Building Reliable Air-Travel Infrastructure Using Empirical Data and Stochastic Models of Airline Networks

Mazhar Arıkan, Vinayak Deshpande, Milind Sohoni,
Building Reliable Air-Travel Infrastructure Using Empirical Data and Stochastic Models of Airline Networks

Mazhar Arıkan
University of Kansas, Lawrence, Kansas 66045, mazhar@ku.edu

Vinayak Deshpande
University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, vinayak_deshpande@unc.edu

Milind Sohoni
Indian School of Business, Gachibowli, Hyderabad 500 032, India, milind_sohoni@isb.edu

Flight delays have been a growing issue and they have reached an all-time high in recent years, with the airlines’ on-time performance at its worst level in 2007 since 1995. A recent report by the Joint Economic Committee of the U.S. Congress chaired by Senator Charles E. Schumer has estimated that the total cost to the U.S. economy because of flight delays was as much as $41 billion in 2007. The goal of this paper is to build stochastic models of airline networks and utilize publicly available data to answer the following policy questions: Which are the bottleneck airports in the U.S. air-travel infrastructure (i.e., airports that cause most delay propagation)? How would increasing airport capacity at these airports alleviate delay propagation? What are the appropriate metrics for measuring the robustness of airline schedules? How could these schedules be made more robust? Which flight in an aircraft rotation is a bottleneck flight (and, hence, deserves managerial attention)? Flight delays are typically attributed to two factors: (i) the randomness in the intrinsic travel time for a scheduled flight (which is the travel time excluding propagated delays), and (ii) the propagation of this randomness through the air-travel network and infrastructure. We model both of these factors that cause travel delays. The contribution of this paper is twofold. First, we develop stochastic models, using empirical data, to analyze the propagation of delays through air-transportation networks. Our stochastic models allow us to develop three important robustness measures for airline networks. Second, our analysis enables us to make policy recommendations regarding managing bottleneck resources in the air-travel infrastructure, which, if addressed, could lead to a significant improvement in air-travel reliability.

Subject classifications: air-travel infrastructure; schedule robustness; empirical analysis; stochastic models.


History: Received September 2010; revisions received October 2011, May 2012; accepted September 2012. Published online in Articles in Advance February 8, 2013.
the airline’s network), and (ii) the propagation of this randomness through the air-travel network and infrastructure. We call the delay caused because of a spillover from the previous flight of the aircraft total propagated delay (see §3 for precise definitions). We model both of these factors that cause travel delays. The randomness in travel time is affected by independent and collaborative decisions made by multiple agencies within the air-transportation network such as airlines, airports, and regulatory agencies like the Federal Aviation Administration (FAA). For example, airport capacity, which is driven by several factors such as number of runways, weather disruptions, etc., has a significant impact on the total travel time. Airline scheduling and resource utilization decisions impact congestion at airports, which in turn affect the total travel time. Because most airline flight networks are tightly coupled, and resources such as airports are shared across networks, any delay caused within the system (either because of a single flight delay or a ground stop at an airport) propagates across the entire transportation network and impacts the performance of the air-transportation infrastructure as a whole. Although delay propagation models have been studied in the past literature, it is important to note that most of these models are deterministic in the sense that a given delay on a specific flight (called the root flight) is propagated through the network by assuming that there are no delays on subsequent flights in the network (Kondo 2008; AhmadBeygi et al. 2008, 2010). In other words, these models replace the block-time random variable by the scheduled block time for subsequent flights in the network. The key contribution of our modeling approach, however, lies in defining and constructing a stochastic model of delay propagation, which convolutes a stochastic root delay with stochastic travel times of subsequent flights in an aircraft rotation to compute the propagated delay in the airline network.

The goal of our research is to answer the following policy questions by building stochastic models and utilizing publicly available data. Which are the bottleneck airports in the U.S. air-travel infrastructure (i.e., airports that cause most delay propagation)? How would increasing airport capacity at these airports alleviate delay propagation? What are the appropriate metrics for measuring the robustness of airline schedules? How could these schedules be made more robust? Which flight in an aircraft rotation is a bottleneck flight (and, hence, deserves managerial attention)?

To answer these questions, we collected publicly available data from several data sources including the FAA, the Bureau of Transportation Statistics (BTS), and aircraft manufacturer websites. Our primary data source is the on-time performance data set published by the BTS. We collected data on approximately 21 million domestic flights flown by major airlines over a three-year period (2005–2007). We first build a stochastic model of intrinsic (i.e., nonpropagated) travel time by analyzing factors that affect this travel time such as route, airport, airline, congestion, etc. This intrinsic travel-time distribution is then combined with a stochastic model of delay propagation to obtain the network impact of a delayed flight. We show that the deterministic approach, used in the prior literature, overestimates the root propagated delay and the total propagated delay in an airline network. These stochastic models are then utilized to define several metrics of airline network robustness.

Our analysis reveals that all airlines have negative block-time buffers (i.e., scheduled block times are not enough to cover actual block times) built in their schedule, potentially driven by the 15 minute buffer used by the DOT in reporting on-time performance. We evaluate airlines based on the magnitude of delay propagation in their networks, and recommend that airlines should judiciously allocate scheduled block-time and ground-time buffers by analyzing the stochastic propagation of delays. We also construct a metric to identify “bottleneck” flights in airline networks. Airlines should focus their managerial attention on such bottleneck flights to improve the robustness of their schedules. Our airport analysis reveals that the top five airports in the United States with the most network impact are Atlanta, Chicago O’Hare, Dallas Fort Worth, Phoenix, and Los Angeles (LAX). Our stochastic model is then applied to study the impact of alleviating congestion at these airports. Using our stochastic models, we build the marginal network delay impact curve as a function of the reduction in congestion (increase in capacity at congested airports). Our analysis reveals that a 15% increase in capacity at these airports will result in an average reduction in network delays by approximately 10%. In addition, we show that the highest network delay reduction is obtained within the first 5% capacity increase.

Our paper is organized as follows. In §2, we briefly review the literature on airline flight delays and schedule robustness. Section 3 explains the data used in the study and the stochastic models developed. Different measures for robustness and service reliabilities based on our stochastic models are introduced in §4. In §5, we develop policy recommendations to improve the reliability of the air-transportation infrastructure. Section 6 concludes the paper with a brief summary of findings and policy implications.

2. Literature Review

This paper draws on three streams of research on airline flight delays and schedule robustness: (i) research identifying factors that affect block times for better block-time estimation, (ii) modeling flight delay propagation, and (iii) developing performance metrics that consider flight schedule robustness.

Factors that have an impact on block times have been a subject of much discussion recently in both academia and popular press. In a Wall Street Journal article, McCartney (2007a, August 13) explains airlines’ decision of using smaller airplanes as one of the reasons of increased congestion and delays at airports. The same author reports that scheduled block times are higher compared to 10 years...
ago although planes can fly faster and navigate better with more advanced technology (McCartney 2007b, May 29), (McCartney 2010, February 4). This block-time inflation (i.e., schedule padding) is explained by increased variability in the air-travel infrastructure, airlines’ efforts to improve their on-time rankings, and slower flying speeds of smaller regional jets preferred by airlines. Recently, Ball et al. (2010) show that in 2007, the total schedule buffer of 9,526 million minutes cost the airlines $3.7 billion and passengers $6 billion. They define the schedule buffer as the extra time added into the schedule by airlines to reduce the impact of potential delays. In this paper, we provide tools to airline schedulers to intelligently allocate block times to flights and reduce the overall cost of schedule padding without disrupting the passenger service. Academic research regarding block-time forecasting mostly focuses on individual segments of block times. Shumsky (1995) uses a linear regression model to predict taxi-out times, and Mueller and Chatterji (2002) compare different distributions for departure, arrival, and en route delays. Hebert and Dietz (1997), Willeman et al. (2003), and Tu et al. (2008) are among others who study the estimation of different segments of block times. In contrast, we construct a probabilistic model of total block time instead of breaking it up into segments such as departure delay, taxi times, air time, etc., and propose several explanatory variables to predict the total block time. Unlike most of the studies mentioned above, which focus on specific airports and shorter time periods, our model uses data on all the domestic flights by all major airlines flown in the United States during 2005 through 2007 consisting of more than 21 million observations.

This study contributes to the robust airline scheduling literature (for a review, see Ball et al. 2007 and Barnhart and Cohn 2004) by developing a stochastic model of delay propagation and using publicly available industry-wide data. Researchers have conducted several studies regarding modeling flight delay propagation. Ahmad Beygi et al. (2008) and Lapp et al. (2008) analyze how a flight delay propagates through the network using a tree structure (i.e., propagation tree). Ahmad Beygi et al. (2010) propose a method that reduces delay propagation by redistributing existing slack in the flight schedule. Lan et al. (2006) and Lan (2003) formulate a delay propagation problem as a mixed-integer programming problem. The computational results based on their model reveal a considerable reduction in total propagated delays and in the number of disrupted passengers. The distinction between our study and these studies is in the modeling of block times of downstream flight legs. We use a stochastic model of block times of all flight legs, whereas these papers assume deterministic block times given a random delay for the root flight. In this paper, we show that the deterministic approach used in the prior literature overestimates the impact of expected total propagated delay in the airline network, and, hence, our stochastic model is better suited for capturing network robustness.

