Office of Inspector General
Audit Report

EFFECTS OF THE TARMAC DELAY RULE ON FLIGHT CANCELLATIONS AND DELAYS

Office of the Secretary of Transportation

Report Number: ST-2017-003
Date Issued: October 26, 2016

U.S. Department of Transportation
Office of the Secretary of Transportation
Office of Inspector General

Memorandum

Subject: **<u>INFORMATION</u>:** Effects of the Tarmac Delay Rule on Flight Cancellations and Delays
Office of the Secretary of Transportation
Report No. ST-2017-003

Date: October 26, 2016

From: Barry J. DeWeese
Assistant Inspector General
for Surface Transportation Audits[1]

Reply to Attn. of: JA-30

To: Office of General Counsel

Since 2009, the U.S. Department of Transportation (DOT) has taken steps intended to reduce the number of lengthy tarmac delays—i.e., passengers remaining onboard an aircraft on the tarmac for extended periods of time. Following a series of high-profile instances of passengers held locked inside aircraft during lengthy tarmac delays, DOT developed the tarmac delay rule (TDR).[2] Effective April 29, 2010, the rule renders airlines liable for fines of up to $27,500 per passenger for incidents of domestic flights spending longer than 3 hours on the tarmac.[3]

After the TDR's implementation, tarmac delays longer than 3 hours fell sharply, but critics claimed that the rule would increase flight cancellations. More cancellations increase passenger costs through missed connections and time wasted. Since then, several studies have investigated the short-term effects of the TDR on cancellations, with somewhat varying results. For example, the Government Accountability Office (GAO) examined the TDR's effects on flight cancellation rates in the summer of 2010, and found it strongly increased them.[4]

[1] The OIG economists who conducted this audit cover all modes of transportation and are situated in the Office of Surface Transportation Audits.
[2] U.S. Department of Transportation, "Enhancing Airline Passenger Protections," 74 Fed. Reg. 68,983 (Dec. 30, 2009).
[3] The TDR included exceptions for aircraft with less than 30 seats and for all aircraft in cases of safety, security, or air traffic control advice. Initially, it applied only to domestic flights originating at large or medium hub airports. (Airport size designations are based on Federal Aviation Administration categories, which depend on the number of passengers boarded onto planes.) On April 25, 2011, OST announced an extension of the same provisions to domestic flights from small hub and non-hub airports, and of slightly different ones to international flights. Both took effect August 23, 2011.
[4] GAO, *Airline Passenger Protections: More Data and Analysis Needed to Understand Effects of Flight Delays* (GAO Report No. GAO-11-733), 2011.

Two other researchers, Hideki Fukui and Koki Nagata, investigated the TDR's effects through 2012 and concluded that investigations of TDR violations were critical to the TDR increasing cancellations in 2010 through 2012.[5] In addition, the Office of the Secretary of Transportation (OST) commissioned a study that found that the TDR increased cancellations in the summer of 2011, but had little to no impact on cancellations in the summer months of 2010 or 2012.[6] None of these studies examined the rule's impacts after 2012.

In the Federal Aviation Administration (FAA) Modernization and Reform Act of 2012,[7] Congress directed our office to assess the impact of DOT's rules on carriers' decisions to delay or cancel flights. To meet this mandate, the House Aviation Subcommittee of the Transportation and Infrastructure Committee and the Senate Committee on Commerce, Science, and Transportation requested that we focus specifically on the TDR. In addition, they requested that we review the OST-commissioned analysis. Accordingly, our objectives were to: (1) assess the effect of the TDR on flight cancellations, (2) assess the effect of the TDR on flight delays,[8] and (3) evaluate the OST-commissioned analysis of the TDR's impacts.

We conducted our audit in accordance with generally accepted Government auditing standards. To assess the impact of the TDR on cancellations, we employed regression and synthetic control (a type of simulation analysis) methods. Our implementation of both methods accounted for factors such as weather, airport congestion, air carrier characteristics, and route competition, helping us to isolate the TDR's impacts. Our assessment of the TDR's effects on delays, on the other hand, is comprised of descriptive data analyses, and so cannot identify causes. We conducted all our analyses using a dataset covering the fourth quarter of 2005 through the fourth quarter of 2014.[9] We are making no recommendations, and our review is not intended to be—nor were we asked to do—a regulatory impact analysis.[10] For more details on our scope and methodology, see exhibits A and B.

---

[5] Fukui, Hideki, and Koki Nagata. "Flight cancellation as a reaction to the tarmac delay rule: An unintended consequence of enhanced passenger protection." *Economics of Transportation* 3 (2014): 29-44.

[6] Econometrica, Inc. "Independent Review and Analysis of the Impact of the Three-Hour Tarmac Delay Rule," submitted to U.S. Department of Transportation January 9, 2014.

[7] Pub. L. No. 112-95 §406 (2012).

[8] We look at tarmac and gate delays for departing aircraft, under the assumption that these are the delays most likely to be influenced by the TDR. In this report, tarmac delays refer to time spent waiting on the tarmac after having left the gate and before taking off. Gate delays refer to the time between an aircraft's expected departure time and its moving onto the tarmac. We do not account for aircraft returning to the gate and then taxiing out again, as data tracking such movements is not available for periods prior to 2008.

[9] For all of our analyses, we used data averaged by month for each carrier on each route. We did not use individual flight data because that would have made analysis of monthly data for several years computationally infeasible. When we use the term "flights" in discussion of our analyses, it should be understood that we are referring to route-carrier combinations.

[10] A regulatory impact analysis, as detailed in Office of Management and Budget Circular A-4, requires additional components such as a cost-benefit analysis.

RESULTS IN BRIEF

The Tarmac Delay Rule (TDR) increased cancellation rates during the first 3 years following its implementation (May 2010–April 2013).[11] After that, the TDR did not increase cancellation rates, and cancellation rates behaved as if the TDR had never been imposed—at least through December 2014, which was the end of our period of analysis. The results from both our analytical methods exhibit this pattern. Further, flights that most frequently experienced longer tarmac delays prior to the rule's implementation had larger cancellation rate increases than other flights after the rule went into effect. For example, the rate increase for the 10 percent of flights with the greatest frequency of 2-hour or longer tarmac delays pre-TDR translated into seven additional cancellations per day in the first 3 years post-TDR. We also found that cancellation rates during the summer months increased relative to those during other seasons, and cancellation rates of flights operated by low-cost carriers increased relative to those for other carrier types.[12]

We examined two types of flight delays—tarmac and gate delays—and found that the TDR was associated with a reduction in tarmac delays, but displayed no obvious association with changes in gate delays. Lengthy tarmac delays underwent a marked decrease once OST announced the rule (December 2009), and have remained at substantially lower levels since. In addition, the frequency of longer tarmac delays fell in greater proportion than the frequency of shorter ones. For example, the percentage of flights with tarmac delays lasting at least an hour fell by 46.1 percent, while the percent of flights with tarmac delays lasting at least 2 hours fell by 70.4 percent. In contrast, while average gate delays fell to lower levels in 2009, they rebounded in 2013–2014. This rebound occurred alongside a 20 percent decline in air traffic from 2007 through 2014.

The OST-commissioned analysis of tarmac delay effects contained a number of limitations that impact its reliability as a basis for making possible policy decisions. First, the analysis potentially introduces bias by only considering cancellation impacts on flights experiencing lengthy tarmac delays after the TDR's implementation. Specifically, the analysis assumes that only cancellations occurring after post-rule tarmac delays of at least 2 hours are attributable to the TDR. This assumption rules out the possibility that air carriers may be avoiding long tarmac delays by proactively cancelling flights when facing conditions that

---

[11] We use April 29, 2010, as the date of the TDR's implementation for all flights because it provides a clear division between pre- and post-rule periods. Only 3.24 percent of route-carrier combinations in our data were consequently included among those covered by the TDR when they were not. Our findings are unchanged when we re-estimate our regressions using the exact TDR implementation dates for flights from airports not included in the initial rule.

[12] According to the International Civil Aviation Organization, a low-cost carrier is an air carrier that has a relatively low-cost structure in comparison with other comparable carriers and offers low fares and rates. The other carrier types in our analysis included legacy carriers and those which largely did not market directly to passengers. In the United States, a legacy carrier is an airline that had established interstate routes by the time of the route liberalization permitted by the Airline Deregulation Act of 1978.

might later produce them—a behavior that could in part explain the marked decline in lengthy tarmac delays. By not allowing for this possibility, the analysis limits the reliability of its results. Second, the analysis fails to account for the impacts of factors other than the TDR, such as weather and congestion, in determining the TDR's effects on cancellations. We have found such factors to be highly significant in explaining variations in airline cancellations. For example, we found that a 1 percent increase in heavy rain[13] at a departure airport added 1.86 percent to cancellation rates on average. As a result, the analysis cannot identify the specific impact of the TDR—separate from other factors—on cancellations.

OUR ANALYTICAL APPROACHES

Cancellation Rates

Many factors, such as weather and airport congestion, affect airline cancellations. Determining the effects of the TDR on cancellation rates requires accounting for the impacts of such factors. For example, the last economic downturn has been credited with contributing to a decrease in air traffic. This decrease was associated with a reduction in airport congestion, which we have found affected cancellation rates.[14] Indeed, average cancellation rates fell after the downturn. Consequently, failure to account for changes in economic activity or airport congestion while investigating changes in cancellation rates could lead to the erroneous conclusion that the TDR decreased cancellation rates.

We sought to avoid such pitfalls by using methods that allowed us to account for numerous relevant factors when determining cancellation rates. These methods allow us to identify changes in cancellation rates relative to what would have occurred in the absence of the TDR. Consequently, the effects we isolate represent changes in cancellation rates attributable to or caused by the TDR.

Our approach to identifying these effects could be thought of in the context of an experiment in which one group undergoes a treatment and another—a "control" group—does not. The treatment here would be the imposition of the TDR. The control group would track the changes that would have occurred in the absence of the rule. The difference in outcomes between the two groups would be attributable to the TDR. Since the rule applied to the vast majority of flights at the same time, there is no natural control group.

Instead, we differentiate groups according to how likely they are to be affected by "a treatment." Specifically, we expect that flights frequently experiencing long

[13] Heavy rain is defined here as precipitation in excess of the 95th percentile for an airport for the period covered by our data.
[14] *Reductions in Competition Increase Airline Delays and Cancellations* (OIG Report No. CR-2014-40), April 2014. OIG reports are available on our Web site: http://www.oig.dot.gov.

tarmac delays in the years prior to the TDR would be more likely to change their behavior following the rule’s imposition than flights that rarely experienced such delays. The impact of the TDR can then be identified as the difference in cancellations between groups experiencing more versus less frequent lengthy tarmac delays. This identification is the basis of our regression analysis and of constructing synthetic control groups for use in a simulation analysis. After examining the TDR’s effects using both the synthetic control group and regression analysis methods, we compared our results. As we will detail below, both methods produced comparable findings regarding cancellation rates.

Delays

In contrast to our analyses of cancellation rates, our analyses of delays are descriptive. They do not control for the effects of other factors. Consequently, they only permit us to discuss changes in delays associated with the TDR, as opposed to those that were caused by the rule.

In both our cancellation and delay analyses, we used the timeline that GAO found characterized the decision-making process for cancelling a flight sitting on the tarmac (2011). According to GAO, an airline and airport control tower begin discussions about potentially cancelling a flight once a tarmac delay has reached 1 hour. The discussions then continue intermittently until the tarmac delay has reached 2 hours. At that point, the flight either needs to take off or start the process of returning to the gate in order to avoid exceeding the 3-hour mark. Accordingly, we analyzed data using these two thresholds, along with an intermediate 90-minute mark, to understand the impact of the TDR.

THE TDR ELEVATED CANCELLATION RATES ONLY DURING ITS FIRST 3 YEARS

The TDR increased cancellation rates during the first 3 years following its implementation. After that, the TDR did not increase cancellation rates, and cancellation rates behaved as if the TDR had never been imposed. This pattern appears consistently in the results from both our analytical approaches. Further, during those 3 years, flights that most frequently experienced longer tarmac delays prior to the rule’s implementation had larger cancellation rate increases than other flights after the rule went into effect. We also found that cancellation rates during the summer months increased relative to those during other seasons, and cancellation rates of flights operated by low-cost carriers increased relative to those for other carrier types.

