dy^m} y^m\right] = \sum_{m=1}^{M} \left(\frac{\frac{\partial \vec{D}}{\partial x^n}}{\frac{\partial \vec{D}}{\partial y^m}}\right) p^m \left[1 + \frac{1}{\epsilon^m}\right], \tag{10}$$

where p^m and ϵ^m are the prices and own-price elasticity of demand for the respective outcomes.[29]

Figures 6(a), 6(b), and 6(c) report the revenue increase from a $1,000 expenditure on Downtime, Flex, and Gates, respectively. Legacy carriers generate substantially less revenues from expenditure on each of the inputs. Consistent with Table 5, increased expenditure on gates generates the greatest revenues. The multi-modal distributions of returns to gates arise for two reasons: differences in utilization across the day and differences in the leasing costs across airports. For both carrier types, there are substantial differences in the returns to boarding gates during peak and off-peak hours. Even though low-cost carriers maintain more consistent use of gates throughout the day, the intensity of use is so much greater at all times that slight differences in utilization over the day results in substantial variation in returns.

for short-haul routes. The median cargo own-price elasticity of demand is -1.3. We use the median estimate for each demand elasticity in our calculations, permitting the passenger elasticity to vary depending on the route distance, a major determinant of the availability of reasonable substitutes for air travel.

[29]The limitations of the DB1B Survey restricts the fare to be the same for all days within a quarter for a carrier, potentially masking variation in revenue returns within each route.

4.4 Estimates of Tradeoffs between Outcomes

While airlines can improve particular operational outcomes by investing in robust-scheduling inputs, they can instead improve some outcomes by sacrificing performance on other outcomes. Our estimated production frontier determines the decline along one outcome dimension necessitated by an improvement along another, if inputs are held constant (Porter 1996, Lapre and Scudder 2004). The tradeoffs carriers are willing to make depend on the value they place on each outcome.

As in Section 4.3, our estimates of the directional distance function can be used to quantitatively measure these tradeoffs. For any two good outcomes, y^m and $y^{m'}$, an efficient carrier can trade off along the frontier between any two outcomes at the rate

$$\frac{dy^m}{dy^{m'}} = -\left(\frac{\frac{\partial \vec{D}}{\partial y^{m'}}}{\frac{\partial \vec{D}}{\partial y^m}}\right). \tag{11}$$

The set of such tradeoffs for each pair of outcomes is presented in Table 6, by carrier-type. For both types of carriers, our estimates satisfy a null-jointness condition, despite not imposing this during estimation. That is, carriers cannot increase a good outcome ($LoadFactor$ or $Cargo$) without an associated increase in a bad outcome ($P(>15)$, $P(>60)$, or $P(>180)$), for any level of inputs.

Our estimates in Table 6 enable us to analyze the tradeoffs between outcomes and develop useful managerial insights. For example, we find that a legacy (low-cost) carrier may trade off an increase of 1% in $P(>15)$, poorer on-time performance, for a 0.31% (0.38%) increase in $LoadFactor$.[30] Using a calculation similar to Equation 10, for each carrier type, we calculate the additional revenue gained by increasing $P(>15)$ by 1%.[31] We find that legacy and low-cost carriers can increase revenues by \$74.33 and \$68.45 per flight, respectively, by accepting a 1% increase in $P(>15)$. Such a tradeoff may be preferable to increasing expenditure on inputs (Table 5). To put this in perspective, for Southwest Airlines in 2007 this amounted to about 0.8% of total revenues per flight, or \$78 million across their entire network (over 1.1 million flights). Additionally, we find that both types of carriers value a 1% reduction in the probability of a flight delay lasting over 180 minutes at more than 4 times as much as one lasting over 15 minutes. This result reinforces the finding of Ramdas et al. (2013) that carriers' market value is most strongly impacted by the frequency of long delays ($P(>180)$). It also rationalizes Southwest's strategy of swapping aircraft and crews to parse one long delay, possibly due to an aircraft incurring a mechanical failure, into many short delays.

[30] Rupp (2009) also documents a tradeoff between $LoadFactor$ and delays.

[31] Note that this calculation accounts for the different mix of aircraft used by carriers, along with the variation across airports and carriers in fares and elasticities. Forbes (2008) performs a similar calculation to measure the impact of delays on fares.

Our estimates of the value carriers place on delays can also guide regulatory policy, which must balance the benefits to passengers of fewer delays against the costs to carriers of limiting the severity and frequency of delays.[32] Currently, as part of the FAA Modernization and Reform Act of 2012 (Section 406(b)), the DOT is conducting such a review of "carrier flight delays, cancelations, and associated causes...". This includes assessing "air carriers' scheduling practices" and "capacity benchmarks at the Nation's busiest airports," as well as providing "recommendations for programs that could be implemented to address the impact of flight delays on air travelers." Our estimates give policy makers insight into the costs that carriers would incur if forced to amend scheduling practices to reduce delays. These can then be weighed against the costs of alternative ways to reduce delays, such as building additional airport facilities to alleviate congestion (Daniel 1995).

