Study of the Effect of World-Wide Air Terrorism on United States Passenger Airline Performance

A. L. S.

May 24, 2015

1 Introduction

This report explores passenger airline travel within the United States in the context of terrorism attacks against airports and aircraft throughout the world. The general hypothesis is that terrorist events, possibly depending on their severity and location (domestic or international), have a significant effect on airline cancellations and delays for air travel within American borders. Data for air travel is from the Department of Transportation Bureau of Transportation Statistics [1], which maintains a database of flight information from 1987 to the present. Only domestic flights are considered. A myriad of flights statistics are examined, and an analysis of the distribution of flight delays and rates of cancellations surrounding air terrorism was performed.

Data on terrorism events are from the University of Maryland's Global Terrorism Database [2]. A distribution of the number of world-wide terror incidents involving aircraft or airports from the time period of 1970 to 2013 is shown in Figure 1.

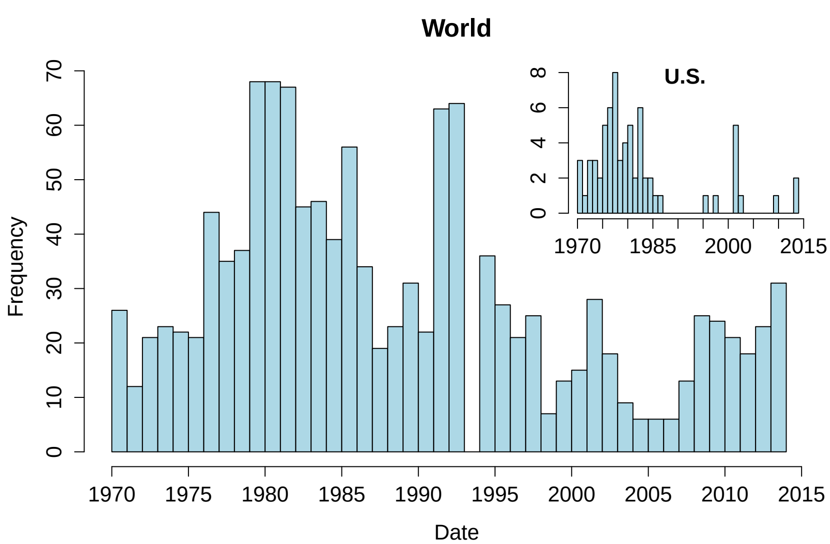

Figure 1: Distribution of terrorism events across the world involving airports or aircraft from 1970 to 2013. Inset: events occurring in the United States.

For this time period across the globe, the distribution appears mildly normal, though the clustering of events is skewed toward the 1980s. If normality is assumed, the mean date is November 1988 with quite a large standard deviation of 11.5 years. For the terror events occurring on U.S. soil, the characteristics are quite different and appear to be a hybrid of two distributions, depending on the time period. From 1970 to 1988, normality is clear. However, after 1988, the events drop sharply until 1995 when only a handful are scattered throughout the years. Within this change in the general temporal pattern, the events of September 11, 2001 are significant outliers, causing the gravest number of fatalities: more than a factor of 10 higher than any other attack of this type.

Keywords: passenger air travel, terrorist events, flight cancellations, flight delays

This analysis focuses on all air-attack events that transpired in the United States after 1987, when flight data became available from Ref. [1]. Air events that occurred world-wide (after 1987) with fatalities greater than 100 are also included to add additional depth and data, though only their effect on travel within the United States is examined. Because of its high profile and impact on security procedures, the foiled "shoe-bombing" event from December 22, 2001 on a flight originating in France (with a destination of Miami) is also incorporated. The events included in this analysis are listed in Table 1.

Table 1: Selected terrorist events from 1987 to the present.

Date	Location	Attack Type	Weapon Type	Fatalities
2001-09-11	New York City, Arlington, Shanksville, United States	Hijacking	Melee	2997
1988-12-21	Lockerbie, Great Britain	Bombing/Explosion	Explosives/Bombs	270
1989-09-19	Unknown, Niger	Bombing/Explosion	Explosives/Bombs	171
1996-11-23	Addis Ababa, Ethiopia	Hostage Taking (Barricade Incident)	Unknown	123
1989-11-27	Bogota, Colombia	Bombing/Explosion	Explosives/Bombs	107
2002-07-04	Los Angeles, United States	Armed Assault	Firearms	3
2013-11-01	Los Angeles, United States	Armed Assault	Firearms	1
2013-04-18	McCook, United States	Bombing/Explosion	Explosives/Bombs	0
2009-12-25	Detroit, United States	Bombing/Explosion	Explosives/Bombs	0
2001-12-22	Paris, France	Bombing/Explosion	Explosives/Bombs	0
1997-07-26	San Francisco, United States	Bombing/Explosion	Explosives/Bombs	0
1995-10-13	New York City, United States	Bombing/Explosion	Explosives/Bombs	0

Effects on air travel within the United States as a whole, aggregating data from all airports and airlines, will be discussed. Separating data by airlines and airports attacked by terrorists from those not directly affected, to observe if any differences arise, is beyond the scope of this report. This could be an interesting question to investigate for future studies.

2 Analysis Overview and Summary Statistics for the Selected Terror Events

For each of the twelve selected terror events, air flight data from the month of attack was selected, as well as the previous month and the subsequent month. This cache of 36 months of data covers nearly 17 million scheduled flights. Table 2 displays summary statistics for this data set. Data are summarized as a whole and also split into sets based on terror event date. Only 1 % of the flight data discussed presently occur on dates with a corresponding entry in the Global Terrorism Database. Despite this small subset, these events can have a significant impact on airline performance, as one would expect. The most significant statistical difference is manifested in the cancellation rate, skyrocketing from under 2 % to over 10 %, an increase by a more than a factor of 5, if all events are considered (the September 11, 2001 outlier is discussed in Section 2.3). Differences in statistics across different subsets of days with terror events compared to days without are discussed in Section 2.4.