The TDR Increased Cancellation Rates in Its First 3 Years, After Which Cancellation Rates Behaved as If the Rule Had Never Been Imposed

We found that the TDR did increase cancellation rates, but only during the first 3 years after it went into effect. Table 1 shows estimates of the TDR's cancellation rate impacts over different time periods through 2014—along with rate changes preceding the rule's implementation.[15] The estimates shown represent the impact of events in each period on the cancellation rate for a carrier on a route having frequent tarmac delays prior to the rule's implementation. We focus on these effects because of our assumption that flights with a history of lengthy tarmac delays would be those most likely affected by the TDR. The changes shown are measured relative to the average cancellation rate in our data prior to August 2009, which was 1.815 percent.

Table 1. Cancellation Rate Effects Over Time

Time Period	Estimated Average Percent Change
August 2009 - December 2009	-0.0137**
January 2010 - April 2010	-0.0351**
May 2010 - April 2011	0.0154**
May 2011 - April 2012	0.0163**
May 2012 - April 2013	0.0128**
May 2013 - April 2014	-0.0575**
May 2014 - December 2014	-0.014**

Source: OIG analysis
Note: Estimates are relative to the average cancellation rate prior to August 2009.
** Significant at the 5 percent level[16]

As table 1 shows, only during the first 3 years following its implementation (May 2010–April 2013) does the TDR elevate average cancellation rates over their pre-August 2009 levels. The changes shown for each of those 3 years represent the increase in the cancellation rate that results from a marginal increase in the pre-rule frequency of tarmac delays lasting 1 hour or more. For example, from May 2010 to April 2011, a route flown by a given carrier would have seen its average cancellation rate increase to 1.830 for a 1 percent increase in the pre-rule frequency of tarmac delays lasting at least an hour. (1.815 + 0.0154 = 1.830.)

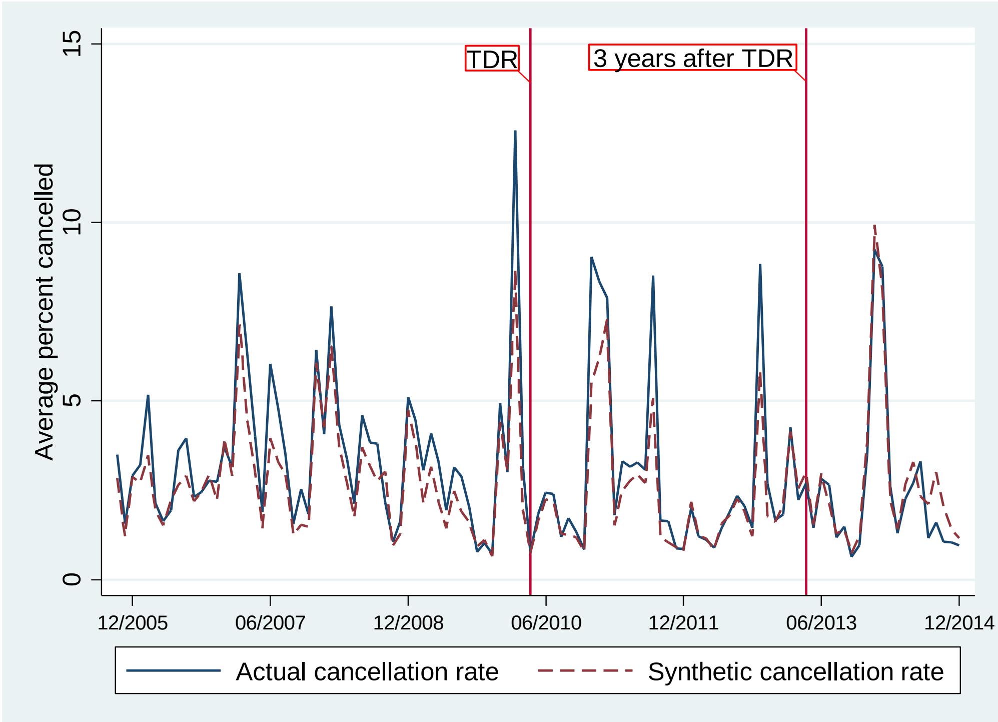
[15] The first period begins with extensive media coverage of instances of extreme tarmac delays. The second period extends from the announcement of the TDR to its implementation.
[16] A statistically significant outcome is one that is unlikely to have occurred by chance. The level of significance is the likelihood the outcome occurred by chance. A level of significance of 5 percent or less is considered indicative of a statistically meaningful relationship.

Table 1 also indicates that after 3 years (starting in May 2013), average cancellation rates for those same route-carrier combinations return to the levels they exhibited from August 2009 until the rule's implementation. In other words, the cancellation rates behaved as if the rule had never come into effect. The negative rate changes shown for the periods preceding the rule's implementation and after April 2013 may capture the impacts of a long-term trend. Throughout the study period, flights on routes with a high frequency of lengthy pre-rule tarmac delays decrease in number, while flights with low frequency of lengthy pre-rule tarmac delays showed a relative increase. Since the frequency of lengthy pre-rule tarmac delays correlates highly with cancellation rates both pre- and post-TDR, this change in the composition of flights would be expected to reduce average cancellation rates over time.

Figure 1, developed using a different approach, supports the results in table 1. It compares the path of actual average cancellation rates for a group with sizeable pre-rule delays—the 10 percent of flights with the highest frequency of tarmac delays lasting at least 90 minutes pre-TDR—with that for its synthetic control group. We chose a group with sizeable pre-rule delays because we expected that it would be most likely to change its behavior post-TDR.

The synthetic control group was constructed to mimic the selected group's responses to factors affecting cancellations prior to implementation of the TDR, such as airport congestion and weather. We then simulated the control group's response to the post-TDR values of those factors to get the path of its cancellation rate over the period following the TDR's implementation. The selected group's actual cancellation rates were affected by the TDR's implementation, while the synthetic control group's cancellation rates were not.

Figure 1. Cancellation Rates for Flights With the Greatest Frequency of Pre-Rule Tarmac Delays of at Least 90 Minutes



Source: OIG analysis

As figure 1 shows, after the TDR was in place, actual cancellation rates only exceed those of the control group during the first 3 years. In the figure, the synthetic cancellation rate (the dotted line) estimates the rate of cancellations if the TDR had never been imposed. The actual cancellation rate (the solid line) reflects the rate of cancellations in reality, including the TDR's impacts. It is expected that there would be some fluctuation between the behavior of the synthetic control group and actual cancellation rates, due to the inability of the synthetic control group to exactly mimic actual cancellation rates. However, it is significant that the graph indicates that the actual cancellation rates exceed those of the synthetic control group only in the first 3 years following the rule's implementation—suggesting that the TDR no longer increased cancellations after 3 years. This same pattern also characterizes the 10 percent of flights with the highest frequency of delays lasting at least 60 and at least 120 minutes pre-TDR.

Flights Frequently Subject to Longer Tarmac Delays Prior to the Rule's Implementation Had Larger Cancellation Rate Increases

Flights that most frequently experienced longer tarmac delays prior to the TDR had larger cancellation rate increases following the TDR's implementation. For example, flights most frequently experiencing tarmac delays greater than 90 minutes pre-TDR saw substantial and statistically significant cancellation rate increases once the rule went into effect. Moreover, the flights most frequently experiencing delays of at least 120 minutes prior to the rule experienced even greater cancellation rate impacts from the rule.

Table 2 shows the TDR's average impacts on cancellation rates over the 3 years following its implementation for the 10 percent of flights experiencing the greatest frequency of tarmac delays lasting at least 60 minutes, 90 minutes, and 120 minutes pre-rule. These impacts are larger than those shown in table 1 because table 2 reports effects for the groups having the highest frequency of pre-rule tarmac delays—which we expected to be more affected by the TDR. In contrast, table 1 averages effects across all groups. Prior to the TDR, the average cancellation rate for each group was approximately 3.4 percent, and was not statistically distinguishable from that of the other two groups.

Table 2. TDR Effects on Cancellation Rates of Flights With the Greatest Frequency of Lengthy Tarmac Delays Pre-Rule

Tarmac Delay of at Least:	Estimated Average Rate Impact
60 minutes	0.39**
90 minutes	0.49***
120 minutes	0.56***

Source: OIG analysis
Note: Effects are relative to cancellation rates prior to implementation of the TDR.
**Significant at the 5 percent level. ***Significant at the 1 percent level.

For the flights most frequently experiencing delays of at least 120 minutes prior to the TDR, the 0.56 percent increase translates into seven additional cancellations daily over the 3 years following the rule's implementation**.**

The TDR Increased Cancellations for Summer and Low-Cost Carrier Flights Relative to Cancellations for Other Groups

The TDR increased cancellations during the summer months relative to cancellations for other seasons. Table 3 shows the TDR's impacts on cancellation rates for summer and winter flights relative to those for March, April and October.[17] Summer flights were clearly affected more than flights in other seasons. (Summer months typically experience higher traffic levels.) Table 3 also shows that the impacts of the TDR on the cancellation rates of low-cost carriers are relatively much stronger and significant than those on the cancellation rates of other carrier types.

[17] Following GAO (2011), we define summer as extending from May through September, and winter as running from November through February.

Table 3. TDR Effects on Cancellation Rates by Season and Carrier Type

Season/Carrier Type	Estimated Average Rate Impact
Summer	0.0182**
Winter	-0.0261**
Low-cost carriers	0.0209**

Source: OIG analysis
Note: Effects are relative to cancellation rates prior to implementation of the TDR.
** Significant at the 5 percent level

LENGTHY TARMAC DELAYS HAVE FALLEN MARKEDLY WHILE GATE DELAYS CHANGES HAVE SHOWN NO CLEAR PATTERN

We examined two types of flight delays—tarmac and gate delays[18]—and found that the TDR was associated[19] with tarmac delay reductions, but did not have a clear association with changes in gate delays. Lengthy tarmac delays underwent a marked decrease once OST announced the TDR (December 2009), and have remained at substantially lower levels since. In contrast, while average gate delays also declined in 2009, they rebounded in the period from 2013 through 2014.

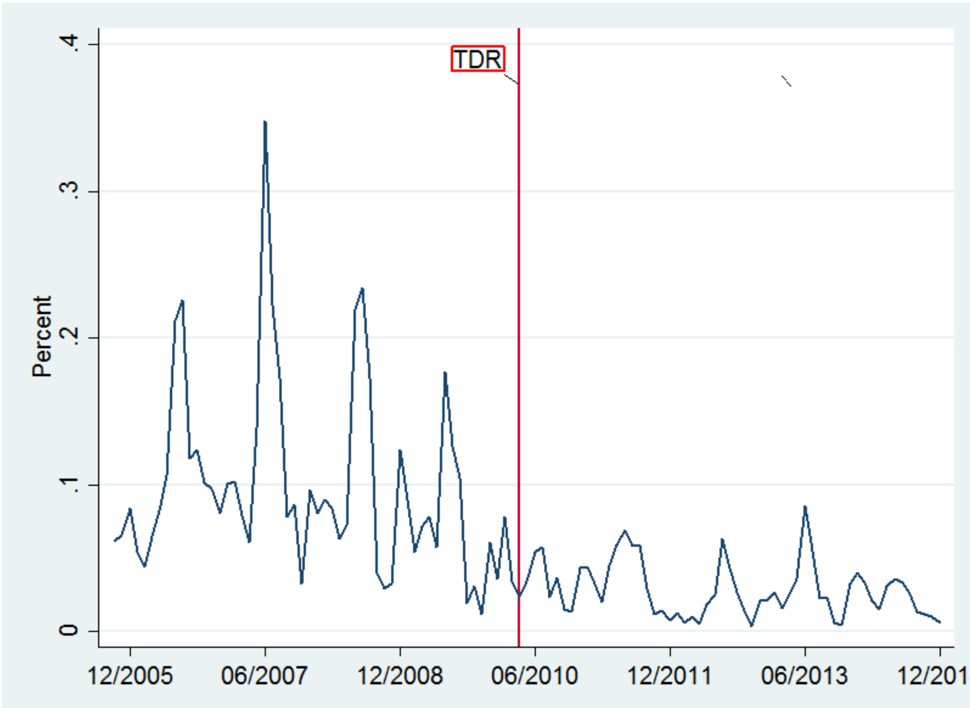
The TDR Has Been Associated With Marked Reductions in Lengthy Tarmac Delays

When OST announced the TDR in December 2009, lengthy tarmac delays fell significantly. For example, the percentage of flights with tarmac delay that exceeded 3 hours prior to the announcement was 0.0172 percent. As of December 2009, the percentage had already declined to 0.0044 percent, and since then the average monthly rate has been 0.0003 percent.