Studies that analyze flight schedule robustness metrics mainly develop new performance measures and compare them to the existing performance measures. Caulkins et al. (1993) argue that the DOT on-time performance metric is deficient in the sense that it does not take the differences across airports into account. In a recent paper, Forbes et al. (2012) find evidence that some airlines misreport some of their flights’ arrival times so that the flight is counted as on time in the DOT on-time ranking. Besides, this misreporting (or gaming) can be aggravated by employee bonus programs that reward employees based on the airline’s position on the DOT ranking. In this paper, we develop new on-time performance metrics that will not provide airlines an undue advantage by utilizing the gaming behavior indicated in Forbes et al. (2012). Bratu and Barnhart (2005) propose two alternative flight-based delay metrics that can be more useful in reflecting passenger disruptions: the percentage of flights that are delayed more than 45 minutes and the percentage of canceled flights. In another study, Wu (2006) generates a schedule reliability index for flights that reflects both inherent and propagated delays. Using flight schedule information of an anonymous airline, it is shown that the proposed sequential optimization algorithm, which reallocates ground times between rotation legs through a simulation model, improves the overall schedule reliability. Carey (1999) also develops new measures of reliability from analytical models that can be applicable to various forms of scheduled transportation. Our approach in this area of research is to distinguish the intrinsic delay attributable to the flight itself and the total propagated delay from previous flights, which is then used to define different measures of schedule robustness.

3. Data and Model

The data for our analysis comes from several sources such as the data published by the BTS, the FAA, and aircraft manufacturers. For our analysis we consider airline on-time performance data for all domestic flights within the United States. The BTS is the primary source for these data. Additional data regarding aircraft types and other details were obtained from the aircraft registry database, the deregistered aircraft file, and the aircraft reference file for aircraft specific variables. These databases are maintained by the FAA. We also used FAA Operations and Performance Data, the DOT’s Domestic Airline Fares Consumer Report Table 6, T-100 Domestic Market data, T-100 Segment (All Carriers) data published by the BTS.

Although we only analyzed domestic flights (and their schedules), we also considered incoming and outgoing international flights while calculating the congestion variables. Inclusion of this data is critically important especially for airports with significant international flight activity. Our data set has approximately 21 million observations over a three-year period from 2005–2007 (one observation for each commercial scheduled flight flown by a major carrier in the United States).
For every flight in our data set, we compute the scheduled block time ($Q_i$) as the difference between the scheduled arrival time of the flight and its scheduled departure time. The actual block time ($D_i^L$) is computed as the difference between the actual arrival time of the flight and its scheduled departure time. The actual block time is comprised of several components including taxi-out time, en route time, and taxi-in time. Each of these components is subject to different causes of delay and the total block-time delay is the sum of all individual component delays. It is noteworthy that our definition of actual block time differs from the traditional definition mentioned earlier in §1 because the latter ignores the impact of flight delays propagated through the system and departure delays associated with the concerned flight.

An airline network is coupled by the resources that are shared across the network. In particular, any given aircraft for an airline typically flies multiple flights over the course of a day, and hence a delay on one flight can potentially spillover, or propagate, from one flight to the next. Each flight is operated by a particular aircraft, which is identified uniquely by its tail number, and an aircraft rotation is a sequence of flights operated by a particular tail number. In our model, an aircraft’s rotation begins with the first revenue flight after a major maintenance, or a layover of more than five hours at an airport, and ends with the last flight operated before the aircraft returns for its next maintenance or remains on the ground for several hours. Throughout our analysis, we assume that airline crews remain with the aircraft. We acknowledge that unavailability of crew schedule information publicly remains a limitation of our analysis. We discuss this assumption in greater detail in §6.

As mentioned earlier, any flight delay is composed of the following two components: (i) that which is attributable (intrinsic) to the flight itself, i.e., intrinsic delay, and (ii) that which was caused by a spillover from a previous flight in the aircraft’s rotation, i.e., total propagated delay. A conceptual representation of the coupling of flights on an aircraft rotation and spillovers is presented in Figure 1. We refer to the actual block time of a flight, ignoring any spillovers, as the intrinsic block time of the flight ($D_i^L$). The time duration between the next flight’s scheduled departure time, on an aircraft rotation, and the earlier flight’s scheduled arrival time is referred to as the scheduled turnaround or ground time ($G_i$). To compute $G_i$, from the on-time performance data set, we first sorted the data by airline, tail number, and scheduled departure time so that all aircraft rotations are grouped together. Then, for each flight $i$, we computed $G_i$ by subtracting the scheduled arrival time of flight $i-1$ from the scheduled departure time of flight $i$. A snapshot of one such aircraft rotation flown on September 24, 2007 by Delta Airline’s aircraft with tail number N980DL is shown in the online Appendix A available at http://dx.doi.org/10.1287/opre.1120.1146. We computed the minimum time to turn an aircraft ($T_i$) by analyzing ground times at different airports for different types

Figure 1. Illustration of spillovers in an aircraft rotation.
of aircraft for each airline. First, we grouped the actual 
ground times for each flight flown in 2005 and 2006 by 
airline, aircraft model, and departure airport. We then 
calculated the fifth percentile value (in minutes) across all 
actual ground times for each airline, aircraft model, and 
departure airport combination. The turnaround time for the 
corresponding flight $i$, in 2007, was assumed to be this 
fifth percentile value. Note that the time to turn an 
aircraft around shows considerable differences across airports 
and across aircraft models. In Tables 7 and 8 in the online 
appendix (Appendix A), we show the average turnaround 
times at some airports and some aircraft models, 
respectively. The buffer time available on ground for flight $i$, $B_i$, 
is calculated by subtracting $T_j$ from $G_i$ for all flights except 
the first flight on the rotation. Note that the $B_i$ value of the 
first flight of any iteration is assumed to be zero. Thus, the 
spillover, $L_i$, from flight $i-1$ to flight $i$ is given by

$$L_i = [(D_{i-1} - (Q_{i-1} + B_i))^+]^+.$$  

(1)

Moreover, for any flight $i$ in the rotation, we have

$$D_i = D_i - (Q_i + B_i),$$

no spillover

$$D_i = (D_i - (Q_i + B_i))^+ + [D_{i-1} - (Q_{i-1} + B_i)]^+.$$  

spill-over=L_i

(2)

The first term in (2) captures a flight’s own (intrinsic) ran-
dom block time and the second term captures any spillover 
incurred because of a delay on a previous flight in the air-
craft rotation.

We use several terms and definitions throughout this 
paper. For ready reference, we provide a detailed list of 
the terms used and their definitions in Table A.1. Keeping 
these definitions in mind, it is easy to see that there are 
three cases to consider between any two consecutive flights 
in an aircraft rotation: (i) previous flight is on time so there 
is no spillover to current flight, (ii) previous flight is not 
on time but there is no spillover to current flight because 
the lateness of the earlier flight can be absorbed by $B_j$, and 
(iii) previous flight is not on time and $B_j$ is not long enough, 
which creates the spillover $L_i$ for flight $i$. The graphical 
representation of these three cases is shown in Figure 1.

Finally, we define the arrival delay of any flight $i$ (AD), 
including any spillover from a previous flight in the aircraft 
rotation, as

$$AD_i = [D_i - Q_i]^+ = [D_i + L_i - Q_i]^+,$$

$$= [D_i - Q_i + [D_{i-1} - (Q_{i-1} + B_i)]^+]^+,$$

$$= [D_i - Q_i + [D_{i-1} - (Q_{i-1} + B_i)]^+] + [\cdots + [D_1 - (Q_1 + B_1)]^+]^+.\tag{3}$$

As can be seen in (3) the arrival delay of flight $i$ is 
dependent on the intrinsic block time of all previous flights 
on the rotation, $D_j$, $1 \leq j < i$, through the spillover vari-
bles $L_j$, $1 < j \leq i$. However, if $L_j = 0$ for some $j < i$, 
then it is equivalent to treating the rotation as beginning 
from flight $j$ instead of the first flight. Using this insight, 
the arrival delay of flight $i$ can be decomposed into the 
arrival delay impact (ADIj-i) of all flights $j < i$ on flight $i$ as follows:

$$ADI_{j,i} = AD_j - AD(L_{j+1} = 0).\tag{4}$$

Here, $AD(L_{j+1} = 0)$ represents the arrival delay of flight $i$ 
conditional on a zero spillover $L_{j+1}$ from flight $j$ to $j+1$. 
Thus, the net impact of any flight $j < i$ on the arrival delay 
of flight $i$ (NIADj,i) can be found as follows:

$$NIAD_{j,i} = AD_{j,i} - AD_{j-1,i}.\tag{5}$$

Thus, the total propagated delay (TPD) by any flight $i$, 
on an aircraft rotation, is defined as the additional delay 
cauased on flight $i$ because of the spillover from the previous 
flight $i-1$, and is computed as

$$TPD_i = [D_i + L_i - Q_i]^+ - [D_i - Q_i]^+$$

$$= AD_i - AD(L_i = 0) = AD_{i-1,i}.\tag{6}$$

Therefore, the total arrival delay of a flight, $AD$, is the 
sum of the intrinsic delay, $[D_i - Q_i]^+$, and the total propa-
gated delay, TPD.