It should be noted that standard deviations are given as the uncertainties in Table 2. This communicates the spread in the data more easily than the standard errors of the means, in which the standard deviations are normalized to the square root of the number of data points. Because the data in this set include a high volume of data, the standard errors of the means are significantly smaller.

2.1 Distribution of Scheduled Number of Flights

As an example of a typical summary statistic, the average number of domestic flights per day for this sample, N_{avg}, is 15227 flights per day with a standard deviation, sd_{avg}, of 1675 flights per day. To extract the 95% confidence interval, CI, of 15128 to 15326 flights per day the following equation was employed:

$$CI = N_{\text{avg}} \pm \texttt{qnorm}(0.975) * sd_{\text{avg}}/\sqrt(n) \tag{1}$$

where n is the total number of flights, $sd_{\text{avg}}/\sqrt(n)$ is the standard error of the mean, and $\texttt{qnorm}(0.975)$ is the R function that yields the position (or quantile) of the standard normal distribution at which 2.5 % of the probability

Table 2: Summary statistics for the data set in Table 1. Uncertainties are standard deviations. Average delay times are in minutes.

Statistic	Dates: All	No Terror Events	Terror Events	Terror Events (non-Sept 11)	Sept 11
Total number of scheduled flights	16795179	16614992	180187	162684	17503
Number of days	1103	1091	12	11	1
Scheduled number of flights per day	15227 ± 1675	15229 ± 1672	15016 ± 1986	14789 ± 1914	17503
Rate of flights cancelled	2.05 %	1.96 %	10.28 %	2.18 %	85.5 %
Rate of flights delayed†	41.3 %	41.2 %	48.6 %	49.2 %	10.7 %
Rate of flights delayed by > 15 min†	16.6 %	16.6 %	22.4 %	22.7 %	2.6 %
Average delay time$^{\dagger *}$	10 ± 28	10 ± 28	14 ± 34	14 ± 34	1 ± 7
Average delay if delayed†	23 ± 39	23 ± 39	28 ± 44	28 ± 44	12 ± 19
Average delay if delayed > 15 min†	51 ± 53	51 ± 53	56 ± 55	56 ± 55	36 ± 27

† Adjusted for cancellations. * Includes delays of 0 minutes for on-time or early departures.

lies above, designating 95 % of the probability within the bounds of ± qnorm(0.975). An assumption of a normal distribution is valid in this case because of the high number of flights per day. However, to be conservative, for the hypothesis tests, a Student t-distribution will be used (which approaches a normal distribution for large n). In this case, the choice of either distribution has no appreciable effect on the interval.

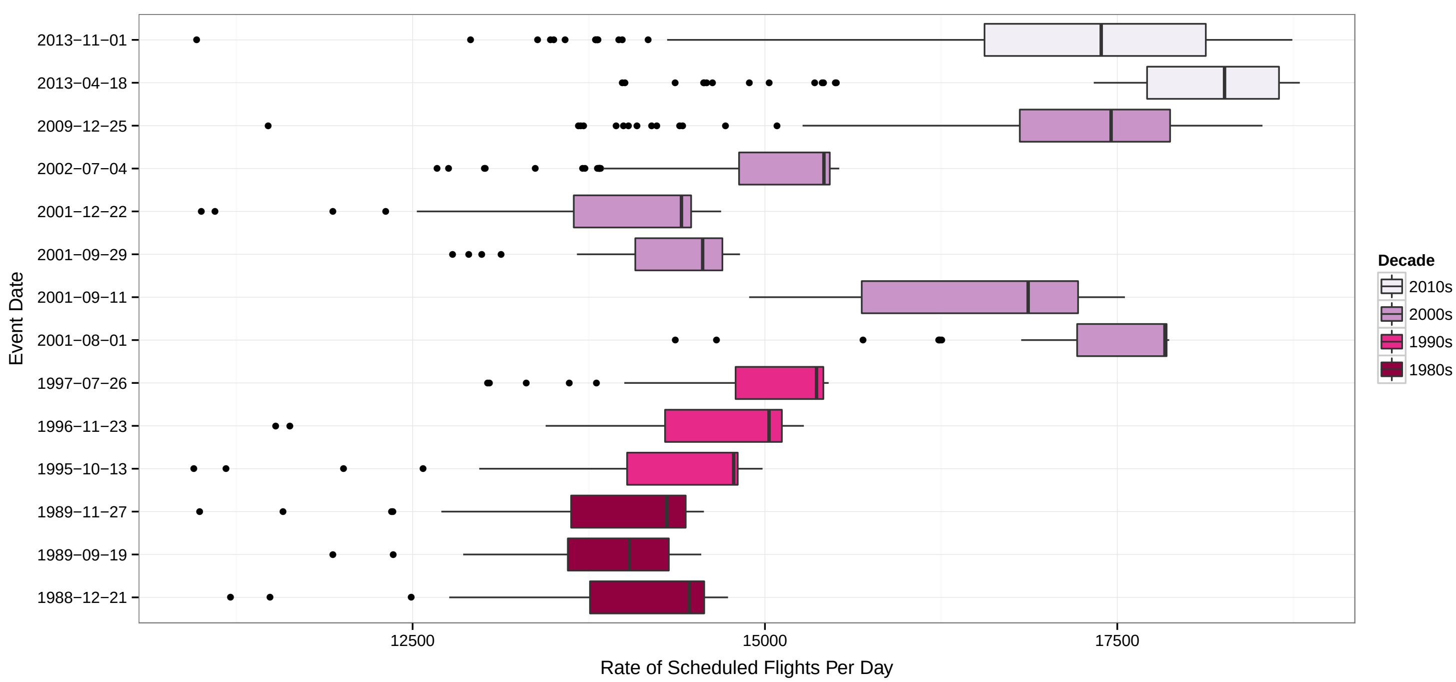

Figure 2: Distribution of daily rate of scheduled flights, grouped by terror event date, for the data set selected for this report. Each event includes three months of flight data (with the exception of the grouping of days surrounding September 11, 2001, discussed in the text).