4.5 Operational Efficiency Measures

Following Agee et al. (2012), we now use our estimates of the directional distance function to quantify each carrier's technical efficiency, i.e., their ability to attain the efficient level of outcomes from a given level of inputs, and then measure the impact on profits of any inefficiencies. Rather than repeating each step of this calculation, our approach is most easily explained using our example from Section 3.3.[33] Consider two of the carriers from this example, A and B, which are labeled in Figure 7. As before, there is one good outcome, y, and one bad outcome, $\widetilde{y}$, and the figure is drawn for a particular level of inputs, x. Carrier A is on the frontier, with a fitted distance of zero, while carrier B is technically inefficient, with a positive fitted distance. Our estimates of what constitutes *frontier* or efficient production enable us to quantify how much better the outcomes for inefficient carriers can be. As shown in Figure 7, holding fixed the level of the good outcome for carrier B, y_B, this carrier can decrease the bad outcome by $\tilde{y}_B - \tilde{y}_F(y_B)$. Similarly, holding fixed the bad outcome, $\tilde{y}_B$, carrier B can increase the good outcome by $y_B - y_F(\tilde{y}_B)$. We use this idea of a shortfall from *frontier* production to calculate our measures of technical efficiency.

In our data, we observe each carrier, j, operating different aircraft, k, in various environments, i, over a long period of time. In addition, we have multiple good and bad outcomes. As a result, we calculate a carrier's $j's$ average technical efficiency for each good outcome, y^m, over different aircraft, k, and environments, i, in which they operate in each year, t, as

[32] Recent work in economics, Snider and Williams (2014), suggests that current federal policy which limits airports' ability to raise capital to expand airport facilities both limits competition and worsens the frequency and length of delays.

[33] We refer the reader to Agee et al. (2012) for the computational details.

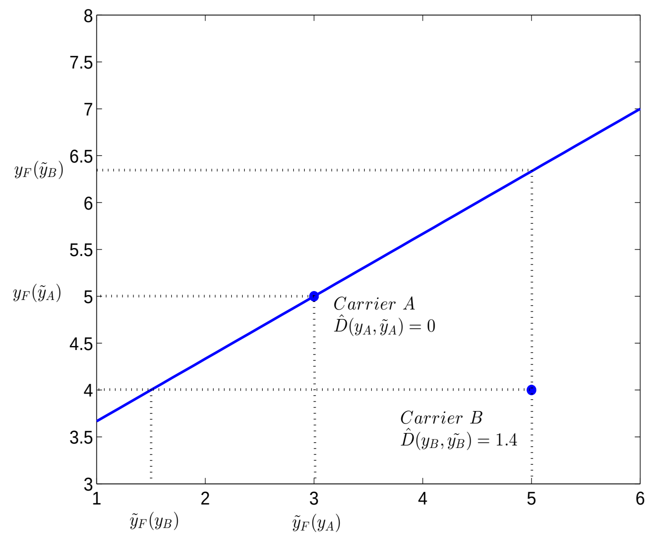


Figure 7: *Technical Efficiency Measures*

$$TE_{jt}^{y^m} = \frac{1}{N_{jt}} \sum_{i,k} \frac{y_{ijkt}^m}{y_{ikt}^{mF}(\mathbf{y}_{ijkt}^{-m}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{x}_{ijkt})},$$

where

$$y_{ikt}^{mF}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}^{-m}, \widetilde{\mathbf{y}}_{ijkt}) = \sup\{y^n : \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \widehat{\gamma}, \widehat{\phi}, \widehat{\eta}) \geq 0\},$$

$\mathbf{y}_{ijkt}^{-m}$ is the vector of other good outcomes (i.e., besides y^n) produced by carrier j, and N_{jt} is the number of observations for carrier j in period t. Our measure of technical efficiency for each good outcome is thus the average percentage shortfall from the relevant *frontier* outcome level across those environments in which the carrier operates in period t. Similarly, for a bad outcome, $\widetilde{y}^l$, we measure each carrier's operational efficiency as

$$TE_{jt}^{\widetilde{y}^l} = 1 - \frac{1}{N_{jt}} \sum_{i,k} \left[\widetilde{y}_{ijkt}^l - \widetilde{y}_{ikt}^{lF}(\mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}^{-l}, \mathbf{x}_{ijkt}) \right],$$