This was even true of tarmac delays that were well short of the 3 hour limit set by the rule. Figures 2 and 3 show the decline in tarmac delays lasting at least 120 and 90 minutes, respectively. These reductions have been maintained through the end of the period covered by our data (December 2014). In addition, longer tarmac delays fell in greater proportion than shorter ones. For example, the percentage of flights delayed at least 60 minutes fell 46.1 percent on average. The comparable decline in flights with longer delays of at least 120 minutes was 70.4 percent.

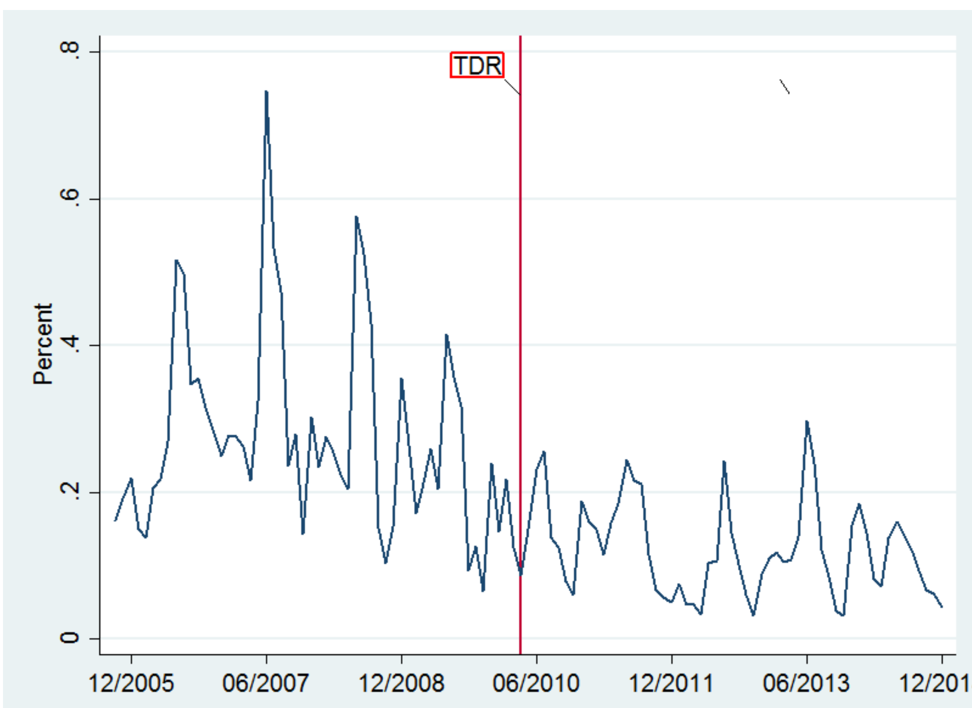
[18] Gate delay is defined as the difference between scheduled and actual departure times from the gate.
[19] In statistics, association is any statistical relationship, whether causal or not, between two random variables or two sets of data. All of the figures and statistics related to delays in this report were developed using data weighted by number of flights.

Figure 2. Percentage of Flights With Tarmac Delays of at Least 120 Minutes



Source: OIG analysis of Bureau of Transportation Statistics (BTS) data

Figure 3. Percentage of Flights With Tarmac Delays of at Least 90 Minutes

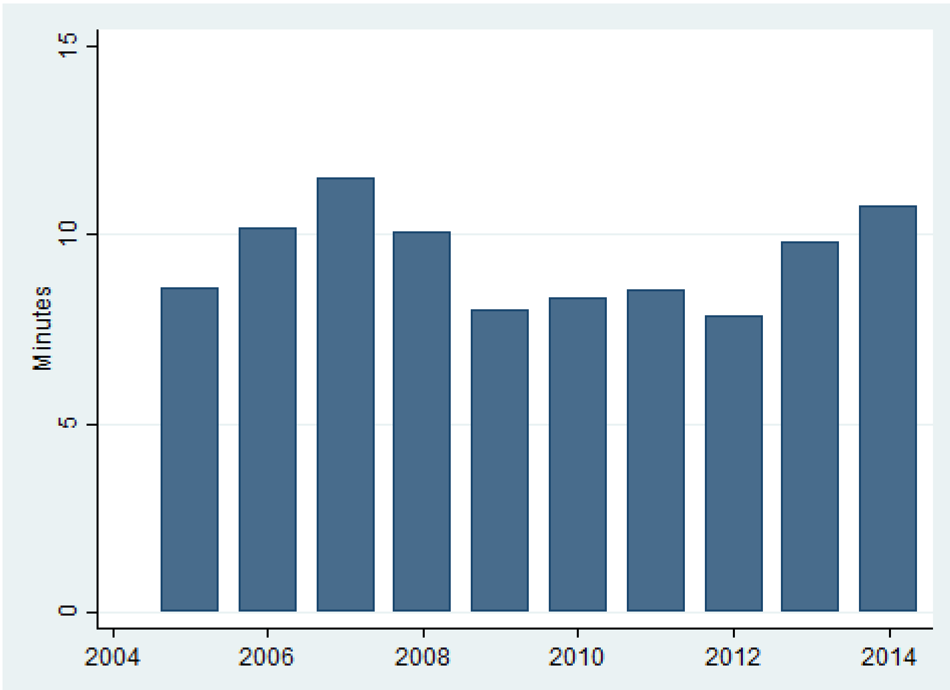


Source: OIG analysis of BTS data

Changes in Average Gate Delays Did Not Have an Obvious Association With the TDR

We found no obvious association between changes in average gate delays and the TDR. Figure 4 indicates that average gate delays were generally lower in 2009 through 2012, but then returned to their former levels. Airports of all sizes exhibited this pattern.

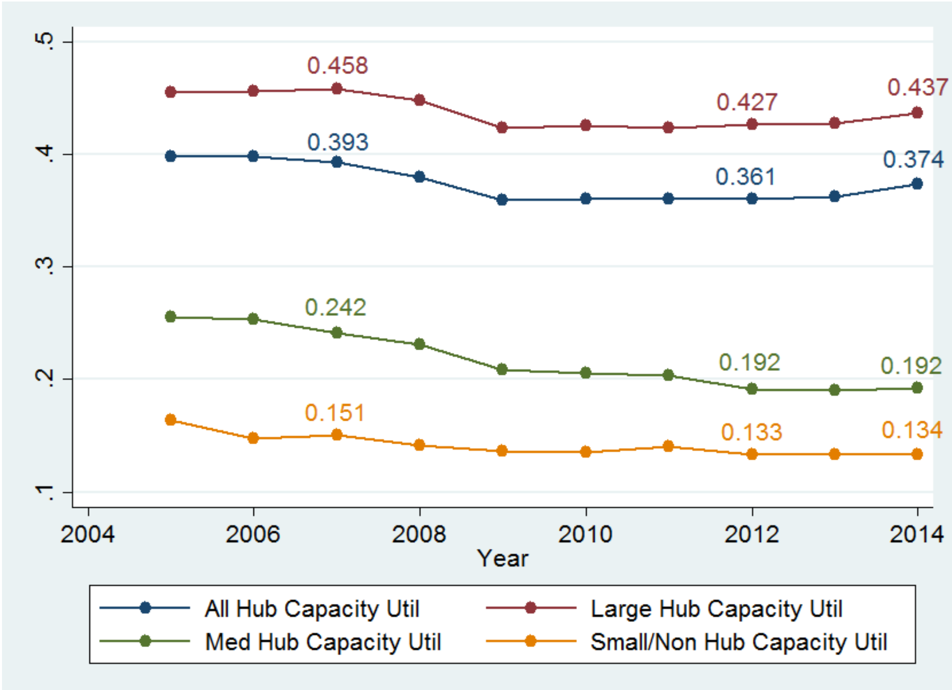
Figure 4. Annual Average Minutes of Gate Delays



Source: OIG analysis of BTS data

At the same time, air traffic fell and remained at lower levels until the end of the study period. The number of annual flights declined 20 percent from 2007 to 2014. Medium-sized airports experienced the greatest reduction in flights (34 percent). Congestion also fell during this period. Figure 5 shows FAA's measures of airport capacity utilization—which are indicators of airport congestion—by airport size.

Figure 5. Annual Average Capacity Utilization by Airport Size



Source: OIG analysis of FAA data

THE OST-COMMISSIONED ANALYSIS HAS LIMITATIONS THAT IMPACT ITS RELIABILITY

Several factors limit the reliability of OST's commissioned analysis of the TDR's effects on flight delays. First, the analysis introduces bias through its assumption that it is only necessary to consider cancellation impacts on flights experiencing lengthy tarmac delays after the TDR's implementation. Second, the analysis fails to account for the impacts of factors other than the TDR, such as weather and congestion, in determining the TDR's effects on cancellations. As a result of these limitations, the analysis does not provide reliable information on which to base potential policy decisions.

The Primary Assumption Underlying the OST-Commissioned Analysis Biases the Results

The OST-commissioned analysis assumes that only cancellations occurring after tarmac delays of at least 2 hours are attributable to the TDR in the post-rule period. However, this assumption is problematic because tarmac delays of this length have dropped dramatically post-TDR. The decrease in post-TDR lengthy tarmac delays raises the question of what air carriers have done to avoid these delays. One possibility is that air carriers proactively cancelled flights facing conditions that might later result in a lengthy tarmac delay.

The primary assumption of the OST-commissioned analysis rules out the possibility that the reductions in lengthy tarmac delays were related to the increased cancellations. In this way, it potentially biases results. In addition, post-rule tarmac delays cannot serve as stable indicators of the potential for the rule to impact cancellations because the rule potentially affected both tarmac delays and cancellations at the same time. Consequently, the study's impact calculations based on these delays are biased.[20]

The OST-Commissioned Analysis Fails To Account for Factors Such as Weather and Congestion

The OST-commissioned analysis does not account for other factors that play a significant role in cancellation rates, such as weather and congestion. We have found that factors such as these are highly significant in explaining variations in airline cancellations. For example, we found that each 1 percent increase in heavy rain at a departure airport increased cancellation rates 1.86 percent on average.

Specifically, the OST-commissioned analysis determines the impact of the TDR by comparing cancellations across airports on groups of days with or without lengthy delays, year for year. However, that comparison does not account for various factors that could be contributing to those delays or cancellations on those particular days. For example, suppose summer 2009 experienced many delays and cancellations on days when there was heavy rain, but in summer 2010, there were many delays and cancellations on the same days, yet no rain. The OST analysis would claim the TDR had a minimal or no effect on cancellations because the number of the cancellations or delays appears the same between the 2 years—even though it is likely the rain in 2009 influenced many decisions to cancel flights. By omitting contributing factors such as weather or other causes of delays, the OST-commissioned analysis cannot identify how many cancellations were due to the TDR.