The number of scheduled flights as a function of event date for the data considered in this analysis is shown in Figure 2. To convey the statistical properties of this parameter, each three month set corresponding to a single event date are grouped together. The September 11, 2001 event data set was split into three groups: August 1 to September 10, September 11 to September 29, and September 29 to October 31. The first group illustrates the high number of flights normally scheduled in this time period. During the second, the rate in scheduled flights after the event changes rapidly and starts to equilibrate at a significantly lower value at the start of the third group.

By mid-2001, the median rate of daily scheduled flights had increased by almost 20 % from the rate in 1987. Over the course of the two weeks after the September 11 attack, the median rate plummeted back to its value in the late '80s. Recovery was not quick but cannot be pinpointed exactly with the data set in question, as it contains gaps in which no air terror activity occurred and therefore were not included in the data set. By mid-2002, this rate had only recovered to 87 %. By 2009, it reached 98 %. It should be noted that the medians of most of these distributions (black lines in each box) are skewed quite high toward their third quartiles (right edge of the each box) and yet still have a few days with rates that deviate quite a far from the first quartiles, evidenced by the black whiskers to the left of each box. By at least 2009, the median rates became less highly skewed and more centered in their distributions, and the overall width of the distributions indicated a marked increase. Possible future extensions of this analyses could include separating out weekends and holidays to probe if and how they change the distributions.

2.2 Distribution of Delays

Some potentially puzzling statistics are the average delay times. The standard deviations reach factors of 3 larger than their mean values for some data sets. The reasoning can be understood by observing the behavior in Figure 3, which displays the distribution of delays split by day over the course of one month. The distributions are highly logarithmic in nature, not Gaussian, as most flights are not delayed by more than a few minutes. However, a non-negligible number of flights are delayed with much larger and widely varying times. The day-to-day shapes of the distribution are similar but do vary in terms of maximum delay. In these cases of skewed distributions, sometimes a more robust statistic to calculate is the median, spanned by the 0.16 and 0.84 quantiles, instead of the mean and standard deviation. The quantiles still yield a wide range of possible values under the 68 % probability envelope for the data considered in this report: 11^{+38}_{-8} minutes, compared to a mean of 27 ± 44 minutes, though at least the lower bound of the median is above the floor distribution value of 0 minutes delayed. A negative delay time (given by the lower bound of the mean in this example) is unphysical in the context of the data. A discussion of the data for September 11 follows in the next subsection.

2.3 September 11, 2001 Statistics

The tragedy of September 11, 2001 is an outlier among terror events. Because of this outlier potentially skewing the statistics, the data in Table 2 are further subdivided into this single event and all other attacks. As one can observe, the September 11 flight cancellation rate of 86 % is the driving force behind the significant increase in the cancellation rate for all events combined. However, the rate for the other events separately is still statistically significant at 2.18 %, which will be discussed in the next subsection. The rest of the statistics also behave differently from non-September 11 events, as the rate of delays and average delay times *decrease* with respect to days without terrorist activity. For this outlier, the cause of this decrease is explained by the dramatic increase in the cancellation rate, as all passenger air travel was halted, and planes remained on the ground, unable to be delayed. This reduction in delay behavior can be seen on the September 11 panel in Figure 3 and is carried over into September 13 and subsequent days, slowly transforming back to pre-attack distribution of delays as a function of time.

2.4 Significance of Differences in Parameters

The significance of each of the rate and average parameters in Table 2 was investigated with two-sided hypothesis testing. The null hypothesis assumed to be true, H_0, is that terror events are not correlated with significant changes to the parameters, x_i; the alternative hypothesis, H_α, is that terror events *do* impact the parameters in a statistically significant way:

$$\begin{aligned} H_0 &: x_i^{\mathrm{ter}} = x_i^{\mathrm{nter}} \\ H_\alpha &: x_i^{\mathrm{ter}} \neq x_i^{\mathrm{nter}} \end{aligned} \tag{2}$$

where "ter" and "nter" indicate terror and non-terror, respectively. For hypothesis tests, a test statistic must be calculated, converted into a probability called a p-value, and compared to a desired confidence level (usually 95 %) or significance level, α (usually 0.05). The p-values must be smaller than α in order to reject the null hypothesis