where

$$\widetilde{y}_{ikt}^{lF}(\mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}^{-l}, \mathbf{x}_{ijkt}) = \inf\{\widetilde{y}^l : \overrightarrow{D}(\mathbf{x}_{ijkt}, \mathbf{y}_{ijkt}, \widetilde{\mathbf{y}}_{ijkt}, \mathbf{w}_{it}, \mathbf{d}_{ijkt}; \mathbf{g}_y, -\mathbf{g}_{\widetilde{y}}, \widehat{\gamma}, \widehat{\phi}, \widehat{\eta}) \geq 0\}$$

and $\widetilde{\mathbf{y}}_{ijkt}^{-l}$ is the vector of other bad outcomes (i.e., besides $\widetilde{y}^l$) produced by carrier j.

The reason for taking the difference of a carrier's outcome and the *frontier* outcome level in the case of bad outcomes (i.e., our delays measures) is that the frontier firm might have zero delays so that a relative measure would not be well-defined. Thus, our measure of a carrier's efficiency with respect to each of the bad outcomes is the reduction in the bad outcome that would be required to reach the frontier, holding constant inputs and other outcomes. This bounds our measures of technical efficiency for both good and bad outcomes between zero and one.[34]

In Table 7, we report the estimates of carriers' average efficiency in 2007, and, since we have actual prices for the two good outcomes, cargo and loadfactor, we also report lost revenue due to the inefficient production of these outcomes.[35] Similar to our analysis of returns to inputs in Equation 10 of Section 4.2, we utilize the elasticity of demand for cargo and passengers to calculate the tradeoff between volume and price.

For good and bad outcomes, the relative measures are bounded between 0 and 1, with the most efficient firm having a value of 1. For example, Delta's relative measure of 0.91 for $LoadFactor$ (a good outcome) implies that, after conditioning on levels of inputs and other outcomes, Delta achieves 91% of the frontier level of $LoadFactor$ defined by American. By our estimates, Delta lost over \$500 million in 2007 due to its load-factor inefficiencies, even after accounting for the reduction in ticket price needed to achieve an efficient level of load factor. The results for cargo are less dramatic, but still substantial, as Delta's inefficiencies in this regard resulted in lost revenue of over \$21 million. Similarly, Delta's technical efficiency measure of 0.98 for $P(>180)$ implies that Delta can increase profits by reducing the frequency of long delays by 0.02, with no additional expenditure on inputs.

The efficiency results in Table 7 for the delay measures demonstrate the value of our directional distance function approach, which characterizes the efficient level for each of the multiple outcomes for any given level of inputs and operating environment. Northwest is the most efficient carrier for each of the delay measures; however, Continental is a very close second. Yet this result is completely counter to Table 1a if one considers the delay measures in isolation, as Continental performs rather poorly across each measure of delays. However, Continental employs much smaller amounts of *Downtime* and *Flex* than other legacy carriers, and is very near the top in terms of gate utilization. Thus, once the level of all inputs, all outcomes, and the operating environment is accounted for, Continental is relatively efficient. American's last-place ranking for each of the bad

[34] Note that our measures of efficiency credit carriers for operating in more difficult environments by normalizing their outcome levels by the *frontier* outcome level in each environment. In difficult environments, the frontier outcome level is characterized by higher levels of bad outcomes and lower levels of good outcomes. Thus, a carrier's performance relative to the best practices in each environment is captured by our measures.

[35] One can not compare all 10 carriers in the earlier years, 1997 and 2001, as some of the low-cost carriers were not yet large enough (1% of domestic enplanements) to be required to report operating data to the DOT. We report efficiency results for 2007 because the gate survey had the highest response rate in this year. Results in 2008 and 2009 are very similar.

outcomes can be explained by its high frequency of each type of delay along with the large amount of inputs that it employs.

Southwest's ranking is also of particular interest, as it is often touted as a successful operating model for the industry. The results in Table 7, which suggest that Southwest is rather average in terms of technical efficiency, can be misleading. Southwest operates in the least-congested and most-concentrated airports in our data (see Table 1b), which are characterized by average levels for good and bad outcomes that are substantially higher and lower, respectively, than at other airports. After controlling for the effect of these well-chosen environments, we find Southwest's performance is in the middle of the pack. Importantly, since 2007, Southwest has substantially increased $LoadFactor$, by slightly more than 10%, and revenues by over 50%, with no accompanying increase in delays. This is consistent with the estimates in Table 7, which suggest Southwest was quite inefficient along this dimension and could increase $LoadFactor$ with no increase in the frequency of delays. Thus, Southwest's management was effective at identifying this inefficiency and reducing it, which is consistent with its reputation as an industry leader in operations (e.g., Shumsky 2006).[36]