As we indicated in §1, in the prior literature, the models 
are deterministic and focus on computing the root propa-
gated delay, which is the delay caused by the root flight. 
Specifically, AhmadBeygi et al. (2008) develop a heuristic 
approach to compute the impact of the root propagated 
delay on all future/subsequent flights that share the same 
aircraft and crew resources. This set of future/subsequent 
flights, which share the same aircraft and crew resources, 
form a propagation tree. To compute the root propagated 
delay, their deterministic approach assumes that all sub-
sequent flights in the propagation tree operate according to 
the planned schedule with no intrinsic delays, i.e., the 
actual block times equal the scheduled block times for these 
flights. The complexity in their model arises from the fact 
that they include crew schedule related delays, too. A simi-
lar deterministic approach is described in Kondo (2008) 
where the root propagated delay, because of a root flight, 
is computed using deterministic block times for all future 
flights included in the aircraft rotation. However, Kondo 
(2008) does not consider crew schedules.

The key contribution of our modeling approach lies in 
constructing a stochastic model of total delay propagation, 
which convolves a stochastic root delay with stochastic 
block times of subsequent flights on an aircraft rotation to 
compute the total propagated delay in the airline network. 
To this end, we first show that a deterministic analysis over-
estimates the root propagated delay and the total propa-
gated delay. We first formally define root propagated delay, 
(RPD), as the net impact of the root flight (which we con-
sider as first flight of the rotation) on the arrival delay of 
flight $i$ using our notation in (5):

$$RPD_i = NIAD_{i,i}.\tag{7}$$
We state the impact of the assumption of all subsequent flights ($j$)—following the root flight—operating as per schedule in Theorem 1.

**Theorem 1.** A deterministic analysis, which assumes $D_j = Q_j$, $1 < j \leq i$,

(i) overestimates the expected RPD$_i$, and also

(ii) overestimates the expected total propagated delay TPD$_i$.

The proofs of all theorems are provided in the online appendix (Appendix B).

Similarly, one can use the deterministic approach to estimate the arrival delay of flight $i$, which consists of the sum of the intrinsic arrival delay and the total propagated delay for flight $i$. The impact of this approach is stated in Theorem 2 in the online appendix (Appendix B). Thus the key difference between our approach and past literature lies in (i) a focus on total propagated delay that captures the impact of all previous flights on an aircraft’s rotation, instead of root propagated delay that only captures the impact of a delay on the first flight; and (ii) a stochastic, instead of deterministic, model of delay propagation. Thus, we relax the assumption that the actual block times on subsequent flights are deterministic, as assumed in prior literature. Theorems 1 and 2 demonstrate that our stochastic approach is well suited for estimating arrival delays and propagated delays within aircraft rotations, and hence developing other network robustness measures. We build on this idea in the rest of this paper.

To elucidate the different notations we introduced in this section and also Theorem 1, we provide an example rotation (with three flights) below in Table 1. In this example, the scheduled block times of all flights are the same whereas the realized intrinsic block times are different. A root delay of 20 minutes on the first flight is propagated into subsequent flights. However, the net impact of the root delay on flight 3 is only five minutes, because the intrinsic block-time of flight 2 and flight 3 is less than the scheduled block time, which creates an additional buffer. In contrast, the deterministic approach would have estimated the impact of root delay on flight 3 at 20 minutes. This example illustrates how the deterministic approach may overestimate the root propagated delay.

Before using the data for our analysis, we cleaned the data to eliminate several bad observations. Details of the data cleaning methodology, with examples of bad data, are available in the online appendix (Appendix A).

### 3.1. A Stochastic Model of Actual Block Time and Delay Propagation

In this subsection we focus on describing two stochastic models that capture the coupling across an airline’s network: (i) a stochastic model of the intrinsic block time, $D_i$, for a scheduled flight that captures factors such as airport capacity and schedule congestion, and (ii) a stochastic model of the actual block time, $D^a_i$, that includes delay propagation through the airline network.

To do so, first, we build on the model developed in Deshpande and Arıkan (2012), where the authors show that the intrinsic block time of a flight follows a log-Laplace distribution. Deshpande and Arıkan (2012) provide a methodology for estimating the mean, $\mu_i$, and the standard deviation, $\sigma_i$, of the log-Laplace distribution. Variables that impact the mean of the intrinsic block-time, $D_i$, distribution include flight route, airline, aircraft type, congestion at origin and destination airport, etc.

The variance of the block-time distribution is composed of three components: an origin airport shock, a destination airport shock, and a carrier dependent idiosyncratic shock. We describe essential details of the regression model developed in Arıkan (2011) in the online appendix (Appendix C).

Second, we develop a new delay propagation model that allows us to capture the ripple effect of flight delays through the network. To analyze the delay propagation in the network, we focus our attention on aircraft rotations. As shown in (2), the total block-time distribution of a flight is a convolution of the intrinsic block-time distribution of a flight and the spillover distribution.

However, the distribution of the actual block time, $D^a_i$, is not easy to characterize, because it is the convolution of a log-Laplace distribution and the spillover distribution. There does not exist a simple distribution that characterizes $D^a_i$. Hence, we make the simplifying assumption that $D^a_i$ also follows a log-Laplace distribution with parameters $\mu_i$ and $\tilde{b}^2_i$. The challenge, however, is to estimate the parameters of the resulting (assumed) log-Laplace distribution. Next, in Proposition 1, we use a novel two-moment approximation to estimate the parameters $\mu_i$ and $\tilde{b}^2_i$ of the total block-time distribution.

**Proposition 1.** Matching the first two moments of the actual block-time distribution in (2) gives

$$\tilde{b}^2_i = (-1 - 2A) + \sqrt{(4A + 1)(A + 1)},$$

where $A = \frac{\text{Var}[D^a_i]}{\text{E}[D^a_i]^2}$ and

$$\mu_i = \text{ln}(\text{E}[D^a_i](1 - \tilde{b}^2_i)).$$

| Table 1. Flight rotation example to illustrate Theorem 1. |
|---|---|---|---|---|---|---|---|---|
| $i$ | $Q_i$ | $B_i$ | $D_i$ | $L_i$ | AD$_i$ | AD$_3(L_i = 0)$ | AD$_{i,3}$ | NIAD$_{i,3}$ | RPD$_3$ | RPD$_3$ (deterministic) |
| Flight 1 | 1 | 100 | 0 | 120 | 0 | 20 | 5 | 5 | 5 |
| Flight 2 | 2 | 100 | 0 | 90 | 20 | 10 | 0 | 5 | 0 |
| Flight 3 | 3 | 100 | 0 | 95 | 10 | 5 | 0 | — | — | 5 |

Our block-time model presented in the online appendix (Appendix C), captures the correlation in block times of flight $i-1$ and $i$ through the common shock term of the common destination/origin airport of flight $i-1/i$. However, an implicit assumption behind the approximation presented in Proposition 1 is that the block-time distributions of consecutive flights are not correlated. We validate this assumption by conducting a simulation study using our data. The details of this study are available in the online appendix (Appendix D) where we show that our independence assumption is computationally reasonable for our large data set.

There are several advantages of using the above two-moment approximation. First, the above approach provides a tractable analytical expression for the actual block-time $D_i$. Second, the above approximation can be applied recursively to a string of flights on an aircraft rotation to compute the distribution of each flight on an aircraft rotation. Thus, it easily captures the cumulative effect of prior accumulated propagated delays on a given flight $i$ by keeping track of a single random variable $D_{i-1}$ of the previous flight, instead of having to keep track of the intrinsic block-time distributions of all previous flights on the rotation. Third, the above computational procedure is extremely efficient. Evaluating the actual block-time distributions and the corresponding on-time arrival probability for each of the 7.4 million flights flown in 2007 took less than an hour on our desktop computer. This is very important, because suggesting policy changes requires estimating the impact of changes to the entire network consisting of millions of flights. This would be very difficult to do with other computational approaches such as simulation of the entire system. A detailed discussion on the accuracy of the above approximation is provided in §3.2.

These aforementioned stochastic models allow us to separate any flight delay into the two components mentioned earlier, i.e., the root delay and the propagated delay. To this end, we define the two respective probabilistic metrics $\text{OTPF}_i$ and $\text{OTPF}_i^L$ as follows:

$\text{OTPF}_i = \Pr(D_i < Q_i)$. On-time probability without including the late aircraft delay $L_i$ (i.e., OTP intrinsic to the flight).