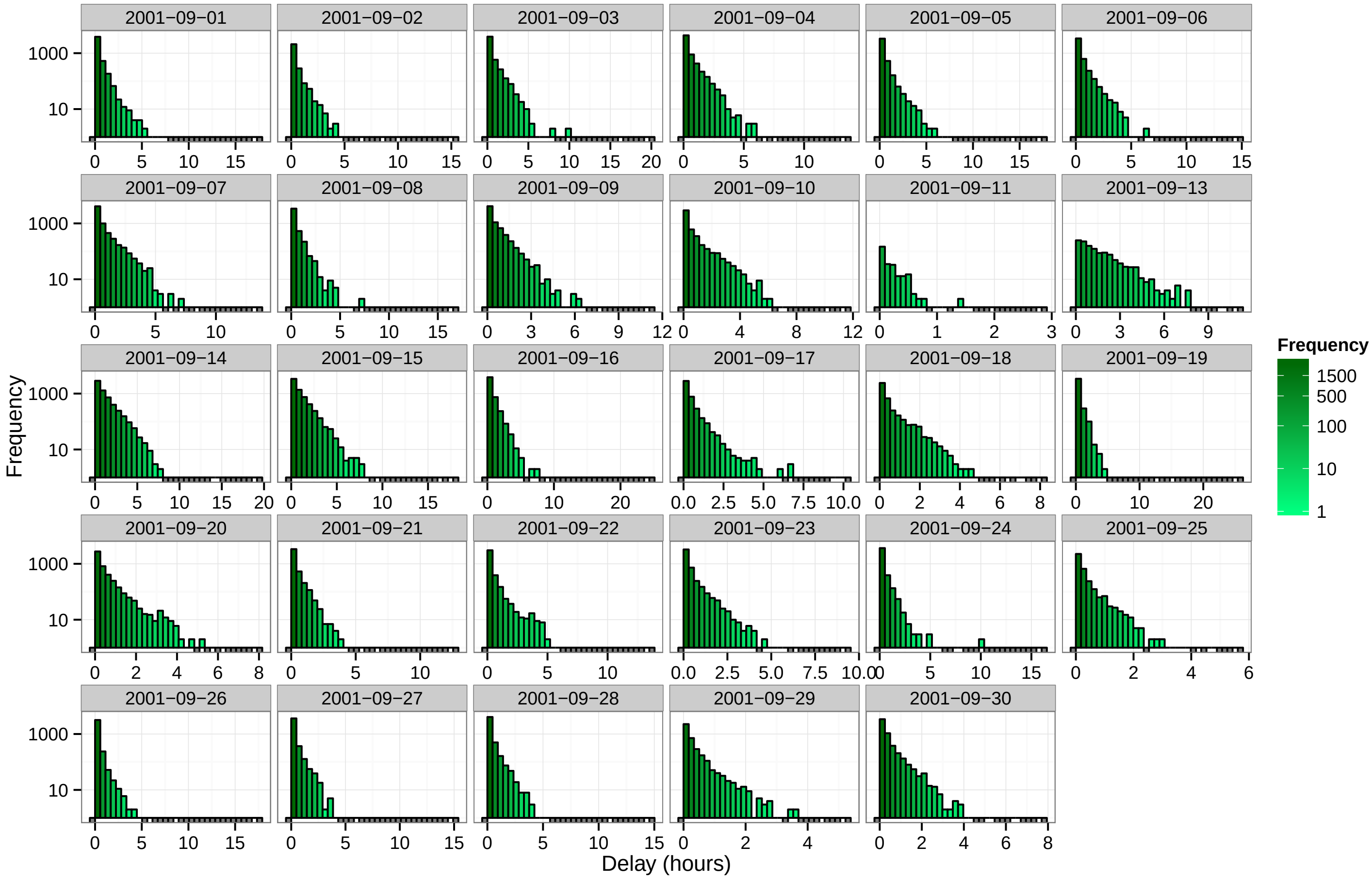

Figure 3: Distribution of delay times for September, 2001. September 12 is not shown because nearly 100 % of flights were canceled. A logarithmic scale is necessary to explore the full distribution of delays, as they are highly skewed. Bins with a single count are shown as counts below the main axis, as a logarithmic transformation yields zero and cannot be adequately plotted on a logarithmic scale. Removing them from the figure would give a false impression of the data.

and state with (95 %) certainty that terror events are correlated with flight parameters. In other words, if the p-value for a specific parameter is determined to be 0.05, the random chance that a p-value that extreme is a statistical fluctuation (when the null hypothesis is true) is only 5 %. Details on how to the test statistics and p-values are calculated are in Appendix A.1.

The p-value results for comparing the three subsets of terror events in Table 2 with the non-terror events for the three rate parameters and three average parameters in Table 2 are essentially all zero. The largest p-value is on the order of 10^{-5} for the September 11 average delay time for delays longer than 15 minutes. Because these p-values are all significantly smaller than 0.05, in each case the null hypothesis is rejected, implying that terror events do affect the flight parameters such as cancellation rate and delay time. One reason that the p-values are so small is due to the large volume of flight data, as the test statistic is directly proportional to $\sqrt{n}$, and a large test-statistic (i.e., far off from the mean in the tails of the distribution) corresponds to a very small p-value.

3 Returning to Normal

Although it may appear obvious that terror events on aircraft and airports will affect flights, perhaps a more interesting question concerns the time period to return to normal conditions (statistically speaking) immediately after such an event and how this duration correlates with event severity and country of origin. Delayed effects potentially triggered by the terror attack are beyond the scope of this discussion. The following analysis aggregated data by day and used fluctuations in the daily cancellation rate immediately following the event to explore this question, though other parameters such as delay time could be probed as well. The cancellation rate discussed in this report is defined as the total number of cancellations relative to the number of scheduled flights for each day. For this dataset [1], only cancellations occurring within seven days of the scheduled flight dates are included. Using the rate per day, not the total number of flights per day, is important as terror events can induce cancellations more than seven days in advance, which would be indistinguishable in this data set simply from fewer scheduled flights. The culprit driving this need is September 11, after which the daily scheduled number of flights drops by as much as ~ 30 % (see Figure 2).