5 Concluding Remarks

We examine how airlines have used robust-scheduling practices to reduce the negative impact of unforeseen service disruptions on profits by estimating a multiple-outcome, multiple-input production frontier. We use estimates of the returns to robust-scheduling inputs and the tradeoffs faced by carriers along this frontier to infer the costs of certain robust-scheduling inputs and how carriers can most profitably target improvements through expenditure on inputs. We also calculate the tradeoffs among outcomes faced by carriers along this frontier to infer the costs of reducing delays. The estimates of a carrier's shortfall from best practices measure how operating inefficiencies impact profits and where each carrier should target improvement.

Our results provide important methodological and empirical insights to both academics and practitioners. We find that the largest per-dollar returns are from expenditure on gates, or from more effectively managing existing gate capacity by smoothing departures across the day to avoid periods of intense utilization. Interestingly, management of this crucial resource by airlines has received little attention in the academic literature (Barnhart et al. 2003), despite being of central focus to regulators both in the evaluation of airline mergers (e.g., the settlements made by American Airlines and US Airways with the US Department of Justice) and in their efforts to improve industry

[36] We include many temporal and carrier-specific controls to help rule out many alterative macro-economic explanations (e.g., year-month fixed effects control for aggregate fluctuations in demand).

performance and enhance passenger welfare (e.g., FAA Modernization and Reform Act of 2012). We also obtain interesting comparative results for flexibility and downtime that rationalize scheduling patterns of carriers. Our empirical methodology for identifying the returns to schedule robustness using retrospective data also reduces the level of uncertainty involved when carriers conduct costly large-scale experiments such as Delta's Operation Clockwork.

Both our application and methodology provide direction for future research. Our application demonstrates the importance of accounting for recovery robustness when examining airlines' scheduling practices. We show that there are significant returns to recovery robustness, which suggests the need for models that formally incorporate robustness criteria. In terms of methodology, we adapt econometric tools recently developed in the economics literature (Atkinson and Primont 2002, Atkinson and Dorfman 2005, and Fare et al. 2005) to study scheduling practices in an extremely complex operating environment. The flexibility of these tools makes them amenable to identifying efficient operating practices in other complex industries like health care.

References

[1] Agee, M., S. Atkinson, and T. Crocker. 2012. Child Maturation, Time-Invariant, and Time-Varying Inputs; Their Interaction in the Production of Child Human Capital. *Journal of Productivity Analysis.* **38** 29-44.

[2] Arguello, M., J. Bard, and G. Yu. 1997. A GRASP for Aircraft Routing in Response to Groundings and Delays. *Journal of Combinatorial Optimization.* **5** 211-228.

[3] Arikan, M., V. Deshpande, and M. Sohoni. 2013. Building Reliable Air-Travel Infrastructure Using Empirical Data and Stochastic Models of Airline Networks. *Operations Research.* **61** 45-64.

[4] Atkinson, S., D. Primont, and M. Tsionas, 2014. Estimation of Efficient Production Using Optimal Directions with Multiple Inputs and Outputs. *Working Paper.*

[5] Atkinson, S. and C. Cornwell. 1993. Measuring Technical Efficiency with Panel Data: A Dual Approach. *Journal of Econometrics.* **59** 257-261.

[6] Atkinson, S. and C. Cornwell. 1994. Parametric Estimation of Technical and Allocative Inefficiency with Panel Data. *International Economic Review.* **35** 231-243.

[7] Atkinson, S. and D. Primont. 2002. Stochastic Estimation of Firm Technology, Inefficiency, and Productivity Growth Using Shadow Cost and Distance Functions. *Journal of Econometrics.* **108(2)** 203-225.

[8] Atkinson, S. and J. Dorfman. 2005. Bayesian Measurement of Productivity and Efficiency in the Presence of Undesirable Outputs: Crediting Electric Utilities for Reducing Air Pollution. *Journal of Econometrics.* **126(2)** 445-468.

[9] Barnhart, C., P. Belobaba, and A. Odoni. 2003. Application of Operations Research in the Air Transport Industry. *Transportation Science.* **37(4)** 368-391.

[10] Berndt, E. and D. Wood. 1975. Technology, Prices, and the Derived Demand for Energy. *The Review of Economics and Statistics.* **57(3)** 259-268.

[11] Bratu, S. and C. Barnhart. 2006. Flight Operations Recovery: New Approaches Considering Passenger Recovery. *Journal of Scheduling.* **9(3)** 279-298.

[12] Brueckner, J. 2002. Airport Congestion When Carriers Have Market Power. *American Economic Review.* **92(5)** 1357-1375.