CONCLUSION

Lengthy tarmac delays significantly inconvenience the flying public; however, flight cancellations are also costly in a number of ways for both passengers and the airline industry. Understanding the impact of DOT's tarmac delay rule on flight cancellations is an important step in determining whether the rule has improved flying conditions for the public. The findings of this audit agreed with those of other analyses in identifying a short-term increase in cancellations following the TDR.[21] However, our audit differed from those analyses in also being able to examine longer term effects—specifically over 4.5 years following the TDR's

[20] This is called endogeneity bias. It occurs when an outcome of interest and a factor used to explain the outcome are determined simultaneously.
[21] Fukui and Nagata (2014), GAO (2011).

implementation. As a result, our analysis was able to show that the TDR decreased lengthy tarmac delays, and that it did increase cancellations—but its impact on cancellations only lasted 3 years.

AGENCY COMMENTS AND OIG RESPONSE

We provided a draft of this report to OST on September 22, 2016. On October 18, the Agency informed us that it did not intend to submit a written response to our report. During the review, we discussed our findings and methodology with OST representatives. Where appropriate, we incorporated OST's comments and input received during our meetings with the Agency. As our report did not include any recommendations, we are not requesting further action from OST.

We appreciate the courtesies and cooperation of representatives of the Office of the Secretary of Transportation during this audit. If you have any questions concerning this report, please call me at (202) 366-5630 or Betty Krier, Program Director, at (202) 366-1422.

#

cc: DOT Audit Liaison, M-1
FAA Audit Liaison, AAE-100

EXHIBIT A. SCOPE AND METHODOLOGY

We conducted our work from February 2015 through September 2016 in accordance with generally accepted Government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

The FAA Modernization and Reform Act of 2012 mandated that our office assess the effect of the rules and regulations of the Department of Transportation on the decisions of air carriers to delay or cancel flights. Staff from both the House Aviation Subcommittee of the Transportation and Infrastructure and the Senate committee on Commerce, Science, and Transportation directed our office to focus on the Tarmac Delay Rule (TDR) and also asked us to evaluate the Office of the Secretary of Transportation's (OST) commissioned analysis of the rule. Accordingly, our objectives were to: (1) assess the effect of the TDR on flight cancellations, (2) assess the effect of the TDR on flight delays, and (3) evaluate the OST-commissioned analysis of the TDR's impacts.

We used two analytical methods to assess the TDR's effects on cancellations. First, we performed an econometric analysis with four models, which allowed us to isolate the effect of the TDR from other factors affecting cancellation. In all models, the dependent variable was the percentage of cancelled flights. We also developed a list of factors, or control variables, to account or control for variation in the airline cancellation behavior that was not associated with the TDR (e.g., weather, congestion, etc.). We developed the list of control variables based on our previous audit addressing competition effects on airline flight delays and cancellations.[22] We also drew from economics literature on factors significantly affecting airline cancellation behavior. The list of control variables was the same for all four models, except for controls designed to capture responses to the TDR by time period, season of the year, and carrier type.[23]

Our second approach used a synthetic control group. This method simulates the behavior of a target group in the absence of a policy (e.g., the TDR). For this purpose, a synthetic control group is formed by combining pre-policy behavior

[22] *Reductions in Competition Increase Airline Delays and Cancellations* (OIG Report No. CR-2014-40), April 2014.
[23] The list of control variables included weather factors (e.g., heavy and freezing rain), profitability variables (e.g., load factor), market competition, labor actions, airport expansions and congestion levels, route and carrier characteristics (e.g., connections at origin and destination, departure schedule), and aircraft characteristics (e.g., capacity levels). All regression models also include dummy variables for each month in the dataset and for the number of observations per route-carrier combination.

and attributes (e.g., cancellation rates, weather statistics) of a donor pool of groups to mimic the same pre-policy behavior and attributes of the target group. Our target group was the top 10 percent of route-carrier combinations that had the highest percentage of excessive delays before TDR implementation.[24] The donor pool consisted of nine groups of the remaining 90 percent of route-carrier combinations. We used pre-TDR cancellation rate data and the list of control variables common to all four econometric models to form the synthetic control group. Under this method, calculating the difference between actual and simulated cancellation rates for the post-TDR periods quantifies the effects of the TDR.

Our analysis used airline data on 2,511 domestic directional airport-pairs[25] from the fourth quarter of 2005 through the fourth quarter of 2014. This dataset was initially formed by combining on-time performance data from the Airline Service Quality Performance (ASQP) database, the airline tickets' information from the Airline Origin and Destination survey (DB1B), and the route-carrier group statistics from the Air Carrier Statistics Database (T-100), all of which the Bureau of Transportation Statistics (BTS) maintains.[26] We complemented these data with weather information from the National Center for Environmental Information and airport capacity utilization data from FAA's Aviation System Performance Metrics (ASPM).[27] Our dataset contained 20 airlines, 15 of which are major carriers.[28] The remaining airlines are national carriers with annual revenues between $100 million and $1 billion.[29]

To ensure sufficient time series variation, we omitted route-carrier groups with data spanning less than one year. For data reliability reasons, we excluded any route-carrier group reporting less than or equal to eight flights per month, and flights that arrived more than 1 hour early or over 6 hours late. We also excluded observations with either an origin or destination outside the continental U.S. Moreover, we limited our analysis to routes connecting to the 70 larger airports for

[24] In our study, each observation represents one route (i.e., a directional airport pair) operated by one carrier. We did all our analysis at this level.

[25] For example, Washington Dulles (IAD) to Raleigh-Durham International (RDU) is considered to be a different route than RDU to IAD.

[26] DB1B is a quarterly 10 percent sample of all airline tickets issued by reporting carriers. All airlines operating any aircraft having over 60 seats are required to report. We used the T-100 database to identify service type, aircraft characteristics and load factors, defined as the monthly average ratio of total passengers and seats.

[27] The capacity utilization data were based on FAA's "called rates," which are the maximum hourly number of arriving and departing flights an airport can safely handle as determined by air traffic control given existing operational conditions, such as wind direction and weather. See exhibit B for further information.

[28] The BTS defines a major carrier as a carrier having annual revenues in excess of $1 billion The major carriers are: ATA (TZ), AirTran (FL), Alaska (AS), American (AA), American Eagle (MQ), Comair (OH), Continental (CO), Delta (DL), Frontier (F9), JetBlue (B6), Northwest (NW), SkyWest (OO), Southwest (WN), US Airways (US), and United (UA).

[29] The national carriers are: Atlantic Southeast (EV), ExpressJet (XE), Mesa (YV), Pinnacle (9E), and Virgin America (VX).

which we obtained FAA airport capacity utilization data.[30] Even though the number of routes covered in our final sample (333,447 monthly observations) only account for about 43 percent of those in our sample prior to the merger with the capacity utilization data, these routes encompass approximately 60.4 percent of the total number of flights in the larger sample.

Our cancellation impact work was reviewed by academics with peer-reviewed publications related to this study: Jan Brueckner (University of California – Irvine), Hideki Fukui (Eihme University), and Koki Nagata (University of Maryland). We addressed their questions and comments.

Our analysis of the impact of the TDR on flight delays was purely descriptive. Specifically, our analysis consisted primarily of calculating correlations and percentage changes, and constructing graphs, while breaking out the data along several different dimensions, such as airport size.

[30] FAA's Aviation Performance Metrics Airport Efficiency System is not publically accessible. It covers 77 airports, but only 73 within the lower 48 states.

EXHIBIT B. ANALYSIS OF CANCELLATIONS IN TECHNICAL DETAIL

This exhibit describes our scope and methodology for our analysis of cancellations in technical detail. It is organized as follows. First, we outline our approach. Second, we describe the sources, construction, and characteristics of our estimation dataset. Third, we detail our model specifications and estimation approaches. The last section discusses our results and sensitivity analyses.

Our Approach

We developed a measure of the historical incidence of lengthy tarmac delays on each route for each carrier. We refer to this measure as our "exposure" variable, under the assumption that increased likelihood of lengthy tarmac delays prior to implementation of the TDR increases a flight's exposure to or probability of being affected by the rule.

The exposure variable provided us with a means of sorting route-carrier combinations into what are effectively control groups. Our use of a continuous exposure variable to distinguish control groups draws upon the approaches documented in Angrist and Pischke (2009), who provide examples of differentiating between control groups using a measure of the fraction affected by a policy change. For example, Kiel and McClain (1995) used distance to an undesired facility as an exposure measure to assess the impact of installing a waste-to-energy incinerator on house prices. Mian and Sufi (2012) followed a similar approach in using ratios of "clunkers" to automobile stock to measure a city's exposure to USDOT's 2009 "Cash for Clunkers" program in order to quantify the impact of that program on auto purchases. Like these authors, we employed our exposure variable in Differences in Differences (DID) models, which we used to estimate the TDR's effects on cancellation rates.

We constructed the exposure variable as the percentage of flights subject to tarmac delays greater than a given threshold length in the period prior to the TDR's implementation. We constructed this variable for each route-carrier combination, i.e., for each combination of a directionally-specific route and carrier servicing that route. The period over which the percentages are calculated runs from the beginning of our dataset through March 2010, the month prior to the TDR's initial implementation. We also generated a version of the exposure variable based on 2007 data alone.[31]

[31] Cancellation behavior by airlines during 2007 is not plausibly affected by a policy implemented in 2010, and 2007 is the pre-tarmac year with the worst record of lengthy flight delays in our dataset. Hence, using 2007 for exposure calculations provides a suitable worst-case scenario for a robustness assessment of the results obtained using the whole pre-rule period as a baseline.

We used a range of thresholds—60, 90, and 120 minutes—to develop exposure measure variants. GAO (2011) found that important points in the decision process for cancelling flights occurred at 60 and 120 minutes into a tarmac delay. Conversations between an airline and airport control tower about how they will handle the delay and a potential return to the gate begin at the 60 minute mark of a tarmac delay. Cancelling a flight on the tarmac before the 3-hour limit is breached requires that the pilot begin implementing procedures to return to the gate once the tarmac delay has reached approximately 2 hours.

The exposure variable will only identify an effect well if: (a) there is variation across route-carrier combinations in the number of instances planes exceed a time threshold on the tarmac, and (b) the exposure variable produces a fairly consistent ranking of route-carrier combinations across the years preceding the TDR. If the first condition is not met, and all route-carriers evidence a uniform distribution of flights exceeding tarmac delay thresholds, then the exposure variable will not provide additional information. In our dataset, we found that the standard deviation of every exposure variable variant is much larger than its mean. If the second condition is not met, the exposure variable would not be able to distinguish meaningful control groups. A rank correlation test found substantial and statistically significant correlation in annual route-carrier rankings by exposure variable value from 2006 through 2009.

We used April 2010 as the date of the TDR's implementation, even though at that time the rule applied solely to flights from airports classified as large or medium hubs by the Federal Aviation Administration (FAA).[32] This provided us with a clear division between pre- and post-policy periods. Route-carrier combinations originating at large or medium hub airports comprise about 78 percent of our regression dataset. Because DOT extended the rule's provisions to all domestic flights as of August 2011, only 3.24 percent of the route-carrier combinations in our dataset were consequently included among those covered by the TDR before their coverage under the rule began.[33]

Data and Statistics

Dataset

To form our baseline dataset, we merged the Bureau of Transportation Statistics (BTS) Airline Service Quality Performance Data (ASQP), Airline Origin

---

[32] FAA classifies airports according to passenger enplanements. All "primary" airports enplane at least 10,000 passengers annually. FAA designates primary airports as large, medium, or small hubs, or as non-hubs. Of system-wide passengers, "large" hubs enplane at least 1 percent; "medium" hubs at least 0.25 percent, but less than 1 percent; "small" hubs at least 0.05 percent, but less than 0.25 percent; and "non-hub" less than 0.05 percent.

[33] Our findings are unchanged when we re-estimate our regressions using the correct TDR implementation date for route-carrier combinations originating from small hub and non-hub airports. These results are available upon request.

and Destination Survey (DB1B), and Air Carrier Statistics Database (T100).[34] The ASQP database contains flight-level performance data for direct flights by air carriers who account for 1 percent or more of total domestic scheduled service passenger revenues, which we use to construct our exposure measures. The DB1B is a quarterly 10 percent sample of scheduled airline tickets from carriers operating any aircraft larger than 60 seats. The T100 details air service type and flight and aircraft characteristics.