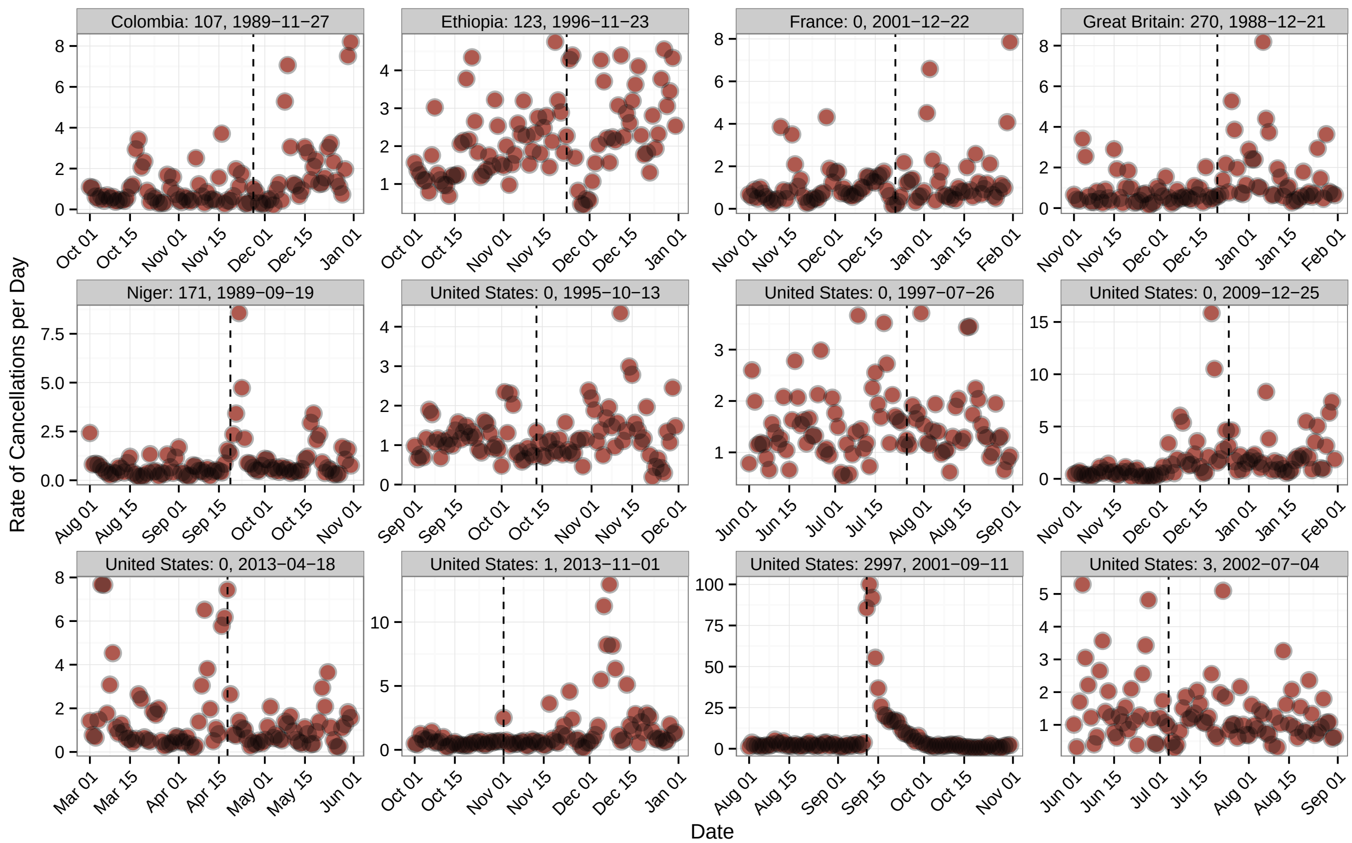

Figure 4: Daily cancellation rates of flights across the United States for the selected terrorist events, labeled by country, number of fatalities, and event date. The dashed line highlights the date of the attack.

Figure 4 illustrates the time-series data for flight cancellation rates surrounding each terror event considered in this report. Number of fatalities per event is also listed. The outlier of September 11, 2001 is once again immediately evident. Five of the twelve events appear to have the cancellation rate directly affected by the terrorism, indicated by increases above statistical fluctuations on the event date. The seven other events do not appear to have increases that are inconsistent with statistical fluctuations.

The time period to return to normal (if there was indeed a departure from normal) was determined by hypothesis testing, assuming the null hypothesis that terror events against aircraft or airports are not correlated with flight cancellation rates. Typical non-paired hypothesis testing would consist of comparing a set of data to a single hypothesized value to see if the measured results differed from this value in a statistically significant way. This analysis flips this methodology. A span of pre-event data is used to establish the “normal” cancellation rate under the null hypothesis. The event, presumably, catapults the data out of this normal region (null hypothesis is rejected). The rate of cancellations on the date of the event and subsequently dates are compared to the set of pre-event data to extract the first date after the event on which null hypothesis can no longer be rejected at the desired significance level. This would be the duration needed to return to normal. If the rate on the event date does not fluctuate outside the pre-event statistics, the analysis would find the duration to be to zero.

For a given set of pre-event data, a battery of t-tests were performed (see Appendix A.2) for the event date and for each subsequent day. In theory, the cancellation rate increases on the date of the terrorism, yielding a

Table 3: Data of the selected terror events along with the number of days after the for flight cancellations to recover to pre-terrorism levels, plotted in Figure 6. Outcome of the hypothesis test is also listed, where rejecting the null hypothesis indicates that the respective terror event can be correlated with a significant disruption in the cancellation rate at the 95 % confidence level.

Date	Country	Fatalities	Time to Return (days)	Reject H_0
2001-09-11	United States	2997	20.7 ± 0.6	True
1988-12-21	Great Britain	270	0	False
1989-09-19	Niger	171	7.3 ± 1.2	True
1996-11-23	Ethiopia	123	0	False
1989-11-27	Colombia	107	0	False
2002-07-04	United States	3	1.7 ± 2.9	False
2013-11-01	United States	1	4 ± 1.7	True
2013-04-18	United States	0	2 ± 1.7	False
2009-12-25	United States	0	0	False
2001-12-22	France	0	2.7 ± 2.3	False
1997-07-26	United States	0	1.3 ± 1.2	False
1995-10-13	United States	0	0.7 ± 1.2	False

very small p-value compared to the significance level of 0.05. This would be an indication that the null hypothesis should be rejected and that terror events *are* correlated with significant changes to flight data. It should be noted that, in theory, the p-values for dates before the event are hypothesized to be large or on the order of 0.05, further evidence that fluctuations in the cancellation rate are purely statistical and within the bounds of the null hypothesis (though it will be shown that this is not always the case given large fluctuation in the rate in general). In order to test the sensitivity of the duration to the pre-event data, three sets of differing time spans were chosen: 30 days, 15 days, and 5 days, each ending the day immediately preceding the event date.

The time period to return to normal conditions was selected as the difference in the date on which the p-value from the hypothesis test became larger than the significance level of 0.05 (thereby no longer rejecting the null hypothesis) and the event date. The final result was the average duration from the three pre-event data sets, with an uncertainty assigned to be the standard deviation.