[13] Brueckner, J. 2005. Internalization of Airport Congestion: A Network Analysis. *International Journal of Industrial Organization.* **23(7-8)** 599-614.

[14] Cachon, G. and C. Terwiesch. 2011. Matching Supply with Demand: An Introduction to Operations Management. *McGraw-Hill.*

[15] Cannon, J., T. Randall, and C. Terwiesch. 2012. An Operations Based Earnings Prediction Model Applied to the US Airline Industry. *Working Paper.*

[16] Caulkins, J. A. Barnett, P. Larkey, Y. Yuan, and J. Goranson. 1993. The On-Time Machines: Some Analyses of Airline Punctuality. *Operations Research.* **41(4)** 710-720.

[17] Chambers, R., Y. Chung, and R. Färe. 1998. Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications.* **98(2)** 351-364.

[18] Chambers, R., R. Färe, S. Grosskopf, and M. Vardanyan. 2013. Generalized Quadratic Revenue Functions. *Journal of Econometrics.* **173(1)** 11-21.

[19] Ciliberto, F. and J. Williams. 2010. Limited Access to Airport Facilities and Market Power in the Airline Industry. *Journal of Law and Economics.* **53(2)** 467-495.

[20] Clausen, J., A. Larsen, J. Larsen, and N. Rezanova. 2010. Disruption Management in the Airline Industry – Concepts, Models and Methods. *Computers & Operations Research.* **37** 809-821.

[21] Daniel, J.. 1995. Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues. *Econometrica.* **63(2)** 327-370.

[22] Deshpande, V. and M. Arikan. 2011. The Impact of Airline Flight Schedules on Flight Delays. *Manufacturing & Service Operations Management, Forthcoming.*

[23] Färe, R., S. Grosskopf, D. Noh, and W. Weber. 2005. Characteristics of a Polluting Technology: Theory and Practice. *Journal of Econometrics.* **126(2)** 469-492.

[24] Forbes, S. 2008. The Effect of Air Traffic Delays on Airline Prices. *International Journal of Industrial Organization.* **26(5)** 1218-1232.

[25] Greene, W. 2005. Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *Journal of Econometrics.* **126(2)** 269-303.

[26] Hart, A. and M. Maynard. 2005. Another Clock Is Ticking at Delta Air Lines. *The New York Times*, New York, January 29, 2005.

[27] Hellerman, R. 2006. Capacity Options for Revenue Management: Theory and Applications in the Air Cargo Industry. *Springer.*

[28] International Air Transport Association. 2008. Air Travel Demand, IATA Economics Briefing No. 9, Measuring the Responsiveness of Air Travel Demand to Changes in Prices and Incomes. *IATA Press.*

[29] Jiang, H. and C. Barnhart. 2009. Dynamic Airline Scheduling. *Transportation Science.* **43(3)** 336-354.

[30] Krinsky, I. and A. Robb. 1986. On Approximating the Statistical Properties of Elasticities. *The Review of Economics and Statistics.* **68(4)** 715-719.

[31] Kumbhakar, S. and C. Lovell. 2000. Stochastic Frontier Analysis. *Cambridge University Press.*

[32] 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** 15-28.

[33] Lapre, M. and G. Scudder. 2004. Performance Improvement Paths in the U.S. Airline Industry: Linking Tradeoffs to Asset Frontiers. *Production and Operations Management.* **13(2)** 123–134.

[34] Li, J. and S. Netessine. 2011. Partnering with Competitors – An Empirical Analysis of Airline Alliances and Multimarket Competition. *SSRN Working Paper #1761211.*

[35] Li, J., N. Granados, and S. Netessine. 2014. Are Consumers Strategic? Structural Estimation from the Air-Travel Industry. *Forthcoming in Management Science.*

[36] Lieberman, M. and R. Dhawan. 2005. Assessing the Resource Base of Japanese and U.S. Auto Producers: A Stochastic Frontier Production Approach. *Management Science.* **51(7)** 1060-1075.

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

[38] Mazzeo, M. 2003. Competition and Service Quality in the US Airline Industry. *Review of Industrial Organization.* **22** 275–296.

[39] Olivares, M., C. Terwiesch, and L. Cassorla. 2008. Structural Estimation of the Newsvendor Model: An Application to Reserving Operating Room Time. *Management Science.* **54** 41-55.

[40] Porter, M. 1996. What Is Strategy? *Harvard Business Review.* Nov.-Dec. 6-78.