Our unit of observation is route-carrier-month. We define a route as a directional origin and destination airport-pair. For example, Washington Dulles (IAD) to Raleigh-Durham International (RDU) is considered to be a different route than RDU to IAD. Airport-pair, instead of city-pair, is used to define a route because different airports located in the same city can experience different cancellation rates.

We converted all data into monthly averages except those based on the DB1B quarterly data, which we assume are constant across the months within a quarter. Our dataset covers the 4th quarter of 2005 through the 4th quarter of 2014 and 20 airlines. Fifteen of these are major carriers, defined by BTS as having annual revenues in excess of $1 billion.[35] The remaining airlines are national carriers with annual revenues between $100 million and $1 billion.[36]

We limited our analysis to routes connecting the 70 larger airports for which we obtained FAA airport capacity utilization data.[37] Those hourly data report the number of operations (e.g., arrivals and departures) an airport handles divided by the maximum number air traffic control determines it can handle safely given runway configuration and weather conditions. Our final dataset contains 2,511 directional airport-pairs or routes, 4,935 route-carrier combinations, and 333,447 monthly observations.[38] Even though the number of routes covered in our

---

[34] The ASQP data is linked to the operating carrier, while each observation in the DB1B is tied to a combination of an operating carrier as well as a marketing carrier. It is possible that in the same month and on the same route, one operating carrier is working for multiple marketing carriers and vice versa; e.g., code-sharing. About 50 percent of the observations have an operating carrier that is different from the marketing carrier. We merged these two datasets by operating carrier, and assigned a unique marketing carrier based on the following rules: (1) For records with multiple operating carriers all linked to the same marketing carrier, we assigned that marketing carrier. (2) For records with multiple marketing carriers linked to the same operating carrier, we assigned the dominant marketing carrier, the one with greater than or equal to 50 percent of the origin and destination passenger share.

[35] The major carriers are: ATA (TZ), AirTran (FL), Alaska (AS), American (AA), American Eagle (MQ), Comair (OH), Continental (CO), Delta (DL), Frontier (F9), JetBlue (B6), Northwest (NW), SkyWest (OO), Southwest (WN), US Airways (US), and United (UA).

[36] The national carriers are: Atlantic Southeast (EV), ExpressJet (XE), Mesa (YV), Pinnacle (9E), and Virgin America (VX).

[37] We obtained airport capacity utilization data from FAA's Aviation Performance Metrics Airport Efficiency System, which is not publically accessible. It covers 77 airports, but only 73 within the lower 48 states.

[38] To ensure sufficient time series variation for estimating each panel, we omit route-carrier groups with data spanning less than one year. For data reliability reasons we exclude any route-carrier combination reporting less than or equal to eight flights per month, and flights that arrived more than 1 hour early or over 6 hours late. We also exclude observations with either an origin or destination outside the continental U.S.

final dataset only account for about 43 percent of those in our dataset prior to the merger with the capacity utilization data, these routes encompass approximately 60.4 percent of the total number of flights in the larger sample.

Control Variables

Our choice of control variables for both the regression and synthetic control group methods builds on our previous statistical work on airline cancellation rates (DOT Office of Inspector General 2014). Most of our control variables fall into three groups: weather factors, profitability measures, and route-carrier specific covariates. Tables B1–B3 present summary statistics of our data.

Our weather data came from the National Center for Environmental Information. We used heavy rain and freezing rain to capture the weather conditions that would most likely cause cancellations. $Heavy\ Rain\ Origin_{jt}$ is the percentage of days in month *t* at the origin airport of route *j* with precipitation in excess of the 95th percentile for that airport over the entire dataset period. $Freezing\ Rain\ Origin_{jt}$ is the percentage of days in month *t* with greater than 95th percentile precipitation and a temperature below 32 degrees at the origin airport of route *j*. Similarly, $Heavy\ Rain\ Dest_{jt}$ and $Freezing\ Rain\ Dest_{jt}$ capture extreme weather at the destination airport of route *j*.

We used load factors to indicate airline profitability by route. A flight's load factor is the total number of passengers divided by the total number of seats. $Load\ Factor_{ijt}$ is airline *i*'s average load factor on route *j* in month *t*.

Our route-specific covariates included measures of competition and controls for airport expansions. Our route-carrier-specific covariates included measures of airport congestion, airport hubbing, aircraft characteristics and labor actions. We used a Hirschman-Herfindahl index (HHI) based on airlines' revenue market share, calculated using DB1B data, to gauge the level of competition in each market or route. Specifically, HHI_{jt} is the index value for route *j* in month *t*. To determine the level of competition in each market, we included both direct flights and those requiring one connection when calculating HHI_{jt}.[39]

$Depart\ Early_{ijt}$, $Depart\ 6\ am\ to\ 12\ noon_{ijt}$, $Depart\ 12\ noon\ to\ 6\ pm_{ijt}$, and $Depart\ 6\ pm\ to\ 12\ midnight_{ijt}$ represent the monthly percentage of route-carrier specific total departures occurring from midnight to 6 am, 6 am to 12 pm, 12 pm to 6 pm, and 6 pm to midnight, respectively. $Depart\ Early_{ijt}$ is our reference group.

[39] Only 2 percent of the passengers in our sample take trips involving two or more connections, and we do not include these in our analysis. Note that the cancellation rate for each route-carrier is calculated using only data for direct flights on that route for that carrier.

We controlled for the impact on cancellations of airport congestion by controlling for airport capacity utilization adapted to the route-carrier level. Specifically, we multiplied the monthly average of each hourly airport capacity utilization rate by the monthly average number of flights for each route-carrier using the airport in that hour. The sum of these products across hours at the origination airport is captured in $Congestion\ Origin_{ijt}$. Similarly, $Congestion\ Dest_{ijt}$ captures the route-carrier specific capacity utilization level at the destination airport.[40] These measures allow two airlines serving the same route to experience different congestion levels at the same airport if one schedules more flights during the morning rush while the other predominantly schedules flights in the afternoon.

We controlled for the effects of airline hubbing with variables constructed using DB1B data. As airline hubbing involves an airline scheduling multiple arrivals and departures to facilitate passenger connections, we included two variables controlling for origin and destination connections. $Connections\ Origin_{ijt}$ tracks the total number of locations to which there are direct flights flown by airline *i* from the origin airport of route *j* in the quarter containing month *t*. $Connections\ Dest_{ijt}$ is the corresponding measure for the destination airport.

We also included information on aircraft characteristics. We assembled information on aircraft model for each route-carrier combination from T100 data, and use aircraft manufacturer Web sites to identify each model's seating capacity. We grouped the aircraft in our dataset into five categories based on number of seats. $Aircraft\ Type\ Group\ 1_{ijt}$ is the monthly percentage of all aircraft used by carrier *i* on route *j* comprised of turbo-prop aircraft and $Aircraft\ Type\ Group\ 2_{ijt}$ is comprised of regional jets having up to 70 seats. $Aircraft\ Type\ Group\ 3_{ijt}$ through $Aircraft\ Type\ Group\ 5_{ijt}$ were generated in a similar fashion with seating increasing for higher group numbers.[41]

Other controls include dummy variables for airport expansions and labor actions. Specifically, one set of dummies indicates the completion of airport runway projects in November 2008 at three major airports: Seattle-Tacoma (SEA), Chicago O'Hare (ORD), and Washington Dulles (IAD). For example, $SEA\ Runway\ Origin_{jt}$ is set to one starting November 2008 if the origin airport on route *j* is the Seattle-Tacoma International Airport, and zero otherwise. The other set of dummies, $Labor\ Strikes_{it}$, indicate periods of labor strikes and slowdowns.[42]

[40] These variables also capture slot control effects. Capacity utilization averaged 62 percent for the slot-controlled airports in our dataset, and 49 percent for all other airports.
[41] Group 3 contains regional jets with 70 to 100 seats. Group 4 is made up of narrow body planes having more than 100 seats. Group 5 includes all wide body aircraft.
[42] We collected information on airline labor actions from: the Wall Street Journal, the Chicago Tribune, the Bureau of Labor Statistics, CNN, USA Today, NBC news, TribLive, and Highbeam. Four events were identified during our

Empirical Models

Regression Models

Similar to other regression analyses applying the DID approach with a continuous variable to distinguish control groups, we used the interaction between the exposure variable and a dummy variable for the post-policy implementation period to measure the impact of the TDR on flight cancellation rates. The structure of our regression model is

$$Cancel_{ijt} = \theta Expo_{ij} * TDR_t + \alpha X_{ijt} + \varepsilon_t + c_{ij} + \mu_{ijt} \ (1)$$

where *i* indexes airline, *j* indexes route, and *t* indexes time. The dependent variable $Cancel_{ijt}$ is the monthly percentage of airline *i*'s direct flights on route *j* that are cancelled. $Expo_{ij}$ represents the exposure measure, which is time-invariant and specific to each route-carrier. TDR_t is a dummy variable that equals one for post-implementation observations and zero otherwise.

The vector X_{ijt} contains explanatory variables that may vary by route-carrier and time. ε_t is a vector of time dummies, which in part serves to absorb shifts in macroeconomic variables, such as the unemployment rate. c_{ij} captures time invariant factors for each route-carrier combination, such as distance. [43] μ_{ijt} represents the idiosyncratic error. The composite error is $v_{ijt} = c_{ij} + \mu_{ijt}$, where the sequence $\{v_{ijt}: t = 1, \ldots, T\}$ is allowed to be serially correlated. The composite errors are assumed to cluster over time within each route-carrier combination but be independent across different combinations.

We allowed for time varying effects on cancellations, including prior to announcement of the TDR, by multiplying pre- and post-rule time periods by the exposure variable. The resulting dynamic version of our model is

$$Cancel_{ijt} = \theta Expo_{ij} \cdot D_t + \alpha X_{ijt} + \varepsilon_t + c_{ij} + \mu_{ijt} \ (2)$$

where D_t is a vector with dummy variables for the following pre- and post-policy periods: (i) August 2009 to December 2009, (ii) January 2010 to April 2010, (iii) May 2010 to April 2011, (iv) May 2011 to April 2012, (v) May 2012 to April 2013, (vi) May 2013 to April 2014, and (vii) May 2014 to December 2014. The first interval captures the period just prior to DOT's announcement of the TDR on

dataset, affecting: (1) Northwest Airlines from October to November 2005; (2) Northwest Airlines in November 2006; (3) US Airways in April 2006; and (4) American Airlines from September to October 2012.

[43] We have omitted a stand-alone exposure variable because c_{ij} captures all time-invariant route-carriers effects. Since our exposure variable is constructed to be both time-invariant and specific to each route-carrier pair, c_{ij} absorbs all the variation that would be associated with it.

December 21, 2009.[44] The second interval controls for the period from rule announcement to its implementation on April 29, 2010. The remaining intervals consist of 1-year periods following implementation, except for the last interval which covers the final 8 months of our dataset.

We examined whether the TDR impact varies seasonally by extending model (1). Designating May through September as summer months, as in GAO (2011), we set the dummy variable $Summer_t$ equal to one for the summer months and zero otherwise. Analogously, we coded the dummy variable $Winter_t$ as equal to one from November through February, and zero otherwise. Our model extended for studying seasonal effects is

$$Cancel_{ijt} = \theta_W Expo_{ij} * TDR_t * Winter_t + \theta_S Expo_{ij} * TDR_t * Summer_t$$

$$+\alpha X_{ijt} + \varepsilon_t + c_{ij} + \mu_{ijt} \text{ (3)}$$

We note that model (3) measures effects relative to the months March, April, and October. For example, if θ_S were positive and significant, then the TDR would have caused an increase in summer cancellation rates *compared to* cancellation rates in March, April, and October.