As an example, p-values are shown in Figure 5 for the September 11 outlier as a function of date for each pre-event set (blue diamonds). The data shown before the event are for reference and context. For the set including the previous 30 days, the p-value on September 11 jumps by over 50 orders of magnitude ($> 10^{50}$). However, a large proportion of dates before that day are higher than 0.05. This may be an indication that pinning down a time period to return to normal when the p-values cross the 0.05 threshold (October 2, 21 day time to return to normal) may be an overestimate. Inflating p-values could be considered in a future analysis, or comparing p-values to a fixed significance level of 0.05 could be replaced by comparing to an average pre-event p-value to determine the time to return to normal.

For the other extreme using only 5 days of previous cancellation data, Figure 5 shows the p-values for a much greater number of pre-event days falling above the significance level threshold and yields a date of October 1 to return to normal (20 day duration). The caveat for this data set is that 5 days is probably not long enough for a statistically significant comparison given the normal spread in cancellations per day. For the middle data set encompassing 15 previous days, the duration to relax to normal conditions is the same as for the set using 30 previous days. Averaging these sets returns a duration of 20.7 ± 0.6 days (though fractional days do not necessarily have meaning since the data for this analysis is aggregated by day).

Figure 6 shows the average length of time to return to normal conditions for September 11 and for all other selected terror events as a function of event date and as a function of the number of fatalities (data is tabulated in Table 3). On the whole, only three terror events of twelve exhibit times to return which are multiple standard deviations away from zero. Three others are statistically inconsistent with zero (error bars do not overlap zero) but only by a small amount. A hypothesis test was performed to more rigorously extract for which event dates the null hypothesis of terror attacks not affecting cancellation rates could be rejected (i.e., air terror events *do* affect air flights). Results are documented in Table 3. The one-sided test compared the p-value of the mean μ and standard deviation s of the measured data to zero to see if the it was less than the significance level of 0.05 (95 %

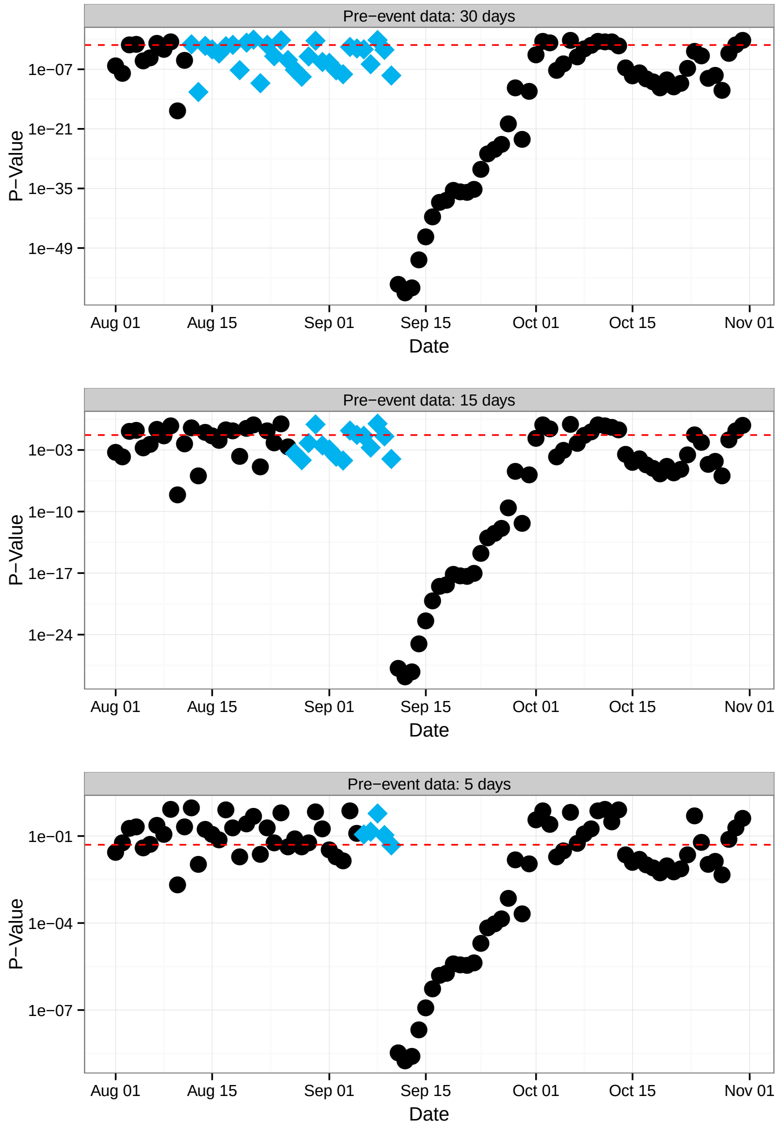

Figure 5: P-values as a function of date for the null hypothesis that cancellation rates are statistically consistent with pre-event values. Data is from the September 11 terror event. Panels show the p-values using each of the three pre-event time spans as the null hypotheses. Respective pre-event data are indicated by blue diamonds. The dashed red line designates the significance level of 0.05. Dates before the event date are shown for reference.

confidence):

$$\mathrm{p-value} = \mathtt{pt}\Big(\frac{\mu - 0}{s/\sqrt{n}}, n-1, \mathrm{lower.tail} = \mathrm{F}\Big) \overset{?}{<} 0.05 \tag{3}$$

where n is the number of dates combined to extract the mean (3), and `pt` is the R function that calculates the probability of a Student's t-distribution that corresponds to the t-statistic $(\mu - 0)/(s/\sqrt{n})$ with $n-1$ degrees of freedom. As discussed previously, if the t-statistic is in the far wings of this distribution, the area of the distribution above this point would be very small, corresponding to a rejection of the null hypothesis if the area is less than 0.05. This occurs in only the three of the events singled out previously by less rigorous methods: September 11, 2001; the event in Niger; and the event in United States on November 11, 2013, which involved a shooting that originated at an LAX TSA checkpoint and continued around the airport. Additional parameters and context are needed to put this data into sharper perspective.