[41] Ramdas, K., J. Williams and M. Lipson. 2013. Can Stock Price Movements Inform Operational Improvement Efforts? Evidence from the Airline Industry. *Manufacturing & Service Operations Management, Forthcoming.*

[42] Rosenberger, J., A. Schaefer, D. Goldsman, E. Johnson, A. Kleywegt, and G. Nemhauser. 2002. A Stochastic Model of Airline Operations. *Transportation Science.* **36** 357-377.

[43] Rupp, N. 2009. Do Carriers Internalize Congestion Costs? Empirical Evidence on the Internalization Question. *Journal of Urban Economics.* **65(1)** 24-37.

[44] Rupp, N. and G. Holmes. 2006. An Investigation into the Determinants of Flight Cancellations. *Economica.* **73(292)** 749-783.

[45] Rupp, N., G. Holmes, and J. DeSimone. 2005. Airline Schedule Recovery after Airport Closures: Empirical Evidence since September 11. *Southern Economic Journal.* **71(4)** 800-820.

[46] Schumer, C. and C. Maloney. 2008. Your Flight Has Been Delayed Again: Flight Delays Cost Passengers, Airlines, and the US Economy Billions. *US Joint Economic Committee.*

[47] Shumsky, R. 1993. Response of U.S. Air Carriers to On-Time Disclosure Rule. *Transportation Research Record.* **1379** 9-16.

[48] Shumsky, R. 2006. The Southwest Effect, Airline Alliances and Revenue Management. *Journal of Revenue and Pricing Management.* **5** 83-89.

[49] Smith, B. and E. Johnson. 2006. Robust Airline Fleet Assignment: Imposing Station Purity Using Station Decomposition. *Transportation Science.* **40(4)** 497-516.

[50] Snider, C. and J. Williams. 2014. Barriers to Entry in the Airline Industry: A Multi-Dimensional Regression-Discontinuity Analysis of the Wendell H. Ford Aviation Investment and Reform Act. *Forthcoming Review of Economics and Statistics.*

[51] Williams, J. 2014. Capacity Investments, Exclusionary Behavior, and Welfare: A Dynamic Model of Competition in the Airline Industry. *Working Paper.*

Table 1a: Descriptive Statistics by Carrier and Year, Outcomes and Prices

	N	Cargo	p^{cargo}	Loadfactor	p^{pax}	P(>15)	P(>60)	P(>180)
American (AA)	21,328	6.350	0.534	0.765	0.196	0.232	0.070	0.005
Continental (CO)	11,348	2.715	0.748	0.736	0.210	0.215	0.078	0.009
Delta (DL)	24,076	4.583	0.599	0.723	0.235	0.161	0.034	0.003
Northwest (NW)	17,158	2.320	0.770	0.727	0.264	0.168	0.036	0.002
United (UA)	18,278	3.921	0.417	0.734	0.209	0.206	0.057	0.004
US Airways (US)	24,992	2.723	0.919	0.714	0.285	0.168	0.034	0.003
Legacy Average	21,328	3.768	0.667	0.732	0.236	0.189	0.049	0.004
Frontier (F9)	603	8.334	0.717	0.768	0.151	0.204	0.032	0.002
JetBlue (B6)	3,823	4.086	0.652	0.788	0.125	0.226	0.077	0.011
AirTran (FL)	7,012	2.639	0.671	0.769	0.150	0.200	0.052	0.005
Southwest (WN)	43,762	3.899	0.920	0.672	0.181	0.171	0.030	0.002
Low-Cost Average	55,200	4.740	0.868	0.693	0.173	0.179	0.036	0.003
1997	53,845	3.732	0.595	0.668	0.266	0.161	0.030	0.002
2001	22,882	3.560	0.639	0.658	0.239	0.160	0.036	0.002
2007	44,405	4.145	0.819	0.768	0.183	0.230	0.063	0.006
2008	25,676	3.855	0.824	0.751	0.187	0.196	0.053	0.005
2009	25,572	3.927	0.854	0.768	0.166	0.174	0.042	0.003
Overall Average	172,380	3.863	0.731	0.719	0.215	0.185	0.045	0.004