$\text{OTPF}_i^L = \Pr(D_i^L < Q_i)$. On-time probability of flight $i$ including the late aircraft delay $L_i$.

We use the stochastic models described in this section to develop network robustness measures and service levels in §4.

3.2. Accuracy of the Second Moment Approximation

In this subsection we focus on testing the accuracy of the second moment approximation proposed in the earlier subsection. Our analysis compares the two-moment model generated on-time arrival probability with the true underlying on-time arrival probability estimated using a simulation experiment. We use randomly selected 14 rotations flown in 2007 from Southwest, Delta, American, Continental, Frontier, JetBlue, and Mesa Airlines (two rotations from each major airline). We picked one short (three or four flights) and one long (between six and nine flights) rotations from each airline, so in total we have 76 flights in this simulation study.

For each flight $i$ in these rotations, 60,000 intrinsics block times are randomly generated from a log-Laplace distribution with parameters $\mu_i$ and $b_i$. Comparing the scheduled arrival time and the simulated actual arrival times, each simulated flight in the rotation is assigned an indicator variable of zero or one depending on whether the flight arrived on time or not (i.e., one represents an on-time arrival and zero represents a delayed flight). For each flight in these rotations, simulated on-time averages are calculated across 60,000 replications using the on-time indicators. Then, these simulated on-time averages (i.e., simulated on-time probabilities) are compared to the estimated on-time probabilities that are calculated by using $\mu_i$ and $b_i$ obtained from our two-moment approximation.

Tables 2 and 3 display the details of the rotations used and the results of the simulation study explained above. The maximum difference between the simulated on-time probabilities and the estimated on-time probabilities is equal to 0.1257 among these 76 flights. The median absolute difference between the simulated on-time probabilities and the estimated on-time probabilities is equal to 0.01765 and the mean absolute difference is equal to 0.0295. The simulated on-time probabilities turn out to be very close (i.e., within 4% range) to the estimated on-time probabilities for most of the flights selected.

The accuracy of the two-moment approximation method has been well studied in the multiechelon inventory literature (Graves 1985). Whereas the airline context is different than multiechelon inventory management, our model has parallels to multiechelon inventory models. Given the computational efficiency of our two-moment approximation, as compared to simulation, for performing calculations for millions of flights in our data set, we proceed with this approximation for the rest of the paper.

4. Robustness and Service Reliability Measures for Airline Networks

DOT measures a flight to be on-time as long as the flight arrives at its destination no more than 15 minutes after its scheduled arrival time. Airlines are typically ranked based on the aggregate (average) on-time performance of all the flights in their network. This DOT metric is relatively easy to compute and provides a basis for measuring airline performance. However, with such a measure, as a basis for comparison, several important issues remain unanswered. We list two important issues below:

1. The DOT aggregate measure does not distinguish between the nature of airline operations, network structures, and markets served by individual airlines.

2. The DOT aggregate measure does not account for the variability in arrival times.

3. The DOT aggregate measure does not consider the impact of different markets served by airlines.

4. The DOT aggregate measure does not consider the impact of different network structures served by airlines.

5. The DOT aggregate measure does not consider the impact of different delay propagation mechanisms served by airlines.

6. The DOT aggregate measure does not consider the impact of different delay propagation mechanisms served by airlines.

7. The DOT aggregate measure does not consider the impact of different delay propagation mechanisms served by airlines.
Table 2. Results of the simulation study (first seven rotations).

<table>
<thead>
<tr>
<th>Pos.</th>
<th>On-time prob. (estimated)</th>
<th>On-time prob. (simulated)</th>
<th>Absolute difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation 1-Southwest Airlines</td>
<td>1</td>
<td>0.7023</td>
<td>0.7001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.7086</td>
<td>0.7310</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.3670</td>
<td>0.4080</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.7952</td>
<td>0.7845</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.5400</td>
<td>0.5629</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.7003</td>
<td>0.7000</td>
</tr>
<tr>
<td>Rotation 2-Southwest Airlines</td>
<td>1</td>
<td>0.8290</td>
<td>0.8266</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.5451</td>
<td>0.5488</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2618</td>
<td>0.2986</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.4772</td>
<td>0.5799</td>
</tr>
<tr>
<td>Rotation 3-Delta Airlines</td>
<td>1</td>
<td>0.4590</td>
<td>0.4597</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.4146</td>
<td>0.4389</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.4146</td>
<td>0.4485</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2608</td>
<td>0.2734</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.1529</td>
<td>0.1602</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.1241</td>
<td>0.1409</td>
</tr>
<tr>
<td>Rotation 4-Delta Airlines</td>
<td>1</td>
<td>0.6692</td>
<td>0.6665</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.6574</td>
<td>0.6617</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.3798</td>
<td>0.3888</td>
</tr>
<tr>
<td>Rotation 5-American Airlines</td>
<td>1</td>
<td>0.5922</td>
<td>0.5909</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.2808</td>
<td>0.3043</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2712</td>
<td>0.3316</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.5735</td>
<td>0.5819</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.2084</td>
<td>0.2542</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.2425</td>
<td>0.3421</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.3615</td>
<td>0.2857</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.3521</td>
<td>0.6578</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.3270</td>
<td>0.3271</td>
</tr>
<tr>
<td>Rotation 6-American Airlines</td>
<td>1</td>
<td>0.5227</td>
<td>0.5228</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.3565</td>
<td>0.3740</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2157</td>
<td>0.2546</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2830</td>
<td>0.2862</td>
</tr>
<tr>
<td>Rotation 7-Continental Airlines</td>
<td>1</td>
<td>0.6397</td>
<td>0.6368</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.4874</td>
<td>0.5298</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.3097</td>
<td>0.3312</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.3636</td>
<td>0.4421</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.3874</td>
<td>0.4454</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.6220</td>
<td>0.6265</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.2992</td>
<td>0.2516</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.1097</td>
<td>0.1116</td>
</tr>
</tbody>
</table>

2. The DOT metric focuses on each flight separately and does not distinguish between the delays attributable to the flight itself and the additional delays because of spillovers propagated (total propagated delays) through the network. Consequently, the aggregate-level DOT metric does not allow us to clearly explain the negative externality imposed by a poorly operated resource on the rest of the network.

In this section our goal is to develop measures that help us describe and quantify the aforementioned effects of (i) intrinsic delays and (ii) total propagated delays. Let $OTP_F^{a,OT}$ represent the aggregate on-time performance computed, for any airline $a$’s network, using the DOT definition of on-time performance, i.e., using the additional 15 minutes buffer beyond the scheduled arrival time of a flight. Recollect our definitions of $OTP_F^a$ and $OTP_F^I$ in §3, which are different from the metrics published by the DOT because we do not use the 15 minute buffer beyond the scheduled arrival time. Let $OTP_F^{a,OT}$ represent the aggregate on-time performance computed for any airline $a$’s network excluding the 15 minute buffer, and $OTP_F^{a,I}$ represent the aggregate intrinsic (excluding the total propagated delays)
on-time performance computed for any airline $a$’s network. Using the distributional parameters $\hat{\mu}_i$ and $\hat{\gamma}_i$ developed in (8) and (9), we rewrite the expressions for the corresponding service metrics (represented by these probabilities) as

$$\text{OTPF}_i^l = \mathcal{L} \left( \frac{\ln(Q_i) - \hat{\mu}_i - \hat{\gamma}_i}{\hat{b}_i} \right),$$  \quad (10)

$$\text{OTPF}_i^{\text{DOT}} = \mathcal{L} \left( \frac{\ln(Q_i + 15) - \hat{\mu}_i - \hat{\gamma}_i}{\hat{b}_i} \right),$$  \quad (11)

$$\text{OTPF}_i^l = \mathcal{L} \left( \frac{\ln(Q_i) - \hat{\mu}_i - \hat{\gamma}_i}{\hat{b}_i} \right),$$  \quad (12)

where $\mathcal{L}$ represents the cumulative distribution function of the Laplace distribution with parameters $(0, 1)^7$ and $\hat{\gamma}_i$ represents the median of all residuals of the regression model described in the online appendix (Appendix C). Let $\mathcal{A}$ represent the set of airlines in our data set and $\mathcal{F}_a$ denote the set of flights operated by a particular airline $a \in \mathcal{A}$. Let $\mathcal{I}_a$ denote the set of aircraft rotations that belong to airline $a \in \mathcal{A}$.

Using (10)–(12), we define network measures of on-time performance for each airline $a$ as follows:

$$\text{OTPF}_a^l = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^l}{|\mathcal{F}_a|},$$  \quad (13)

$$\text{OTPF}_a^{\text{DOT}} = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^{\text{DOT}}}{|\mathcal{F}_a|},$$  \quad (14)

$$\text{OTPF}_a^l = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^l}{|\mathcal{F}_a|}.$$  \quad (15)

To begin, we demonstrate the inflation of on-time performance because of the use of the 15 minute buffer in the DOT definition of on-time performance. In Figures 2 and 3, we illustrate how airline performances would vary if the metric $\text{OTPF}_a^l$, i.e., without the use of the additional buffer of 15 minutes, were used instead of the metric using the 15 minute buffer ($\text{OTPF}_a^{\text{DOT}}$).