September 11 remains a stark outlier, as its time to return is nearly 3 times higher than that of the next highest event (Niger). This behavior is expected because of the extreme severity of September 11, which changed air travel in our country and the around the world forever. An unplanned closure of all airspace in the United

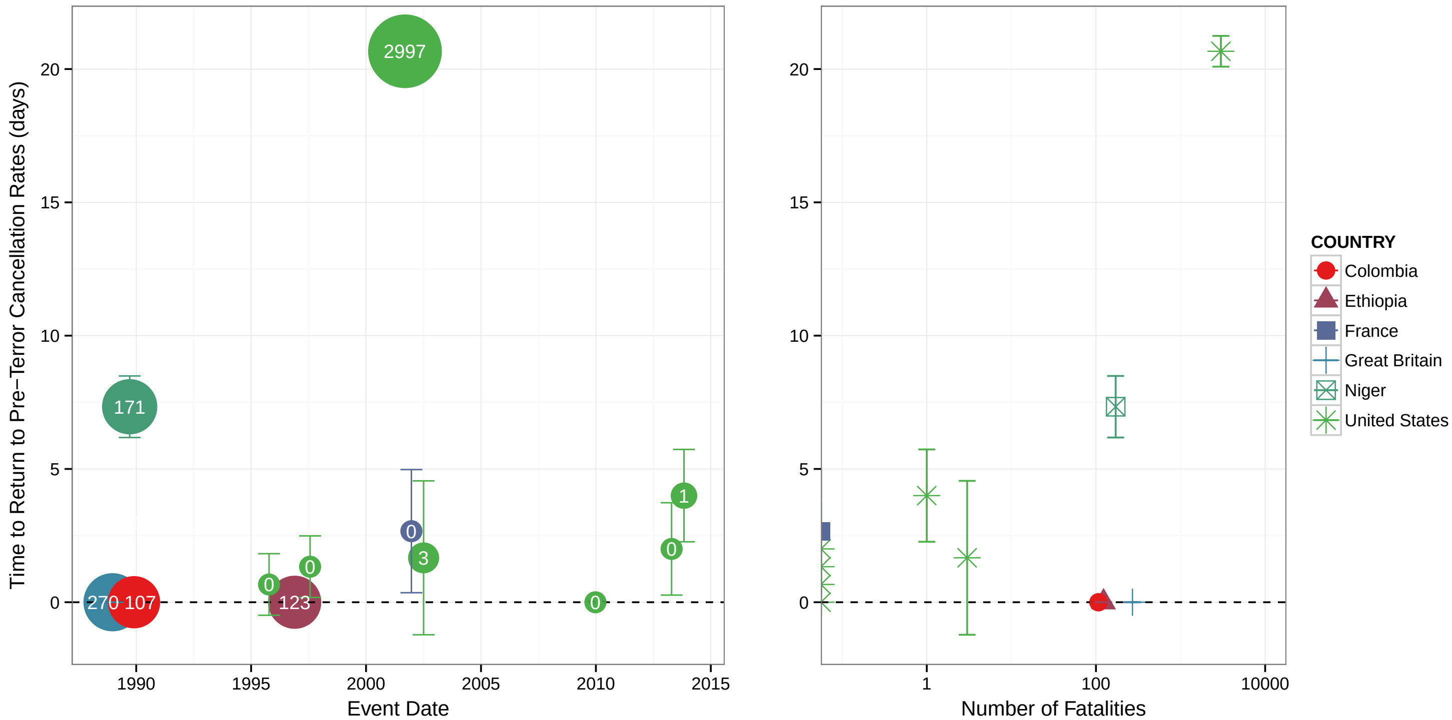

Figure 6: Number of days after a terrorist event for flight cancellations to recover to pre-terrorism levels, tabulated in Table 3. The left panel shows the temporal spread of the results, where the size of the points is proportional to the (logarithmic) number of fatalities (also overlaid). The right panel displays results a function of the number of fatalities, and data along the y-axis are events with zero fatalities.

States had been unprecedented. 86 % of flights were canceled that day, followed by 99.994 % of flights scheduled on September 12. One lone passenger airplane was allowed to depart (operated by Continental Airlines from Los Angeles to Houston). The attacks of September 11 massacred more than 10 times as many lives as the next highest in fatalities from air-related terror events. Each of these factors helps to contextualize this outlying point, as recovery from a tragedy of this magnitude takes time. According to this metric analyzed in this section, the time is 21 days for aircraft and airlines to return operating schedules within their normal statistical parameters, rescaled to reflect fewer scheduled flights. Of course, for those directly affected by September 11, the recovery time can never be quantified.

Refocusing on the other events, the time to return is zero days for several, which indicates that there were no significant changes in cancellation rate from normal cancellation conditions. This is expected when comparing this data with Figure 4, as they do not appear to have spikes in their rates on the date of the event. The flight data from dates surrounding the Ethiopia event have a constant amount of spread in the rate across the time period considered. Rate spikes can be observed in the flight data on days surrounding the events in Columbia, Great Britain, and the United States (Christmas Day, 2009), but each spike occurs at least six days before or after the identified terror attack and are most likely caused by factors not accounted for in this analysis. Therefore, one would predict a time to return equal to zero days.

Another interesting feature of the results is that events against aircraft and airlines in other countries do not seem to affect air travel significantly in the United States in 60 % of the cases, even when the destination is American soil, in the case of the Great Britain event. The events of Niger, however, did affect U.S. air travel, disrupting flights for over a week, after a flight to France from the Republic of the Congo exploded over the Nigerian Sahara Desert. The event from France is the case of the “shoe bomber”, whose destination was Miami, and triggered a statistically significant higher cancellation rate for nearly 3 days.