Table 1b: Descriptive Statistics by Carrier and Year, Inputs/Controls and Prices

	N	HHI	Congestion	Downtime	Flex	GatesSchedDep	p^{gates}
American (AA)	21,328	1,162.299	4.664	33.653	47.146	0.302	76,822.39
Continental (CO)	11,348	908.623	4.622	30.235	29.319	0.279	78,018.92
Delta (DL)	24,076	637.729	4.423	34.697	43.067	0.425	55,837.79
Northwest (NW)	17,158	940.266	4.164	39.014	46.708	0.377	46,307.91
United (UA)	18,278	923.240	4.990	30.734	37.432	0.277	111,683.2
US Airways (US)	24,992	683.176	4.376	33.351	38.105	0.274	71,529.22
Legacy Average	21,328	858.093	4.527	33.801	41.074	0.326	72,364.43
Frontier (F9)	603	862.673	6.783	37.042	75.353	0.134	86279.08
JetBlue (B6)	3,823	1,950.074	4.235	21.320	24.176	0.147	77842.26
AirTran (FL)	7,012	938.669	4.627	14.737	42.470	0.124	48223.67
Southwest (WN)	43,762	3,495.454	4.409	6.071	49.681	0.138	51779.13
Low-Cost Average	55,200	3,034.879	4.451	8.566	47.279	0.136	54319.28
1997	53,845	1,060.596	4.297	24.442	42.382	0.242	NA
2001	22,882	1,564.261	4.505	25.814	39.832	0.271	NA
2007	44,405	1,441.198	4.818	26.396	43.158	0.301	71140.59
2008	25,676	2,144.538	4.564	26.169	44.322	0.258	60243.3
2009	25,572	2,194.430	4.321	26.705	45.948	0.255	60244.63
Overall Average	172,380	1,555.150	4.502	25.720	43.061	0.266	65310.38

Table 2: Directional Distance Function Coefficients

	Low-Cost	Legacy Interactions
LoadFactor	-0.273	-0.502
	(-48.741)**	(-38.917)**
Cargo	-0.448	-0.083
	(-46.803)**	(-5.155)**
P(>15)	0.166	0.176
	(44.424)**	(27.389)**
P(>60)	0.035	0.089
	(8.623)**	(12.075)**
P(>180)	0.078	0.151
	(16.087)**	(17.018)**
GatesSchedDep	0.036	0.019
	(6.933)**	(2.074)**
Flex	0.048	0.059
	(5.683)**	(4.254)**
Downtime	0.015	0.011
	(2.711)**	(1.183)
Congestion	0.0376	-0.0145
	(5.328)**	(-1.582)
HHI	0.020	0.058
	(3.104)**	(6.374)**

Note: Asymptotic t-statistics in parentheses. Significance at the .05 (.01) level denoted by * (**), respectively.

Table 3: Tradeoffs Between Inputs (n=rows, m=columns): $\frac{dx^n}{dx^{n'}} = -\left(\frac{\frac{\partial \vec{D}_t^-}{\partial x^{n'}}}{\frac{\partial \vec{D}_t^-}{\partial x^n}}\right)$

Legacy Carriers			
	Flex	Downtime	GatesSchedDep
Flex		-3.132	-0.010
		(-3.147)**	(-5.970)**
Downtime	-0.319		-0.003
	(-3.147)**		(-2.847)**
GatesSchedDep	-96.384	-301.850	
	(-5.970)**	(-2.847)**	

Low-Cost Carriers			
	Flex	Downtime	GatesSchedDep
Flex		-1.063	-0.002
		(-2.559)**	(-4.893)**
Downtime	-0.941		-0.002
	(-2.559)**		(-2.471)**
GatesSchedDep	-414.262	-440.230	
	(-4.893)**	(-2.471)**	

Note: Asymptotic t-statistics in parentheses. Significance at the .05 (.01) level denoted by * (**), respectively.

Table 4: Returns to Inputs (n=rows, m=columns): $\frac{dy^m}{dx^n} = -\left(\frac{\frac{\partial \vec{D}_t^-}{\partial x^n}}{\frac{\partial \vec{D}_t^-}{\partial y^m}}\right)$

Legacy Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Flex	0.011 (11.078)**	0.027 (10.942)**	-0.100 (-11.092)**	-0.139 (-8.164)**	-0.018 (-9.756)**
Downtime	0.004 (3.200)**	0.009 (3.220)**	-0.032 (-3.199)**	-0.044 (-3.149)**	-0.006 (-3.146)**
GatesSchedDep	1.079 (7.098)**	2.618 (7.140)**	-9.531 (-7.057)**	-13.405 (-6.816)**	-1.755 (-6.680)**

Low-Cost Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Flex	0.005 (5.683)**	0.039 (5.462)**	-0.104 (-5.601)**	-0.225 (-4.703)**	-0.026 (-5.165)**
Downtime	0.005 (2.711)**	0.036 (2.683)**	-0.098 (-2.684)**	-0.212 (-2.547)**	-0.024 (-2.589)**
GatesSchedDep	2.271 (6.933)**	16.046 (6.780)**	-43.057 (-6.873)**	-93.281 (-5.284)**	-10.663 (-6.077)**

Note: Asymptotic t-statistics in parentheses. Significance at the .05 (.01) level denoted by * (**), respectively.

