The Plight of the Late Flight

An Analysis of Delay Improvement over 25 Years

Prepared for ST599 - Big Data Analysis by:
Andrew Bernath
Heather Kitada
Ethan Edwards

Due date: 5/10/2014

Introduction and Motivation

Numb with frustration, you glare at the clock for the 30th time. A longing glance at the attendant provides no consolation; they can't usher you onto a plane that hasn't arrived yet. *Oh well,* you think, *back to Angry Birds*. Thank goodness for smartphones. Few can boast that flight delays have not affected them in some way and therefore, it is worthwhile to ask whether the air travel delay for any particular airline is increasing or decreasing over time. The purpose of this report is to address this question through analysis of 25 years of flight data from the Bureau of Transportation Statistics (BTS). Additionally, this report will explore whether sampling flights from airlines stratified by airport size provides a reasonable estimate of mean delay.

Methods Summary

For a traveler, a natural choice for a measure of total flight delay is the arrival delay, defined as the difference between the scheduled and actual arrival times. This delay is of foremost interest to a traveler as it corresponds with their ultimate goal of arriving at their destination on time. The arrival delay also has the advantage that a flight which arrives at or earlier than its scheduled time would be considered on time or early, regardless of a late departure.

The BTS data represents the entire population of flights in the United States between 1987 and 2013, so that population means and standard deviations are directly calculable. In addition to computing the population metrics, a sampling scheme was developed to estimate average delay. It was assumed that, due to coordination of air traffic control efforts, flights originating from airports of similar traffic volume would have similarities in delay patterns. Therefore, in addition to partitioning by carrier and year, the data were further stratified into three categories corresponding to origin airports with "small," medium," and "large" flight volumes. These strata were defined by taking the average number of flights from every airport over the last 25 years, ordering them, and dividing them into thirds for the three categories. Proportional samples were then taken from each strata based on the number of flights. Sample means and variances were then weighted and finite population correction was taken into account to appropriately create the desired population estimates.

Using population and sample means, the improvement in delay from year x to year y is given as the change in mean arrival delay, $\mu_y - \mu_x$. The median change in mean arrival delay for each airline was the primary metric used to assess a carrier's improvement in arrival delay.

Results and Findings

The median change in mean arrival delays for airlines that had records for more than 20 years can be found in Table 1 (*below, left*).

Obstacles

Determination of a metric with which to characterize air travel delay is not a trivial task. A compelling alternative to the choice of arrival delay is the combination of carrier and late aircraft delays, which represents delay that a carrier is able to control. Unfortunately, inconsistencies in the data prevented this metric from being useful. Delays other than arrival and departure are missing for years prior to 2003, possibly due to changes in monitoring and reporting. Several airlines only had a few years of data and one airline (Pacific Southwest Airlines, PS) had no data due to purchase by USAir in 1987.[1] Additionally, from 2007 to 2008, there is a substantial jump (upwards of a 5-6 fold increase) in several delay categories. It is likely that this was a result of several airlines making substantial changes in flight schedules due to rising fuel prices and the onset of the economic downturn.[2]

Several challenges arose from the sampling scheme. Much of this stemmed from the lengthy computing time necessary to collect a representative sample. Initially, all airports were intended to be proportionally sampled but this required an unreasonable computing time. Even after revision to three size categories,

sampling took several hours to complete. Surprisingly, the sampling computations took considerably longer to execute than those of the population.

Applications and Future Research

The data used in this study is part of a larger BTS flight data set, in which there is much potential for improvement on these results. Relaxing the assumption that arrival delay best characterizes total delay would allow us to understand how various sources of delay (such as weather, late aircraft, National Air System, etc.) affect travel and how they are changing through time. Furthermore, it has already been suggested that a measure of delay from sources that the carrier has direct control over would be useful both for carriers seeking to improve services and for passengers seeking to choose an optimal carrier. Another assumption of this report is that stratification by origin airport volume captures similarities in delay patterns. Stratification can also be performed using geographic region, time of year (months, seasons), non-stop vs. connected flights, etc and compared for accuracy. Finally, incorporating estimates of "padding" (the practice of intentionally setting extended flight times to decrease apparent delay) by airlines may provide a more accurate estimate of delay.

[1] Trinkle, Kevin. PSA History. Pacific Southwest Airlines History website. http://www.psa-history.org/about_psa/history. Published 1995. Accessed May 10, 2014.

[2] Office of the Inspector General. *Aviation Industry Performance: A Review of the Aviation Industry, 2008-2011*. Washington, DC: United States Department of Transportation; 2012. Number: CC-2012-029.