First, note that there exists a significant difference in the two OTPF metrics as shown in Figures 2 and 3. The average OTPF, without including a 15 minute buffer, across all airlines is 47.75%, whereas the average OTPF metric including a 15 minute buffer is 73.70%. The DOT OTP metric can be misleading to passengers who are not aware about the definition of OTP used by DOT. Its impact can be particularly severe on passengers who are planning itineraries with short connection times (e.g., 30 minutes).

As an example, we found flights in our data set whose OTP performance drops from 80% to 50% by removing the 15 minute buffer, which could have a significant impact on passengers making short connections. A careful analysis needs to be conducted on what metric should be reported by the DOT to the public.

Also, as is observable above, some airlines are able to use the 15 minute buffer, available to them under the DOT measure, quite effectively. For example, the OTP performance for Southwest Airlines improves under $\text{OTPF}_a^{\text{DOT}}$. On the contrary, airlines such as JetBlue, improve their position, vis-à-vis other airlines, under $\text{OTPF}_a^l$. Our observations suggest that providing an additional buffer does not address the issue of operational challenges faced by airlines, which are inherent to their network structures and markets served. The additional buffer may provide some airlines a leeway in hiding some of their inefficiencies and improving their position in the DOT on-time performance rankings. Consequently, $\text{OTPF}_a^l$ suffices as an on-time performance metric without providing airlines undue advantage.

Next, we define robustness measures for an airline flight network. Each of these measures quantify the delay externality imposed on a flight, or a resource using a set
of flights, because of tight coupling of the flights within an airline’s network.

4.1. The Average Drop in On-Time Arrivals Because of Spillovers

We begin by defining a robustness metric, $\text{OTPD}_a$, that measures the average network impact (drop) on a flight’s on-time performance because of spillovers (i.e., late arrival of aircraft from its previous flight) for any airline $a \in \mathcal{A}$:

$$\text{OTPD}_a = \text{OTPF}_a^I - \text{OTPF}_a^L \quad \forall a \in \mathcal{A}. \quad (16)$$

We illustrate the impact of spillovers ($L_i$’s) in Figure 4. Earlier, in Figure 3 we plotted $\text{OTPF}_a^I$ for each airline in our data set. In Figure 4, we plot $\text{OTPD}_a$ to illustrate the impact of spillovers. It can be inferred from Figure 4 that the impact of spillovers should not be overlooked. The average drop in on-time performance of almost all airlines is more than 5% and it is 8% or more for more than half of all airlines in our data set. Thus, from an airline’s perspective, the need of identifying and controlling total propagated delays to improve operational performance is an important issue. This fact is well known among practitioners, however, later in §5, we discuss how airlines can address this issue by allocating ground and block-time buffers differentially.

4.2. The Fraction of Expected Delay Attributable to Spillovers

Earlier in §4.1 we analyzed the impact of spillovers. We next quantify the magnitude of the total propagated
delays. For any airline \( a \in \mathcal{A} \) and any flight \( i \in \mathcal{F}_a \), this robustness metric, \( \text{FEPD}_a \), measures the average fraction of total expected arrival delay, observed on flight \( i \), which is attributable to the spillover effect from previous flights in the aircraft rotation. Essentially, it allows us to separate the average magnitude of the expected delay externality imposed on any flight \( i \):

\[
\text{FEPD}_a = \frac{\mathbb{E}(D^L_i - Q_i) + - \mathbb{E}(D_i - Q_i) +}{\mathbb{E}(D^L_i - Q_i) +} \quad \text{where } i \in \mathcal{F}_a. \quad (17)
\]

Using (17) we define \( \text{FEPD}_a \) as the average fraction of expected total arrival delay attributable to propagated delay effects, for each airline \( a \in \mathcal{A} \), and is computed as

\[
\text{FEPD}_a = \frac{\sum_{i \in \mathcal{F}_a}(\mathbb{E}(D^L_i - Q_i) + - \mathbb{E}(D_i - Q_i) +)}{\sum_{i \in \mathcal{F}_a}(\mathbb{E}(D^L_i - Q_i) +)} \quad \forall a \in \mathcal{A}. \quad (18)
\]

We plot \( \text{FEPD}_a \) in Figure 5 for all airlines in our data set. Figure 5 shows that the fraction of average delay, attributable to spillover effects, ranges from 24% to 39% for most airlines.

### 4.3. Bottleneck Flights

Recollect our definition of an aircraft rotation in §3. Let us denote an aircraft rotation by \( r \). We first define the notion of a bottleneck flight within an aircraft rotation. Let \( \text{OTPF}_{i,j} \) denote the on-time probability of flight \( j \), within a particular aircraft rotation, assuming that the aircraft rotation begins with flight \( i < j \) instead of the first flight in that rotation. Consequently, \( \text{OTPF}_{i,j} = \text{OTPF}_{j}^L \) and \( \text{OTPF}_{i,j} = \text{OTPF}_{i}^L \). Now suppose \( N(r) \) denotes the set of flights in the aircraft rotation \( r \) (string of flights in rotation \( r \)), then the network impact of flight \( i \) (based on the drop in on-time arrival probability), \( \text{NIP}_i \), is defined as the cumulative drop in on-time probability of subsequent flights in an aircraft rotation that is attributable to flight \( i \), and is computed as follows:

\[
\text{NIP}_i = \sum_{j=i+1}^{N(r)} (\text{OTPF}_{i+1,j} - \text{OTPF}_{i,j}^L). \quad (19)
\]

We now define the bottleneck flight of any rotation in terms of the network impact based on expected arrival delay. Recollect our definition of the net impact of flight \( i < j \) on the arrival delay of flight \( j \) that was represented by \( \text{NIAD}_{i,j} \) in §3. Let \( \mathbb{E}[\text{NIAD}_{i,j}] \) denote the expected value of this net impact on the arrival delay. Then, the network impact of flight \( i \) (based on the expected arrival delay), \( \text{NID}_i \), is defined as the cumulative increase in the expected arrival delay of subsequent flights in an aircraft rotation that is attributable to flight \( i \), and is computed as follows:

\[
\text{NID}_i = \sum_{j=i+1}^{N(r)} \mathbb{E}[\text{NIAD}_{i,j}]. \quad (20)
\]

A bottleneck flight, within a rotation aircraft \( r \), is then defined as the flight with the largest network impact. Let \( \text{BNFL}_r \) denote the position of the bottleneck flight within its rotation, i.e.,

\[
\text{BNFL}_r = \arg\max_{1 \leq i < |N(r)|, i \in r} \text{NIP}_i. \quad (21)
\]

Using our data set, we compute the frequency with which a flight at a particular position in any flight rotation, across all rotations for all airlines, is the bottleneck flight, as defined by (21). We denote this frequency value as \( \text{BNFL}_{rp} \), where the index \( p \) denotes the position of a flight in the aircraft rotation.

**Figure 5.** Fraction of expected delay because of propagated delays (\( \text{FEPD}_a \)) for different U.S. airlines.
Figure 6. BNFL\(_p\), computed across all airlines, for different flight positions.

10 15 20 25
(\%)
123456789 1 0 1 1 1 2 1 3 1 4 1 5 1 6
rotation. Anecdotal evidence (Ross and Swain 2007) suggests that a natural question, for airline managers, to ask is, do these bottleneck flights correspond to earlier flights in the day? We count the bottleneck flights, across all airlines, within a particular departure time slot \(t\) and denote the corresponding frequency of occurrence as BNFL\(_t\). Figure 6 plots the frequency of occurrence, BNFL\(_p\), for different flight positions, \(p\), within any aircraft rotation. As is observable in this figure, the first four flights in any rotation have a higher probability of being the bottleneck flights. Therefore, early morning flights are not significantly more likely to be bottleneck flights. Plausible reasons for midday flights becoming bottleneck flights are the increased schedule density and reduced airport capacity. Our observation suggests that airlines must tightly manage almost all their flights throughout a day to reduce their delay impact. This requires well-controlled and better managed ground and air operations.

5. Analysis and Policy Implications

The goal in this section is to compare major airlines and airports in the United States based on the robustness measures developed in §4 and to develop a few recommendations for improving the overall operational performance of the domestic air transportation system. We begin by analyzing airline on-time performance and related policy implications.

5.1. Airline On-Time Performance Analysis

On-time performance is a combination of total propagated delays and intrinsic flight delays. Earlier, in §4 we argued that it is important to separate the effect of total propagated delays and showed how we could decompose their impacts. From an airline’s perspective it is important to control both of these effects to improve the overall on-time performance.

To capture the impact of such propagated delays, in §§4.1 and 4.2, we developed two robustness measures: OTPD\(_a\) and FEPD\(_a\). These metrics allow us to measure airlines’ performances based on the average total propagated delays observed within their networks, i.e., how “tightly coupled” are these airline networks. By tightly coupled we mean how much slack is available, in the form of block- and ground-time buffers within the airline network, to absorb propagated delays.