We studied the possibility that some carrier types (e.g., low-cost carriers) respond more aggressively to the possibility of violating the TDR by considering an additional extension of model (1). Letting $Lowcost_i$ denote a dummy variable that equals one if airline *i* is a low-cost carrier and zero otherwise, our extension of model (1) to examine cancellation behavior by airline type is

$$Cancel_{ijt} = \theta_{ELC} Expo_{ij} * Lowcost_i + \theta_{ELCT} Expo_{ij} * Lowcost_i * TDR_t$$

$$+\theta_{LC} Lowcost_i + \alpha X_{ijt} + \varepsilon_t + c_{ij} + \mu_{ijt} \text{ (4)}$$

where θ_{LC} controls for average cancellations of low-cost carriers relative to other carriers (e.g., legacy carriers) and θ_{ELC} measures the differentiated relevance of exposure for low-cost carriers relative to other carrier types. With these controls in place, θ_{ELCT} is interpreted as the relative impact of the TDR on low-cost airline cancellation behavior, relative to all other carriers.

The dependent variables in models (1) through (4) are bounded by zero and one. Thus, estimating these models using simple linear regression is not appropriate as they may yield predictions lying outside the [0,1] interval. Moreover, estimation using simple linear regression is not appropriate if the relationship between at least

[44] This interval intends to capture cancellation behavior reactions to an extreme tarmac delay incident highly publicized in the media, which occurred on August 8, 2009. The incident involved Continental Flight 2816 from Houston, TX, to Minneapolis, MN. After diversion to Rochester, MN, the airplane sat grounded on the tarmac for over 6 hours.

one covariate and the dependent variable is nonlinear.[45] In addition, several route-carrier combinations have no cancellations for many months in our dataset.

As a result, many observations of the dependent variable assume an extreme value (zero) while the remaining observations exhibit values in the middle of the [0,1] interval. The fractional probit model of Papke and Wooldridge (1996, 2008) successfully deals with these issues not addressed by common limited dependent variable methods (e.g., logit, probit) while allowing for nonlinearities in the relationship of cancellation rates with covariates. Therefore, models of the percentage of cancelled flights are appropriately estimated using the fractional response method. We also estimated (1) using linear regression as a comparison allowing us to assess the relevance of controlling for bounded dependent variables and allowing for nonlinearities in the relationship with covariates.

Synthetic Control Group Method

We used a variant of the DID approach—the synthetic control group method—to examine TDR effects on route-carrier combinations with a high probability of experiencing lengthy tarmac delays. If the TDR's impact was concentrated in increased cancellations of route-carrier combinations previously prone to these longer delays, this effect could be overlooked when quantifying policy effects for the entire population of route-carrier combinations, as we do with the regression approach.

The synthetic control group method allowed us to address this concern directly. To implement it, we combined groups of route-carrier combinations less likely to be affected by the TDR into a single, synthetic group having approximately the same statistics (e.g., cancellation rates and average weather conditions) as route-carrier combinations likely affected by the TDR in the pre-implementation period. We define the group of route-carrier combinations likely affected by the TDR as those in the top decile of the exposure measure for a particular delay threshold. The synthetic control group is created using route-carrier combinations in the remaining nine deciles. We measured the TDR effect by computing the average difference between cancellation rates in the post-rule period of route-carrier combinations in the top decile of the exposure measure and the simulated average cancellation rates of the synthetic control group. Our implementation of this method follows Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010).

[45] In previous work we found nonlinear relationships between cancellation rates and at least one of the covariates mentioned above. See *Reductions in Competition Increase Airline Delays and Cancellations* (OIG Report No. CR-2014-40), April 2014.

Covariates

The covariates used in our models are as follows. The explanatory variable vector X_{ijt} in all our regression models consists of the control variables previously discussed. Pre-TDR means of these variables are also used in the application of the synthetic control group method. We used somewhat different sets of alternative exposure measure specifications with the two different methods as described below. In addition, our regression models include three other covariate sets: (i) unbalanced panel adjustment terms, (ii) a control function, and (iii) correlated random effects terms. We provide further discussion of these measures below.

For the sake of robustness, we checked whether the use of alternative exposure measure specifications impacts our results. Specifically, we compared results generated using 60 versus 90 minutes as the delay threshold in constructing the exposure variable, as well as those generated using the whole pre-TDR period of our dataset versus the year 2007. We also tried an exposure measure based on a delay threshold of 120 minutes in our regression models, but found it led to numerical instability. A similar problem was experienced with a 90-minute threshold when basing the exposure variable on data from the whole pre-TDR period.[46] Consequently, we considered the 90- and 120-minute threshold exposure measures only when using the synthetic control group method.

Some route-carrier pairs of our dataset were not observed in all periods, resulting in an unbalanced panel dataset. In order to control for this source of heteroscedasticity, we followed Wooldridge (2010b) by counting the number of time periods in the data for each route-carrier pair. We formed dummy variables for each count and added them as regressors. We also allowed the log of the variance to be a linear function of these variables.

The endogeneity of HHI in empirical models of airline ticket prices is well documented, raising the concern that HHI is also endogenous in empirical models of airline flight cancellations. We addressed this concern using lagged HHI to instrument for HHI in the current period. We carried this out using the control function (CF) approach in both the fractional response and linear regressions because it offers advantages over alternative instrumental variable methods in nonlinear models.[47]

The propensity for flight cancellation may vary due to route-carrier specific factors not directly observed in the data, but correlated with the error term. Using a modified Hausman-Taylor test, we found that this is indeed a concern in our

---

[46] We believe these instabilities stem from the large shares of zero value observations. Only 1 percent of pre-TDR observations had tarmac delays greater than or equal to 120 minutes. The comparable figure for 90-minute delays was 3 percent.
[47] These advantages are discussed in Wooldridge (2010a), pp. 126 and 589.

estimations. Because our regressions are nonlinear, we modeled route-carrier heterogeneity using correlated random effects (CRE). However, CRE requires specification of the form of the correlation (Wooldridge, 2010a). We modeled the term c_{ij} as a linear function of all exogenous variables' time averages and a Gaussian error term.

Results

Regression Model Results

DID Exposure Measure Results

Table B4 details our results estimated using equation (1) for two fractional probit models and one linear specification. The fractional probit specifications include an exposure measure calculated using a 60-minute tarmac delay threshold. One variant of the exposure measure is based on 2007 data alone, while the other is based on the entire pre-TDR dataset. The linear specification uses the latter exposure measure.

The estimated coefficient on the interaction between the exposure variable and the post-policy period dummy is positive in all three estimations, but in only two is it significant at even the 10 percent level. Notably, these results apply to the whole population of domestic flights averaged over the entire post-rule period. They do not differentiate impacts by time since the TDR's implementation or by propensity for more extreme tarmac delays.

Table B5 presents our results derived estimating equation (2), our dynamic regression model, using fractional probit and including two different exposure variable specifications. Both exposure variables are again calculated using a 60-minute threshold; one based on 2007 data alone, and the other based on the entire pre-TDR dataset. In each estimation, the exposure measure is interacted with dummy variables for the pre- and post-policy periods.

Except for those capturing effects in the first 3 years after the TDR implementation, all interactions of the exposure variable with time interval dummies have negative and statistically significant coefficients. Notably, the interaction terms pertaining to the 3 years immediately after the policy became effective are positive and significant at the 1 percent level. Therefore, compared with the period before August 2009, route-carrier combinations with greater frequency of tarmac delays lasting at least 60 minutes experienced relatively lower cancellation rates until the rule came into effect. Then, during the first 3 years after the rule was implemented, those same route-carrier combinations saw statistically significant cancellation increases. Afterwards, flight cancellations returned to their previous behavior.

A rationale for the negative coefficients is provided by the marked trend in the kinds of flights supplied over time. Figure B1 shows that a change in the composition of flights was ongoing throughout the period covered by our dataset. Specifically, flights with high frequency of lengthy pre-rule tarmac delays decreased in number while those with low frequency of lengthy pre-rule tarmac delays exhibited a relative increase. Since the frequency of lengthy pre-rule tarmac delays correlates highly with cancellation rates both pre- and post-TDR, this change in the composition of flights would be expected to reduce average cancellation rates over time. The cancellation rate averaged over the entire pre-August 2009 period would then be expected to be higher than the cancellation rate in subsequent periods.

The effects estimated using the fractional response estimates of model (2) are consistent with the results for model (1), where the positive but insignificant coefficient presumably arises from mixing the immediate post-policy positive effects with the negative effects in other periods.

Model Specification Results

In table B4, the two fractional response models display coefficient estimates similar in sign and statistical significance, but different in magnitude. So the choice of exposure variable affects the size of the estimated impacts. We consider using data from the entire pre-TDR period to calculate the exposure variable preferable to only using 2007 data. A priori, it is more representative, as 2007 stands out as the worst year for tarmac delays in the pre-policy period. We focus our conclusions on the results based on the exposure measure using data from the entire pre-TDR period. In so doing we are choosing the results showing the largest effects.

The fractional response models' estimation results appear to be more reasonable than the linear regression's. While most control variable coefficients were of the expected sign in all three specifications in table B4 (e.g., extreme weather events increased cancellations), there are discrepancies between the fractional response models and the linear regression in regards to statistical significance. For example, destination airport congestion is statistically significant at 1 percent in the fractional response models, yet it is statistically insignificant at 10 percent in the linear regression model. In addition, coefficient estimates for some controls differ in sign between the linear and fractional response models (e.g., morning flight indicator, origin airport congestion), taking counterintuitive signs in the linear model. Consequently, we focus on the fractional probit results.

The results in table B4 also confirm our prior expectation of the HHI being endogenous. Specifically, the CF term is significant at the 1 percent level. As noted earlier, the CF approach corrects for this problem.

Control Variable Results

As mentioned above, most control variable coefficients had the expected effect on cancellations, as shown in table B4. For example, labor strife substantially increased cancellations, while the completion of new runway projects reduced them. Departure time variables' coefficients indicate that the probability of a flight being cancelled is greater the later in the day it is scheduled to depart. In addition, the negative coefficients on load factor and the larger airplane groups indicate that flights with more passengers are less likely to be cancelled.

Several coefficients are not statistically significant, such as the coefficients on $Connections\ Origin_{ijt}$, and $Congestion\ Origin_{ijt}$. However, the coefficient on $Congestion\ Dest_{ijt}$ is significant and positive, indicating that airport congestion at the destination is more relevant than at the origin.

Low-Cost and Seasonal Model Analysis

Our results show that the TDR had a greater impact on cancellations during the summer months, relative to other seasons, and on low-cost carriers, relative to other carrier types. Tables B6 and B7 show results estimated using models (3) and (4). All exposure variables were constructed using a 60-minute delay threshold. As a robustness check, we again used both the whole pre-TDR dataset and the year 2007 for exposure calculation.

For model (3), we found that the coefficient for the interaction between the exposure variable and the product of the low-cost airline and post-rule period dummies is positive and statistically significant at the 1 percent level. Therefore, compared with other carrier types (e.g., legacy carriers), the TDR led low-cost carriers to have a relative increase in cancellation rates.

For model (4), we found that the coefficient on the product between the exposure measure, summer, and post-rule dummies is positive and statistically significant at the 1 percent level. Thus, the TDR effect on summer cancellation rates was greater than its effect on airline behavior in March, April, and October. We also observed that the coefficient on the product between the exposure measure, winter, and post-rule dummies is negative and statistically significant at the 1 percent level. That is, for the post-rule period, winter months had a negative variation in average cancellation rate compared to March, April, and October. The marginal effect results for both models are presented in tables B6 and B7. Since these results benchmark against other carriers or months, the absolute magnitude of these impacts cannot be identified without imposing additional assumptions.

Synthetic Control Group Results

Using the synthetic control group approach, we confirmed that airlines at higher risk of extended flight delays prior to the TDR implementation were more likely to experience cancellations in the post-implementation period. We considered three exposure measures based on the tarmac delay information of each route-carrier pair for the whole pre-policy period. The exposure measures vary by the threshold above which a flight is tagged as being excessively delayed on the tarmac. Specifically, we used thresholds of 60, 90, and 120 minutes. Use of the 60-minute threshold exposure variable facilitates comparisons with our regression results. The remaining thresholds progressively approach the 3-hour limit delineated in the TDR and serve the purpose of checking whether the policy's effects vary in magnitude and significance as the extremity of the delays increases.