Table 5: Returns to Inputs Per $1,000 Dollars (n=rows, m=columns): $\frac{\frac{dy^m}{dx^n}}{p^n} = -\frac{\left(\frac{\frac{\partial \vec{D}_t^-}{\partial x^n}}{\frac{\partial \vec{D}_t^-}{\partial y^m}}\right)}{p^n}$

Legacy Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Flex	0.013 (4.917)**	0.030 (5.690)**	-0.097 (-8.335)**	-0.171 (-7.404)**	-0.018 (-9.021)**
Downtime	0.012 (4.814)**	0.029 (6.906)**	-0.223 (-5.127)**	-0.391 (-4.913)**	-0.040 (-7.632)**
GatesSchedDep	0.038 (11.276)**	0.092 (10.547)**	-0.298 (-10.965)**	-0.521 (-9.158)**	-0.054 (-9.970)**

Low-Cost Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Flex	0.043 (4.957)**	0.296 (4.485)**	-0.769 (-9.800)**	-1.853 (-8.142)**	-0.180 (-7.916)**
Downtime	0.039 (5.792)**	0.266 (5.960)**	-0.350 (-11.036)**	-0.843 (-8.849)**	-0.082 (-8.757)**
GatesSchedDep	0.069 (11.743)**	0.475 (11.392)**	-1.232 (-9.655)**	-2.967 (-9.171)**	-0.289 (-10.706)**

Note: Asymptotic t-statistics calculated using the methodology of Krinsky and Robb (1986) are in parentheses. Significance at the .05 (.01) level denoted by * (**), respectively.

Table 6: Tradeoffs Between Outputs (n=rows, m=columns): $\frac{dy^m}{dy^n} = -\left(\frac{\frac{\partial \vec{D}_t^-}{\partial y^n}}{\frac{\partial \vec{D}_t^-}{\partial y^m}}\right)$

Legacy Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Cargo		-2.428 (-62.305)**	8.837 (63.659)**	12.429 (16.015)**	1.627 (26.272)**
Loadfactor	-0.412 (-62.305)**		3.640 (51.798)**	5.120 (14.698)**	0.670 (21.199)**
P(>15)	0.11316 (63.659)**	0.27471 (51.798)**		-1.40641 (-13.725)**	-0.18411 (-21.857)**
P(>60)	0.080 (16.015)**	0.195 (14.698)**	-0.711 (-13.725)**		-0.131 (-14.441)**
P(>180)	0.615 (26.272)**	1.492 (21.199)**	-5.432 (-21.857)**	-7.639 (-14.441)**	

Low-Cost Carriers					
	Cargo	Loadfactor	P(>15)	P(>60)	P(>180)
Cargo		-7.066 (-48.741)**	18.960 (44.424)**	41.076 (8.623)**	4.695 (16.087)**
Loadfactor	-0.142 (-48.741)**		2.683 (43.485)**	5.813 (8.390)**	0.665 (15.913)**
P(>15)	0.053 (44.424)**	0.373 (43.485)**		-2.166 (-7.867)**	-0.248 (-14.940)**
P(>60)	0.024 (8.623)**	0.172 (8.390)**	-0.462 (-7.867)**		-0.114 (-7.498)**
P(>180)	0.213 (16.087)**	1.505 (15.913)**	-4.038 (-14.940)**	-8.748 (-7.498)**	

Note: Asymptotic t-statistics in parentheses. Significance at the .05 (.01) level denoted by * (**), respectively.

Table 7: Firm Avg. Technical Efficiencies for Each Output (2007)

	Cargo		Load Factor				
Airline	TE	Lost Revenue	TE	Lost Revenue	P(>15)	P(>60)	P(>180)
American (AA)	1.00	0	1.00	0	0.84	0.67	0.96
JetBlue (B6)	0.62	1,544,864	0.89	115,650,446	0.92	0.83	0.98
Continental (CO)	0.31	24,240,330	0.80	422,265,260	1.00	0.99	1.00
Delta (DL)	0.66	21,628,126	0.91	506,729,441	0.90	0.80	0.98
Frontier (F9)	0.97	45,836	0.98	6,723,847	0.85	0.69	0.96
AirTran (FL)	0.41	719,688	0.84	238,225,503	0.95	0.90	0.99
Northwest (NW)	0.27	2,990,038	0.79	579,610,554	1.00	1.00	1.00
United (UA)	0.47	16,892,102	0.85	735,163,514	0.95	0.89	0.99
US Airways (US)	0.43	35,515,292	0.84	791,773,454	0.95	0.90	0.99
Southwest (WN)	0.49	32,528,488	0.83	1,120,583,911	0.95	0.89	0.99
Avg.	0.57	61,146,448	0.88	1,769,304,334	0.94	0.87	0.99