Thus, the first issue airlines face is the allocation of ground- and block-time buffer within aircraft rotations.

Figure 7. The BNFL\(_t\), computed across all airlines, for different departure times.
Different airlines, depending on their network structures and the nature of service provided, allocate these buffers differentially. In Figures 8 and 9 we show the average ground-time buffer and block-time buffer for different airlines in our data set. The trade-offs for allocating block-time and ground-time buffers are complex. Increasing block-time buffer can increase planned costs substantially and reduce resource utilization, while decreasing block time could directly impact the flight’s on-time performance, i.e., $\text{OTPFL}_i$. Increasing ground-time buffers is relatively less expensive but decreases resource utilization. Moreover, allocation of these buffers also depends on several other factors pertaining to the airline’s flight schedule (network). For example, consider the case of Southwest Airlines.

Southwest Airlines allocates low ground-time buffers yet maintains sufficient block-time buffers. Southwest Airlines is predominantly a point-to-point carrier, i.e., does not sell a large fraction of passenger itineraries with connections, and mostly operates to and from lesser congested airports. This allows Southwest Airlines to tightly control its ground-time buffers to maintain high resource (such as aircraft and ground equipment) utilization. Allocating a slightly higher buffer on block times allows Southwest Airlines to maintain a high $\text{OTPFL}_i$. At the other extreme, consider the example of Northwest Airlines. Northwest Airlines maintains a much larger ground-time buffer yet does not allocate a large buffer on its block time. Plausible reasons are the hub-and-spoke network, congested hub airports, and a larger
fraction of connecting passenger itineraries at hub airports. For example, hub-and-spoke networks create banks in their schedule to enable passenger connections, which naturally leads to higher ground-time buffers.

The second important issue concerning airlines is the identification of bottleneck flights. Earlier, in §4.3, we developed a metric to identify a bottleneck flight by isolating the impact of a particular flight in propagating delays through the network. As we showed in Figures 6 and 7, bottleneck flights are spread throughout the day, i.e., need not correspond to the earlier flights in the day. Ross and Swain (2007) indicate that airline planners focus on the “first” flights of the day to prevent propagated delays. Although that is a good strategy, focusing on bottleneck midday flights is essential, too.

From a policy perspective, we make several recommendations based on our analysis. Our analysis shows that airlines carry very small block-time buffers for their flights, with all airlines averaging negative block-time buffers. This implies that, on the average, the scheduled block-time allocated for a flight is not sufficient to cover the actual block time. We speculate that this scheduling behavior is partly in response to how on-time performance is measured and reported by the DOT, i.e., by adding a 15 minute buffer to the scheduled arrival time. This analysis shows that the DOT should carefully evaluate which metric gets reported to the public and whether a 15 minute buffer is necessary in measuring OTP.

Moreover, the inclusion of such buffers can be misleading from a passenger’s standpoint. For example, consider a two-flight passenger itinerary wherein both the flights are of short duration and the connection time between flights is relatively small. In this case, it is possible that both individual flights may perform very well when evaluated independently using the DOT on-time performance metric, however, the passenger itinerary on-time performance may be extremely poor because of insufficient connection time. As a percentage of actual flight time, the total passenger delay can be large when both the flights in the itinerary are short haul flights. Consequently, it may be useful to consider a passenger-centric performance metric to compare airline performance in addition to single flight on-time performance measures. For example, comparing airline networks based on the on-time performance (on-time completion probability) of multileg passenger itineraries may be useful. Such a metric would also help distinguish carriers serving markets with a larger proportion of multileg itineraries versus carriers serving markets with single leg itineraries. In this regard, because our stochastic model generates the entire probabilistic distribution of actual block time, it can be very useful in computing passenger connection probabilities and multileg on-time performance measures.

If providing accurate information to passengers is of importance, particularly to those with short connections, an immediate conclusion is to modify the measurement of on-time performance of flights so that on time truly means on time, i.e., a flight should be considered delayed if it arrives past its scheduled arrival time. If this metric were to be adopted by the DOT, airlines could react by simply adding buffers to all flights (which can significantly increase their cost, as shown by Ball et al. 2010). However, if airlines carefully balance their buffers, by focusing on bottleneck flights as discussed below, a change in the DOT metric can help achieve better OTP performance without increasing buffers and associated cost.

Our model also provides guidelines to airlines on which bottleneck flights they should focus in order to minimize the network impact. We constructed a network impact metric for each flight \( t \) using our stochastic model (i.e., NID). To make their schedules more robust, airlines could allocate larger resources (crew, block times, etc.) to flights that have large network impact and less resources to nonbottleneck flights. From a policy standpoint, airlines could consider allocating ground- and block-time buffers differentially for bottleneck flights. Because airlines are typically budget constrained, in terms of the overall scheduled block-time minutes across the network, reducing the block- and ground-time buffers around nonbottleneck flights and reallocating the same to bottleneck flights could improve the overall network performance without incurring additional costs because of increased block-time buffers.

Finally, our stochastic model is well suited to isolate the impact of spillovers on schedule robustness. Separating intrinsic block time from total block time helps identify the spillover accurately for any aircraft rotation. This is important for airline practitioners because it is well known that spillovers are highly dependent on the sequence of flights flown in an aircraft rotation.

### 5.2. Airport Congestion Analysis

Airports are a critical resource in air transportation networks and impact airline operations substantially. To understand their impact on the total propagated delays through the entire airline network, we compute the average values of overall on-time performance of flights, and intrinsic on-time performance of flights, within every departure time window \( t \). We represent these by \( \overline{OTP}_t \) and \( \overline{OTP}_t \), respectively, computed across all airlines in our data set, where \( t \) represents a particular departure time window. These plots are shown in Figures 10 and 11, respectively. As is readily observable from Figure 10, the average on-time performance steadily decreases throughout the day until 6:00 p.m. and begins to improve thereafter. However, if we remove the impact of spillovers from prior flights (Figure 11), the performance of flights improves between 9:00 A.M. and 1:00 P.M. These observations suggest two important issues: (i) total propagated delays get worse as the day progresses, and, (ii) even after adjusting for spillovers, on-time performance worsens as the day progresses. Airport and airspace congestion together with the scheduling practices of most airlines contribute to this performance. In this section, we focus on the impact of
airport congestion. To this end, we compare airports in our data set on two dimensions: (i) largest network impact (volume), and (ii) largest network impact per flight. Let $\mathcal{F}$ denote the set of airports in our data and $\mathcal{F}_s$ represent the set of all flights departed from or arrived to airport $s \in \mathcal{F}$ in 2007. Then, the total network delay for each airport $s$, $\text{TND}_s$, based on the total network impact of all flights $i \in \mathcal{F}_s$, $\text{NID}_i$ (see §4.3 for definition), is computed as follows:

$$\text{TND}_s = \sum_{i \in \mathcal{F}_s} \text{NID}_i.$$  \hfill (22)

We also normalized the above total network delay metric by considering the total number of flights affected to obtain a per flight network delay for each airport $s$ ($\text{PFND}_s$), i.e.,

$$\text{PFND}_s = \frac{\text{TND}_s}{|\mathcal{F}_s|}. \hfill (23)$$

We calculated the total network delay ($\text{TND}_s$) and per flight network delay ($\text{PFND}_s$) separately for each airport in our data set using all 2007 domestic U.S. flights and then sorted all airports from highest to lowest in each network delay.
scale. Figures 12 and 13 display the top 10 airports in each list.

As is observable from Figures 12 and 13, the top five airports in the United States based on TNDs are ATL (Atlanta Hartsfield-Jackson), ORD (Chicago O’Hare), DFW (Dallas Fort Worth), PHX (Phoenix Sky Harbor), and LAX (Los Angeles); and the top five airports that have the largest PFNDs are JFK (New York JFK), HNL (Honolulu), EWR (Newark Liberty), SEA (Seattle Tacoma), and ORD (Chicago O’Hare). From a resource allocation perspective, for improving overall operational performance of the flight networks of all airlines, alleviating the congestion at these airports is essential.

To analyze the impact of airport congestion on the airline networks, we picked top four airports, namely, ORD, ATL, DFW, and PHX based on TNDs. In addition, we added two more airports, EWR² and JFK,¹⁰ considering their position in the PFND metric. We increased these airports’ capacities by 15%, separately one-by-one, and then measured the network impact of increasing their capacities. First, by increasing the capacity at any airport, all flights arriving to and departing from that airport experience reduced delays because of lower congestion at that airport. In addition to this direct impact, our model also captures the indirect impact of increasing the airport capacity, i.e., reduced spill-overs on subsequent flights flown by the aircraft on its subsequent route. Specifically, to compute how increasing capacity affects delays (and network impact) we did the following: We first recomputed the congestion variables for each flight in the network affected by the capacity change. This necessitates recomputation of the actual block-time distributions of all the flights in the network (all airlines), i.e., we need to recompute the total network impact of each flight \( i \) (NID) and then recalculate the total network delay of each airport \( s \) (TND) based on these new estimates.