In common with other applications of this method (e.g., Abadie et al 2010, Abadie and Gardeazabal 2003), the optimal weight for most groups is zero and weight is concentrated on a few members of the donor pool. In all cases, the 9th decile was the group receiving the most weight. Intuitively, route-carrier pairs with exposure levels in this decile are closer in their characteristics to the 10th decile route-carrier combinations.

Figures B2, B3, and B4 exhibit average cancellation rates over time of route-carrier combinations in the top 10 percent (i.e., 10th decile) of each exposure measure along with those for the synthetic control group. All three figures show that the synthetic group replicates reasonably well the actual cancellation rates prior to the TDR implementation. However, the synthetic and actual series diverge subsequently, particularly when the actual cancellation rates peak.

The results produced by this method are consistent with those from the regression analysis. First, the actual series only exceeds the synthetic series in the first 3 years post-rule in all three graphs. Second, table B8 shows that, for the all the exposure variables, the average difference between the actual and synthetic cancellation rate is positive and statistically significant in the first 3 years post-rule

Both the magnitude and statistical significance of the average difference between the two series increase for higher excessive delay thresholds in the first 3 years post-rule. For the exposure measure calculated using a 60-minute threshold the estimated average difference is 0.39 percent and significant at the 5 percent level. This estimate implies that, in the first 3 years post-rule, these routes experienced approximately five extra cancellations per day in addition to cancellations that would happen in the absence of the rule (out of the top 10 percent route-carrier pairs prone to delays above 90 minutes, which accounted for 1,393,449 flights for that time interval).[48] For the exposure measure calculated using a 90-minute

[48] The differences in flights and implied additional cancellations are due to both flight frequency per route-carrier and decile composition for each exposure measure. For example, consider two route-carrier pairs, A and B, with 50 and 100 scheduled flights per month, respectively. A has 5 percent delays beyond 2 hours on record while B has 10 percent delays beyond 90 minutes (but all below the 2-hour mark). Then B has a higher percentage of delays beyond 90 minutes than A, yet A has a higher percentage of delays beyond 120 minutes than B (which has zero percent by

threshold the estimated average difference is higher, 0.49 percent, and more significant (now at the 1 percent level). From the 3 years following May 2010, these routes experienced approximately six extra cancellations per day in addition to cancellations that would happen in the absence of the rule (out of the top 10 percent route-carrier pairs prone to delays above 90 minutes, which accounted for 1,360,282 flights for that time interval). The magnitude of average difference is even larger, 0.56 percent, and significant at the 1 percent level for the 120-minute threshold. This estimate implies that, in the first 3 years following the TDR's implementation, approximately seven additional flights were cancelled per day (out of the 10th decile of flights more prone to delays in excess of 2 hours, which consisted of 1,387,644 flights).

References

Abadie, A. and J. Gardeazabal (2003) The Economic Costs of Conflict: A Case Study of the Basque Country, *American Economic Review*, 93 (1), 113-132.

Abadie, A., A. Diamond and J. Hainmueller (2010) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, *Journal of the American Statistical Association*, 105, No. 490, 493-505.

Angrist, J. D. and J.S. Pischke (2008) *Mostly Harmless Econometrics: an Empiricist's Companion*. First Edition, Princeton University Press.

Fukui, H. and K. Nagata (2014) Flight Cancellation as a Reaction to the Tarmac Delay Rule: An Unintended Consequence of Enhanced Passenger Protection, *Economics of Transportation*, 3, 29-44.

Kiel, K. A. and K. T. McClain (1995) House Prices during Siting Decision Stages: The Case of an Incinerator from Rumor through Operation, *Journal of Environmental Economics and Management*, 28, 241-255.

Mian, A. and Sufi, A., (2012) The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program, *The Quarterly Journal of Economics*, 1107-1142.

Papke, L. E. and J. M. Wooldridge (1996) Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates, *Journal of Applied Econometrics*, 11, 619-632.

---

construction). If both A and B lie in the 10th decile of exposure with 90 minute delay threshold, then they count for 150 flights per month. Since B must be in the first decile of exposure when a 120 minutes threshold is used, then only the 50 flights of A are counted.

Exhibit B. Analysis of Cancellations in Technical Detail

Papke, L. E. and J. M. Wooldridge (2008) Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates, *Journal of Econometrics*, 145, 121-133.

U.S. Government Accountability Office (2011) *Airline Passenger Protections: More Data and Analysis Needed to Understand Effects of Flight Delays. Report to Congressional Requesters*, Report Number GAO-11-733.

U.S. Department of Transportation Office of Inspector General (2014). *Reductions in Competition Increase Airline Delays and Cancellations*. Report Number CR-2014-40.

Wooldridge, J. M. (2010a) *Econometric Analysis of Cross Section and Panel Data*, Second edition, MIT Press.

Wooldridge, J. M. (2010b).*Correlated Random Effects Models with Unbalanced Panels*. Manuscript, Michigan State University.

Exhibit B. Analysis of Cancellations in Technical Detail

Table B1. Summary Statistics: Entire Analysis Period

Variables	Mean	Std. Dev	Min	Max
Percentage of Delays >=60 min in 2007	0.013	0.030	0	0.466
Percentage of Delays >=90 min in 2007	0.004	0.009	0	0.130
Percentage of Delays >=120 min in 2007	0.001	0.004	0	0.057
Percentage of Delays >=60 min in all pre-TDR periods	0.011	0.021	0	0.307
Percentage of Delays >=90 min in all pre-TDR periods	0.003	0.006	0	0.091
Percentage of Delays >=120 min in all pre-TDR periods	0.001	0.002	0	0.036
Percent of Cancelled Flights	0.014	0.027	0	0.667
Revenue Based HHI	0.547	0.209	0.144	1
Heavy Rain Origin	0.043	0.048	0	0.355
Heavy Rain Dest	0.024	0.053	0	0.483
Freezing Rain Origin	0.043	0.048	0	0.355
Freezing Rain Dest	0.024	0.053	0	0.483
Load Factor	0.795	0.112	0.052	1
Labor Strikes	0.004	0.065	0	1
Departure 6 am to 12 noon	0.403	0.256	0	1
Departure 12 noon to 6 pm	0.376	0.262	0	1
Departure 6 pm to 12 midnight	0.199	0.226	0	1
Connections Origin	104.718	34.584	1	188
Connections Dest	104.455	34.635	1	187
Congestion Origin	0.482	0.201	0.005	1.058
Congestion Dest	0.477	0.207	0.004	1.079
SEA Runway Origin	0.009	0.094	0	1
ORD Runway Origin	0.020	0.139	0	1
IAD Runway Origin	0.008	0.089	0	1
SEA Runway Dest	0.009	0.093	0	1
ORD Runway Dest	0.020	0.140	0	1
IAD Runway Dest	0.008	0.090	0	1
Aircraft Type Group 1	0.003	0.058	0	1
Aircraft Type Group 2	0.187	0.390	0	1
Aircraft Type Group 3	0.171	0.376	0	1
Aircraft Type Group 4	0.635	0.482	0	1
Aircraft Type Group 5	0.005	0.067	0	1

Exhibit B. Analysis of Cancellations in Technical Detail

Table B2. Summary Statistics: Pre-TDR Period

Variables	Mean	Std. Dev	Min	Max
Percentage of Delays >=60 min in 2007	0.014	0.029	0	0.466
Percentage of Delays >=90 min in 2007	0.001	0.003	0	0.057
Percentage of Delays >=120 min in 2007	0.004	0.009	0	0.130
Percentage of Delays >=60 min in all pre-TDR periods	0.011	0.022	0	0.307
Percentage of Delays >=90 min in all pre-TDR periods	0.003	0.006	0	0.091
Percentage of Delays >=120 min in all pre-TDR periods	0.001	0.002	0	0.036
Percent of Cancelled Flights	0.015	0.027	0	0.563
Revenue Based HHI	0.549	0.208	0.144	1
Heavy Rain Origin	0.042	0.047	0	0.290
Heavy Rain Dest	0.026	0.055	0	0.483
Freezing Rain Origin	0.042	0.047	0	0.290
Freezing Rain Dest	0.026	0.055	0	0.483
Load Factor	0.776	0.118	0.052	1
Labor Strikes	0.005	0.072	0	1
Departure 6 am to 12 noon	0.405	0.260	0	1
Departure 12 noon to 6 pm	0.380	0.267	0	1
Departure 6 pm to 12 midnight	0.197	0.228	0	1
Connections Origin	105.709	33.712	1	185
Connections Dest	105.421	33.717	1	186
Congestion Origin	0.499	0.200	0.008	1.058
Congestion Dest	0.494	0.207	0.004	1.079
SEA Runway Origin	0.006	0.077	0	1
ORD Runway Origin	0.014	0.117	0	1
IAD Runway Origin	0.006	0.076	0	1
SEA Runway Dest	0.006	0.076	0	1
ORD Runway Dest	0.014	0.117	0	1
IAD Runway Dest	0.006	0.076	0	1
Aircraft Type Group 1	0.004	0.062	0	1
Aircraft Type Group 2	0.214	0.410	0	1
Aircraft Type Group 3	0.173	0.378	0	1
Aircraft Type Group 4	0.604	0.489	0	1
Aircraft Type Group 5	0.005	0.074	0	1

Exhibit B. Analysis of Cancellations in Technical Detail

Table B3. Summary Statistics: Post-TDR Period

Variables	Mean	Std. Dev	Min	Max
Percentage of Delays >=60 min in 2007	0.013	0.030	0	0.466
Percentage of Delays >=90 min in 2007	0.001	0.004	0	0.057
Percentage of Delays >=120 min in 2007	0.004	0.009	0	0.130
Percentage of Delays >=60 min in all pre-TDR periods	0.010	0.021	0	0.296
Percentage of Delays >=90 min in all pre-TDR periods	0.003	0.006	0	0.091
Percentage of Delays >=120 min in all pre-TDR periods	0.001	0.002	0	0.034
Percent of Cancelled Flights	0.013	0.026	0	0.667
Revenue Based HHI	0.545	0.210	0.167	1
Heavy Rain Origin	0.044	0.048	0	0.355
Heavy Rain Dest	0.022	0.051	0	0.345
Freezing Rain Origin	0.044	0.048	0	0.355
Freezing Rain Dest	0.022	0.051	0	0.345
Load Factor	0.818	0.100	0.237	1
Labor Strikes	0.003	0.056	0	1
Departure 6 am to 12 noon	0.400	0.250	0	1
Departure 12 noon to 6 pm	0.372	0.256	0	1
Departure 6 pm to 12 midnight	0.203	0.222	0	1
Connections Origin	103.525	35.570	1	188
Connections Dest	103.292	35.675	1	187
Congestion Origin	0.462	0.201	0.005	1.049
Congestion Dest	0.456	0.205	0.017	1.058
SEA Runway Origin	0.012	0.110	0	1
ORD Runway Origin	0.027	0.162	0	1
IAD Runway Origin	0.011	0.103	0	1
SEA Runway Dest	0.012	0.109	0	1
ORD Runway Dest	0.027	0.163	0	1
IAD Runway Dest	0.011	0.104	0	1
Aircraft Type Group 1	0.003	0.053	0	1
Aircraft Type Group 2	0.153	0.360	0	1
Aircraft Type Group 3	0.168	0.374	0	1
Aircraft Type Group 4	0.672	0.470	0	1
Aircraft Type Group 5	0.003	0.058	0	1