The effect of event severity on the daily cancellation rate is not clear cut. Although the most severe event (September 11) did trigger the longest interruption of normal flight operations, the next most severe (Great Britain with 270 fatalities) generated none. The events in Ethiopia and Columbia both caused more than 100 deaths, yet no deviation from pre-event cancellation rate was observed in United States flight data. The only other event in this data set matching that severity level is the one from Niger, where 171 people perished. Flight disruptions in the United States of over a week can be correlated with this event, as mentioned previously.

Fitting models to this data, even simple linear ones, would imply that predictions based on severity or country might be possible from this data. Considering the spread and the number of data points with a zero time of disruption, any fit would not be statistically meaningful, unless other data could be wrangled to show evidence as to why certain points should not be folded into a fit, or why others not considered in this analysis should be added. This is beyond the scope of this report.

4 Conclusions

In summary, terror attacks against aircrafts and airports both domestically and around the world are sometimes correlated with disruptions in air travel in the United States but not in the majority of cases considered in this analysis. September 11, 2001 is a major outlier with cancellation rates not returning to pre-event levels until nearly a month after the event, and effects on the number of flights scheduled lingered much longer. Of the twelve events discussed, half of them displayed times to return to pre-terror cancellation rates that were statistically above zero, though only three events were multiple standard deviations away from zero and caused a rejection of the null hypothesis that terror and cancellation rates are uncorrelated. A clear relationship with severity or country was not observed. Yet, the overarching statistics of total rates of flights canceled and delayed and changes in average delay times are affected in a statistically significant way. Flights are delayed with a rate of ~ 40 %, and delay times are highly skewed and exhibit logarithmic behavior, leading to standard deviations often more than double their means.

For future studies of this data set, one could examine the effects of air terror events on different airlines and across airports, where special consideration could give given to airlines and airports targeted in attacks. International flights to or from the United States could be considered, as well as global flight data to explore the effects of air terror in other countries. Additional terrorism events could be incorporated to increase the statistics, possibly even including events not associated with air travel. The cause of the spikes observed in Figure 4 that were not associated directly with a terror event could be investigated: what events, if any, can be correlated with those dates? Delay rates could be analyzed using the same method that was applied to cancellation rates in Section 3. Finally, delays and cancellations based on weather could be filtered out (though only after 2003, when this information became available from Ref [1]). One can only hope the trend of less and less terrorist activity with regards to passenger airlines as a function of time continues . . .

A Appendix

A.1 Hypothesis Testing for Section 2.4

For the rate parameters, a binomial distribution may be assumed, as the result for each flight is either a "yes" or "no" (e.g., if the flight in question was cancelled or not). The variance for each instance is given by $p_i(1-p_i)$, where p_i is the unknown "true" probability of a "yes" or "no" for each parameter, i. Using the Wald formulation, p_i can be approximated with its known estimated value, $\hat{p}_i$, which are the respective values given in Table 2. The test statistic, t_{stat}, and subsequent p-value can be calculated as:

$$\begin{aligned} t_{stat} &= \frac{\hat{p}_i^{\text{nter}} - \hat{p}_i^{\text{ter}}}{\sqrt{\hat{p}_i^{\text{nter}}(1-\hat{p}_i^{\text{nter}})/n}} \\ \mathrm{p-value} &= 2\,\texttt{pt}(t_{stat},\, n-1) \end{aligned} \tag{4}$$

where n is the total number of flights on respective terror events, $n-1$ are the numbers of degrees of freedom, and `pt` is an R function that calculates the probability from a Student's t-distribution corresponding to the test statistic quantile. The factor of 2 arises from a two-sided hypothesis test that the rates are simply unequal, rather than strictly lower or higher than the null rates, so the 2 accounts for the probability for the quantile to be on either side of the distribution.

A similar hypothesis test was used for the average delays, though here the data consist of vectors of delay times with one entry per flight. The test statistic and corresponding p-value calculated by:

$$\begin{aligned} t_{stat} &= \frac{\bar{x}_i^{\text{nter}} - \bar{x}_i^{\text{ter}}}{\sqrt{\frac{sd(x_i^{\text{nter}})^2}{n_{\text{nter}}} + \frac{sd(x_i^{\text{ter}})^2}{n_{\text{ter}}}}} \\ \text{p} - \text{value} &= 2\,\texttt{pt}(t_{stat},\, n_{\text{nter}} + n_{\text{ter}} - 2) \end{aligned} \tag{5}$$

where the $\bar{x}_i$'s are the means of the parameters, sd calculates their standard deviations, and n_i are the respective number of flights. In the analysis, this calculation was implemented by the R function `t.test` with unequal variances and no pairing [`t.test(`x_i^{nter}`,`x_i^{ter}`,paired=F,var.equal=F)`, where the x_i represent vectors of parameter i].

A.2 Hypothesis Testing for Time to Return to Normal of Section 3

For the hypothesis test to determine the duration of abnormal flight cancellation behavior after a terror event, comparisons were made between a set of pre-test data with an average number of cancellation rate, $\bar{R}^{\text{nter}}$, and the rate of cancellations on a single day sometime after the terror event, $R(day)^{\text{ter}}$:

$$t_{stat}(day) = \frac{\bar{R}^{\text{nter}} - R(day)^{\text{ter}}}{sd(R^{\text{nter}})/\sqrt{n_{\text{nter}}}} \tag{6}$$

This is implemented by a one-parameter t-test in R: `t.test(`R^{nter} `-` $R(day)^{\text{ter}}$`)`, where R^{nter} is a vector of cancellations per day with a length equal to the duration of the pre-event period (e.g., 30 days, 15 days, or 5 days).

References

[1] United States Department of Transportation Bureau of Transportation Statistics. http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time, 2015. [Online; accessed 16-May-2015].

[2] University of Maryland. http://www.start.umd.edu/gtd/search/, 2013. [Online; accessed 16-May-2015].