In Table 4, we provide the total network delay reduction of increasing the capacity by 15% at these six airports.

As can be seen, the total network delay is different at different airports. The total network delay reduction in flight delays by adding capacity at Atlanta Hartsfield-Jackson airport is the largest both in terms of total minutes and percentage reduction. Atlanta Hartsfield-Jackson airport is followed by Chicago O’Hare airport.

We also conducted a sensitivity analysis for Chicago O’Hare airport (ORD) to examine the effects of different capacity changes on the total network delay. The percentage of capacity increase and corresponding total network delay is displayed in Figure 14. Although a careful observation of this figure reveals diminishing returns for each
additional capacity increase, the decreasing trend is not far from linear. Specifically, by using linear interpolation, we observe that for each additional 1% increase in capacity (up to 5% capacity increase), the total drop in expected arrival delays is about 30,000 minutes; and each additional 1% capacity increment between 5% and 10% capacity increase levels leads to about 26,000 minutes total reduction in expected arrival delays. This kind of sensitivity analysis can be replicated for other airports using our model and a cost-benefit model can be established as a decision support tool for policy makers. Although we do not include the cost of adding capacity at each airport, our model provides a new tool for evaluating investments in either capacity expansions at each airport, or in evaluating the benefits of new technologies (e.g., modernized landing and tracking technologies, En Route Automation Modernization program (ERAM), etc.) assessed by the FAA under its Air Traffic Control (ATC) and Next Generation (NextGen) modernization programs. Because, our model is at a micro (flight) level, the benefits of such capacity expansion proposals on network delays can be easily evaluated (ATC Modernization and NextGen: Near-Term Achievable Goals 2009).

6. Concluding Remarks

Flight delays have been a growing issue and they have reached an all-time high in recent years, with the airlines’ on-time performance at its worst level in 2007 since 1995. A recent report by the Joint Economic Committee of the U.S. Congress (Schumer and Maloney 2008) has estimated that the total cost to the U.S. economy because of flight delays was as much as $41 billion in 2007.

The goal of this paper is to build stochastic models and use publicly available data to examine the impact of airline network structures and schedules on the reliability of the air-travel infrastructure. Our analysis focuses on answering the following questions: Which are the bottleneck airports in the U.S. air-travel infrastructure (i.e., airports that cause most delay propagation)? How would increasing airport capacity at these airports alleviate delay propagation? What are the appropriate metrics for measuring the robustness of airline schedules? How could these schedules be made more robust? Which flight in an aircraft rotation is a bottleneck flight (and, hence, deserves managerial attention)?

We make several policy recommendations based on our analysis. Our analysis suggests that the DOT on-time metric can significantly inflate true on-time performance and can be misleading, particularly to passengers with short connections. If providing accurate information to passengers is desirable, then the DOT OTP metric should be modified so that on time really means on time, i.e., the 15 minute buffer provided in the on-time metric should be eliminated. A careful cost benefit analysis of this proposal, as discussed in §5, needs to be conducted. We also provide an analysis of alleviating the impact of congestion at Atlanta Hartsfield-Jackson and Chicago O’Hare airports would provide the largest network benefits to the domestic air-travel system.

From the airlines perspective, our stochastic model is well suited to isolate the impact of spillovers on schedule robustness. Separating intrinsic block time from total block time helps identify the spillover accurately for any aircraft rotation. This is important for airline practitioners because it is well known that spillovers are highly dependent on the sequence of flights flown in an aircraft rotation. Our model also provides guidelines to airlines on which bottleneck flights they should focus to minimize the

### Table 4. Change in TND, of six airports with 15% capacity increase.

<table>
<thead>
<tr>
<th>Airport</th>
<th>Current total network delay, TND, (in minutes)</th>
<th>Total network delay after 15% capacity increase (in minutes)</th>
<th>Total network delay reduction (in minutes)</th>
<th>Network delay reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORD</td>
<td>4,101,722</td>
<td>3,703,083</td>
<td>398,639</td>
<td>9.72</td>
</tr>
<tr>
<td>ATL</td>
<td>4,707,561</td>
<td>4,207,576</td>
<td>499,985</td>
<td>10.62</td>
</tr>
<tr>
<td>DFW</td>
<td>2,414,278</td>
<td>2,226,295</td>
<td>187,984</td>
<td>7.79</td>
</tr>
<tr>
<td>PHX</td>
<td>1,956,013</td>
<td>1,800,993</td>
<td>155,019</td>
<td>7.93</td>
</tr>
<tr>
<td>EWR</td>
<td>1,827,784</td>
<td>1,653,668</td>
<td>174,116</td>
<td>9.53</td>
</tr>
<tr>
<td>JFK</td>
<td>1,574,898</td>
<td>1,411,594</td>
<td>163,305</td>
<td>10.37</td>
</tr>
</tbody>
</table>

Figure 14. Total network delay of Chicago O’Hare airport at different capacity levels.
total network impact. To make their schedules more robust, airlines should allocate larger resources (e.g., crew, block times, etc.) to flights that have a large network impact.

As indicated in Ball et al. (2010), airlines have been padding their schedules and this consequently increases the planned costs of airline operations. Airlines should intelligently allocate their resources (e.g., block times) to flights. As a future avenue of research, our stochastic model can be used to allocate ground time and block times optimally across different flights without exceeding a budget constraint (i.e., total block time that can be allocated across the entire network). This optimization model can provide potential benefits in terms of reducing total network arrival delays. Another potential future research idea is to compute the passenger on-time arrival probability that is driven by airline network structures, as well as connecting itineraries taken by passengers. Because a sample of this data is also publicly available, a new metric that computes passenger on-time arrival probability can be developed and a passenger-centric schedule robustness measure can be computed.

Our modeling approach has a few limitations. We did not explicitly model delays because of crew and ground resources/gates/baggage. The lack of available data precludes us from incorporating such constraints into our models. This limitation may not be as severe as one thinks for three reasons. First, our block-time probabilistic distribution models are agnostic to the reason behind a flight disruption and hence the aggregate impact of crew related disruptions are captured in our model. Second, most domestic carriers have reserve crews readily available to cover for crew disruptions (see Sohoni et al. 2006). Finally, many airlines, recognizing the potential for crew related disruptions, now often build crew rotations so that crew members stay with the same aircraft for most of their duty periods (see Mercier et al. 2005 and Ehrgott and Ryan 2002). Our analysis, however, can be sharpened at the flight level if crew information was available.

Overall, the contribution of this paper is twofold. First, we develop stochastic models, using empirical data, to analyze the propagation of delays through air-transportation networks. A key feature of our analysis is that we utilize data available on the entire U.S. domestic network consisting of all airports and all major airlines, instead of focusing on a few airports or a few airlines. Another key feature of our model is that we conduct analysis at a micro (individual flight) level, instead of an aggregate level. Such a model is useful in evaluating the impact of changes made to one part of the system on the entire network. Also, our stochastic models allow us to develop three important robustness measures for airline networks. Second, our analysis enables us to make policy recommendations regarding managing bottleneck resources in the air-travel infrastructure, which if addressed, could lead to improvement in air-travel reliability.

**Appendix**

**Table A.1.** Description of variables.

<table>
<thead>
<tr>
<th>Label</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_i$</td>
<td>Scheduled block time of flight $i$</td>
<td>Scheduled arrival time of flight $i$ -- Scheduled departure time of flight $i$</td>
</tr>
<tr>
<td>$G_i$</td>
<td>Scheduled ground/turnaround time of flight $i$</td>
<td>Scheduled departure time of flight $i$ -- Scheduled arrival time of the previous flight $i - 1$</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Minimum time to turn the aircraft around from its previous flight for flight $i$</td>
<td>$G_i - T_i$</td>
</tr>
<tr>
<td>$B_i$</td>
<td>Buffer time available on ground for flight $i$</td>
<td>$[D_{i-1}^L - (Q_{i-1} + B_i)]^+$</td>
</tr>
<tr>
<td>$L_i$</td>
<td>Spillover caused by the previous flight on the aircraft’s rotation</td>
<td>Actual arrival time of flight $i$ -- Scheduled departure time of flight $i$</td>
</tr>
<tr>
<td>$D_i^L$</td>
<td>Actual block time of flight $i$ including the spillover $L_i$</td>
<td>$D_i^L - L_i$</td>
</tr>
<tr>
<td>$D_i$</td>
<td>Time to complete the flight excluding the spillover $L_i$</td>
<td>$Pr(D_i^L &lt; Q_i)$</td>
</tr>
<tr>
<td>OTPF$_i$</td>
<td>On-time arrival probability of flight $i$ including the spillover $L_i$</td>
<td>$Pr(D_i &lt; Q_i)$</td>
</tr>
<tr>
<td>OTPF$_i$</td>
<td>On-time arrival probability of flight $i$ excluding the spillover</td>
<td>$[D_i^L - Q_i]^+$</td>
</tr>
<tr>
<td>AD$_i$</td>
<td>Arrival delay of flight $i$</td>
<td>AD$<em>i - AD</em>{i-1} (L_{i+1} = 0)$</td>
</tr>
<tr>
<td>ADI$_{i,j}$</td>
<td>Arrival delay impact of all flights $j &lt; i$ on flight $i$</td>
<td>ADI$<em>{i,j} - ADI</em>{i-1,j}$</td>
</tr>
<tr>
<td>NIAD$_{j,i}$</td>
<td>Net impact of flight $j &lt; i$ on the arrival delay of flight $i$</td>
<td>AD$_i - [D_i - Q_i]^+$</td>
</tr>
<tr>
<td>TPD$_i$</td>
<td>Total propagated delay of flight $i$ (the additional delay caused on flight $i$ because of spillover $L_i$)</td>
<td></td>
</tr>
</tbody>
</table>
### Table A.1. Continued.

<table>
<thead>
<tr>
<th>Label</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPD&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Net impact of the first flight of the rotation on the arrival delay of flight &lt;i&gt;i&lt;/i&gt; (root propagated delay)</td>
<td>NIAD&lt;sub&gt;1,i&lt;/sub&gt;</td>
</tr>
<tr>
<td>OTPF&lt;sub&gt;a&lt;/sub&gt;&lt;sup&gt;DOT&lt;/sup&gt;</td>
<td>Aggregate on-time performance of airline &lt;i&gt;a&lt;/i&gt;’s network using DOT definition of on-time arrival performance</td>
<td>See (14)</td>
</tr>
<tr>
<td>OTPF&lt;sub&gt;a&lt;/sub&gt;&lt;sup&gt;L&lt;/sup&gt;</td>
<td>Aggregate on-time performance of airline &lt;i&gt;a&lt;/i&gt;’s network excluding the 15 minutes buffer</td>
<td>See (13)</td>
</tr>
<tr>
<td>OTPF&lt;sub&gt;a&lt;/sub&gt;&lt;sup&gt;1, j&lt;/sup&gt;</td>
<td>Aggregate intrinsic on-time performance of airline &lt;i&gt;a&lt;/i&gt;’s network</td>
<td>See (15)</td>
</tr>
<tr>
<td>OTPD&lt;sub&gt;a&lt;/sub&gt;</td>
<td>Average network impact (drop) on a flight’s on-time performance because of spillovers for airline &lt;i&gt;a&lt;/i&gt;</td>
<td>See (17)</td>
</tr>
<tr>
<td>FEPD&lt;sub&gt;a&lt;/sub&gt;</td>
<td>Average fraction of expected total arrival delay attributable to propagated delays for airline &lt;i&gt;a&lt;/i&gt;</td>
<td>See (18)</td>
</tr>
<tr>
<td>OTPF&lt;sub&gt;a&lt;/sub&gt;&lt;sup&gt;L&lt;/sup&gt;&lt;sub&gt;&lt;i&gt;i&lt;/i&gt;, &lt;i&gt;j&lt;/i&gt;&lt;/sub&gt;</td>
<td>On-time arrival probability of flight &lt;i&gt;j&lt;/i&gt; assuming that the aircraft rotation begins with flight &lt;i&gt;i&lt;/i&gt; instead of the first flight in that rotation</td>
<td>See (19)</td>
</tr>
<tr>
<td>(N(r))</td>
<td>Set of flights in the aircraft rotation &lt;i&gt;r&lt;/i&gt;</td>
<td>(</td>
</tr>
<tr>
<td>NIP&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Cumulative drop in the on-time arrival probability of subsequent flights in an aircraft rotation that is attributable to flight &lt;i&gt;i&lt;/i&gt;</td>
<td>(\sum_{j=i+1}^{N(r)} (\text{OTPF}<em>{i, j}^{2} - \text{OTPF}</em>{i}^{L}))</td>
</tr>
<tr>
<td>NID&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Cumulative increase in the expected arrival delay of subsequent flights in an aircraft rotation that is attributable to flight &lt;i&gt;i&lt;/i&gt;</td>
<td>(\sum_{j=i+1}^{N(r)} \text{E}[\text{NIAD}_{i,j}])</td>
</tr>
<tr>
<td>BNFL&lt;sub&gt;r&lt;/sub&gt;</td>
<td>Position of the bottleneck flight within its rotation &lt;i&gt;r&lt;/i&gt;</td>
<td>(\arg \max_{1 \leq i &lt;</td>
</tr>
<tr>
<td>BNFL&lt;sub&gt;p&lt;/sub&gt;</td>
<td>Cumulative frequency of the position of the bottleneck flight (p) within its rotation across all flights</td>
<td>(\text{arg max}_{1 \leq i &lt;</td>
</tr>
<tr>
<td>BNFL&lt;sub&gt;t&lt;/sub&gt;</td>
<td>Cumulative frequency of occurrence of a bottleneck flight within a particular one-hour time slot &lt;i&gt;t&lt;/i&gt;</td>
<td>(\text{arg max}_{1 \leq i &lt;</td>
</tr>
<tr>
<td>(\mathcal{F}_{s})</td>
<td>Set of all flights departed from or arrived to airport &lt;i&gt;s&lt;/i&gt;</td>
<td>(\sum_{i \in \mathcal{F}<em>{s}} \text{NID}</em>{i})</td>
</tr>
<tr>
<td>TND&lt;sub&gt;s&lt;/sub&gt;</td>
<td>Total network delay attributable to airport &lt;i&gt;s&lt;/i&gt;</td>
<td>(\sum_{i \in \mathcal{F}<em>{s}} \text{NID}</em>{i})</td>
</tr>
<tr>
<td>PFND&lt;sub&gt;s&lt;/sub&gt;</td>
<td>Per flight network delay attributable to airport &lt;i&gt;s&lt;/i&gt;</td>
<td>(\sum_{i \in \mathcal{F}<em>{s}} \text{NID}</em>{i})</td>
</tr>
</tbody>
</table>

### Endnotes

1. Airlines have reported on-time performance to DOT since 1987. Monthly reports are released in the Air Travel Consumer Report.
2. This data contains detailed information regarding departure (arrival) capacities and actual number of departures (arrivals) of Aviation System Performance Metrics airports (77 U.S. airports that have significant activity).
3. Contains information regarding the average fares for all city-pair markets that average at least 10 passengers each day.
4. T-100 data sets contain information regarding seats available and the number of enplaned passengers for each origin destination pair during any month.
5. The details of the calculation of congestion variables are available in the online appendix (Appendix E).
6. Explicit representations for \(\text{E}[\mathcal{D}_{i}^{2}]\) and \(\text{Var}[\mathcal{D}_{i}^{2}]\) are available in the online appendix (Appendix B). \(\bar{\mu}_{i} \) and \(\bar{h}_{i}^{2}\) for any flight <i>i</i> can be computed using (8) and (9) recursively starting from the first flight on any rotation, where \(\bar{\mu}_{1} = \bar{\mu}_{1} \) and \(\bar{h}_{1}^{2} = \bar{h}_{1}^{2}\).
7. \(\mathcal{L}(x) = 0.5[1 + \text{sgn}(x)(1 - \exp(-|x|))]\).
8. We excluded Hawaiian Airlines with \(\text{FEPD}_{a}\) value of 0.59 from this figure as it is an outlier on this chart.
9. EWR is the third airport behind JFK and HNL in terms of PFND<sub>s</sub>, and the sixth airport in terms of TND<sub>s</sub>.
10. JFK is the top airport based on PFND<sub>s</sub> but is not in the top five airports list based on TND<sub>s</sub>.

### Acknowledgments

The authors are grateful for the thoughtful and constructive feedback provided by the editor, Stefanos Zenios; the area editor, Moshe Kress; and three anonymous referees. The authors thank organizers and participants of the Stanford Operations Research for the Public Interest Conference for their helpful comments. The authors thank seminar participants at the University of Michigan, Northwestern University, University of Chicago, Washington University at St. Louis, University of Kansas, University of North Carolina at Chapel Hill, and Purdue University. The authors...
also thank airline operations research practitioners and participants of the AGIFORS Nice conference for providing a practical perspective. The authors thank Rajneesh Pramanik for computational assistance for this paper, as well as the Indian School of Business for research support.

References


Mazhar Arikan is an assistant professor in the Finance, Economics, and Decision Sciences Area at the University of Kansas, School of Business. His research interests include supply chain management, empirical operations management, logistics and transportation systems management, and healthcare operations.

Vinayak Deshpande is an associate professor in the operations area at the Kenan-Flagler Business School at the University of North Carolina at Chapel Hill. His research interests are in the areas of customer-centric supply chain management, supply chain collaboration/security/privacy, service parts management, and inventory management. This paper reflects his current research interest in airline flight schedules and its impact on flight delays.

Milind Sohoni is an associate professor of operations management at the Indian School of Business, Hyderabad, India. His research interest spans large-scale robust airline optimization, supply chain management, and operations management in emerging economies. He has several years of corporate experience in operations research consulting and research.