Exhibit B. Analysis of Cancellations in Technical Detail

Table B4. Model of Percentage of Cancelled Flights

Model	Fractional Probit				Linear Regression
Exposure variable	Delay (>=60min) rel. freq. in 2007		Delay (>=60min) rel. freq., all pre-TDR periods		
Estimates	Model Coefficients	Marginal Effects	Model Coefficients	Marginal Effects	Model Coefs/ Marginal Effects
CF	-0.0808***		-0.111***		-0.00746***
	(0.0168)		(0.0218)		(0.00128)
Exposure*TDR	0.0337	0.0031	0.103*	0.0067*	0.00447
	(0.0285)	(0.0025)	(0.0560)	(0.0035)	(0.00651)
HHI	0.0177*	0.0017*	0.0327**	0.0023**	0.00148
	(0.00960)	(0.0009)	(0.0133)	(0.0009)	(0.00103)
Heavy_Rain_Origin	0.189***	0.0186***	0.264***	0.0186***	0.0180***
	(0.0281)	(0.0009)	(0.0363)	(0.0009)	(0.000964)
Frozen_Rain_Origin	0.435***	0.0428***	0.601***	0.0423***	0.101***
	(0.0622)	(0.0011)	(0.0798)	(0.0011)	(0.00187)
Heavy_Rain_Dest	0.176***	0.0173***	0.247***	0.0174***	0.0172***
	(0.0263)	(0.0009)	(0.0343)	(0.0009)	(0.000958)
Frozen_Rain_Dest	0.399***	0.0392***	0.554***	0.039***	0.0966***
	(0.0572)	(0.0011)	(0.0738)	(0.0011)	(0.00188)
Load Factor	-0.197***	-0.0194***	-0.259***	-0.0183***	-0.0191***
	(0.0289)	(0.0008)	(0.0353)	(0.0007)	(0.000799)
Labor Strikes	0.0315***	0.0034***	0.0506***	0.004***	0.00371***
	(0.00762)	(0.0007)	(0.0107)	(0.0007)	(0.000821)
Depart_6am_to_12_noon	0.000852	0.0001	-0.00110	-0.0001	0.000848
	(0.00848)	(0.0008)	(0.0109)	(0.0008)	(0.000899)
Depart_12_noon_to_6pm	0.0290***	0.0029***	0.0384***	0.0027***	0.00360***
	(0.0100)	(0.0009)	(0.0128)	(0.0008)	(0.000955)
Depart_6pm_to_12_midnight	0.0388***	0.0038***	0.0508***	0.0036***	0.00401***
	(0.0113)	(0.001)	(0.0143)	(0.0009)	(0.00105)
Connections_Origin	-0.000202**	-1.98E-05**	-0.000151	-1.06E-05	-1.22e-05
	(9.18e-05)	(8.53E-06)	(0.000117)	(8.10E-06)	(8.66e-06)
Connections_Dest	5.09e-06	5.01E-07	0.000131	9.20E-06	1.44e-05*
	(8.62e-05)	(8.49E-06)	(0.000117)	(8.20E-06)	(8.73e-06)
Congestion_Origin	0.0142	0.0014*	0.0216*	0.0015*	-0.000606
	(0.00882)	(0.0008)	(0.0117)	(0.0008)	(0.000789)
Congestion_Dest	0.0345***	0.0034***	0.0451***	0.0032***	0.000892
	(0.0101)	(0.0009)	(0.0132)	(0.0008)	(0.000798)

Model	Fractional Probit				Linear Regression
Exposure variable	Delay (>=60min) rel. freq. in 2007		Delay (>=60min) rel. freq., all pre-TDR periods		
Estimates	Model Coefficients	Marginal Effects	Model Coefficients	Marginal Effects	Model Coefs/ Marginal Effects
SEA_Runway_Origin	0.0126	0.0013	0.0122	0.0009	4.49e-05
	(0.00841)	(0.0009)	(0.0115)	(0.0008)	(0.000499)
ORD_Runway_Origin	-0.0578***	-0.0048***	-0.0825***	-0.0049***	-0.0112***
	(0.00943)	(0.0003)	(0.0125)	(0.0003)	(0.000860)
IAD_Runway_Origin	-0.00377	-0.0004	-0.00935	-0.0006	-0.00101
	(0.00820)	(0.0008)	(0.0108)	(0.0007)	(0.00124)
SEA_Runway_Dest	-0.00536	-0.0005	-0.00796	-0.0006	-0.000659
	(0.00964)	(0.0009)	(0.0130)	(0.0009)	(0.000521)
ORD_Runway_Dest	-0.0529***	-0.0045***	-0.0762***	-0.0046***	-0.0115***
	(0.00853)	(0.0003)	(0.0114)	(0.0003)	(0.000911)
IAD_Runway_Dest	-0.00809	-0.0008	-0.0174*	-0.0012*	-0.00245*
	(0.00778)	(0.0007)	(0.0102)	(0.0007)	(0.00127)
Aircraft Type Group 3	-0.0348***	-0.0032***	-0.0428***	-0.0029***	-0.00449***
	(0.00692)	(0.0004)	(0.00864)	(0.0004)	(0.000650)
Aircraft Type Group 4	-0.0598***	-0.006***	-0.0713***	-0.0051***	-0.00649***
	(0.0100)	(0.0006)	(0.0119)	(0.0005)	(0.000698)
Aircraft Type Group 5	-0.0558***	-0.0046***	-0.0698***	-0.0042***	-0.00629***
	(0.0121)	(0.0006)	(0.0155)	(0.0006)	(0.00103)
Constant	-2.450***		-2.841***		-0.0146
	(0.392)		(0.467)		(0.0387)
Observations	298,993		333,447		333,447

Robust standard errors in parentheses.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Exhibit B. Analysis of Cancellations in Technical Detail

Table B5. Dynamic Model of Percentage of Cancelled Flights

Exposure variable	Percentage of delays (>=60min) in 2007		Percentage of delays (>=60min) in all pre-TDR periods	
Estimates	**Coefficients**	**Marginal Effects**	**Coefficients**	**Marginal Effects**
Exposure*1(Aug2009 to Dec 2009)	-0.0725*	-0.005*	-0.279***	-0.0137***
	(0.0429)	(0.0028)	(0.0901)	(0.0041)
Exposure*1(Jan2010 to Apr 2010)	-0.131***	-0.0192***	-0.342***	-0.0351***
	(0.0452)	(0.0056)	(0.0856)	(0.0078)
Exposure*1(May2010 to Apr2011)	0.110***	0.0118***	0.195***	0.0154***
	(0.0395)	(0.0035)	(0.0684)	(0.0049)
Exposure*1(May2011 to Apr2012)	0.121**	0.0091***	0.310***	0.0163***
	(0.0474)	(0.0031)	(0.0911)	(0.0042)
Exposure*1(May2012 to Apr2013)	0.107**	0.0095***	0.204***	0.0128***
	(0.0427)	(0.0033)	(0.0775)	(0.0045)
Exposure*1(May2013 to Apr2014)	-0.376***	-0.039***	-0.790***	-0.0575***
	(0.0897)	(0.0055)	(0.150)	(0.0079)
Exposure*1(May2014 to Dec2014)	-0.173**	-0.0122**	-0.285**	-0.014**
	(0.0804)	(0.0052)	(0.143)	(0.0068)
Observations	298,993		333,447	

Robust standard errors in parentheses.
Includes same control variables as the model shown in Table B4.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

.

.

Table B6. Model of Percentage of Cancelled Flights by Season

Model	**Fractional Probit, Seasonal**			
Exposure variable	**Delay (>=60min) rel. freq in 2007**		**Delay (>=60min) rel. freq., all pre-TDR periods**	
Estimates	**Coefficients**	**Marginal eff.**	**Coefficients**	**Marginal eff.**
Exposure*Summer*TDR	0.116***	0.0091***	0.324***	0.0182***
	(0.0408)	(0.0026)	(0.0811)	(0.0036)
Exposure*Winter*TDR	-0.120***	-0.0146***	-0.301***	-0.0261***
	(0.0410)	(0.0041)	(0.0799)	(0.0058)
Observations	298,993		333,447	

Robust standard errors in parentheses.
Includes same control variables as the model shown in Table B4.

*** Significant at the 1% level. ** Significant at the 5% level. *Significant at the 10% level

Table B7. Model of Percentage of Cancelled Flights by Carrier Type

Model	**Fractional Probit, Low Cost**			
Exposure variable	**Delay (>=60min) rel. freq in 2007**		**Delay (>=60min) rel. freq., all pre-TDR periods**	
Estimates	**Coefficients**	**Marginal eff.**	**Coefficients**	**Marginal eff.**
Exposure*lowcost*TDR	0.319***	0.0146***	0.539***	0.0209***
	(0.119)	(0.004)	(0.155)	(0.0055)
Observations	298,993		333,447	

Robust standard errors in parentheses.
Includes same control variables as the model shown in Table B4.

*** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table B8. Estimated Real-Synthetic Cancellation Rate Differences (in the First 3 Post-TDR Years)

Exposure*	Mean difference	t-statistic	p-value	95% confidence interval	
Delay>=60min	0.00394	2.15651	0.037996	0.0002317520	0.007676085
Delay>=90min	0.00490	3.04392	0.004413	0.0016310110	0.008163042
Delay>=120min	0.00561	3.22609	0.002722	0.0020800650	0.009141647

Includes same control variables as the model shown in Table B4.
*All exposure measures based on the entire pre-TDR sample.

Figure B1. High- and Low-exposure Flights per Route-carrier

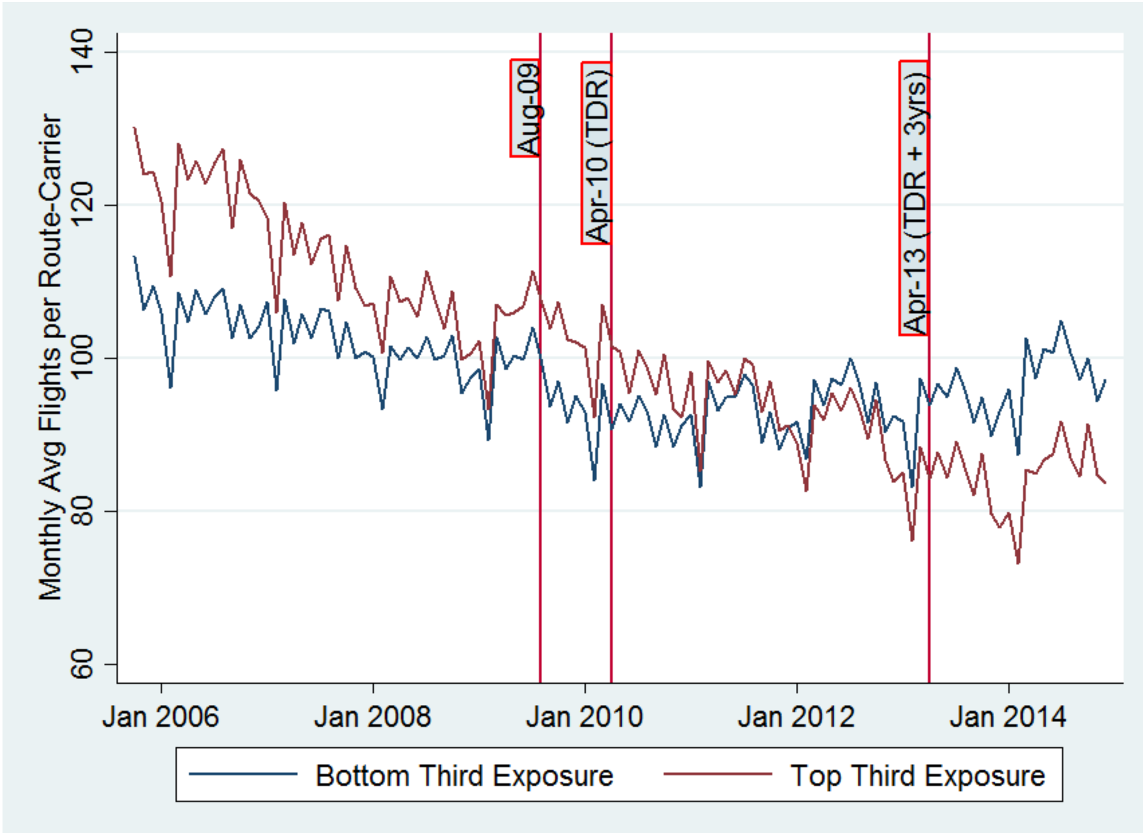


Exhibit B. Analysis of Cancellations in Technical Detail

Figure B2. Synthetic Group Results (Exposure: Delay >=60 min)

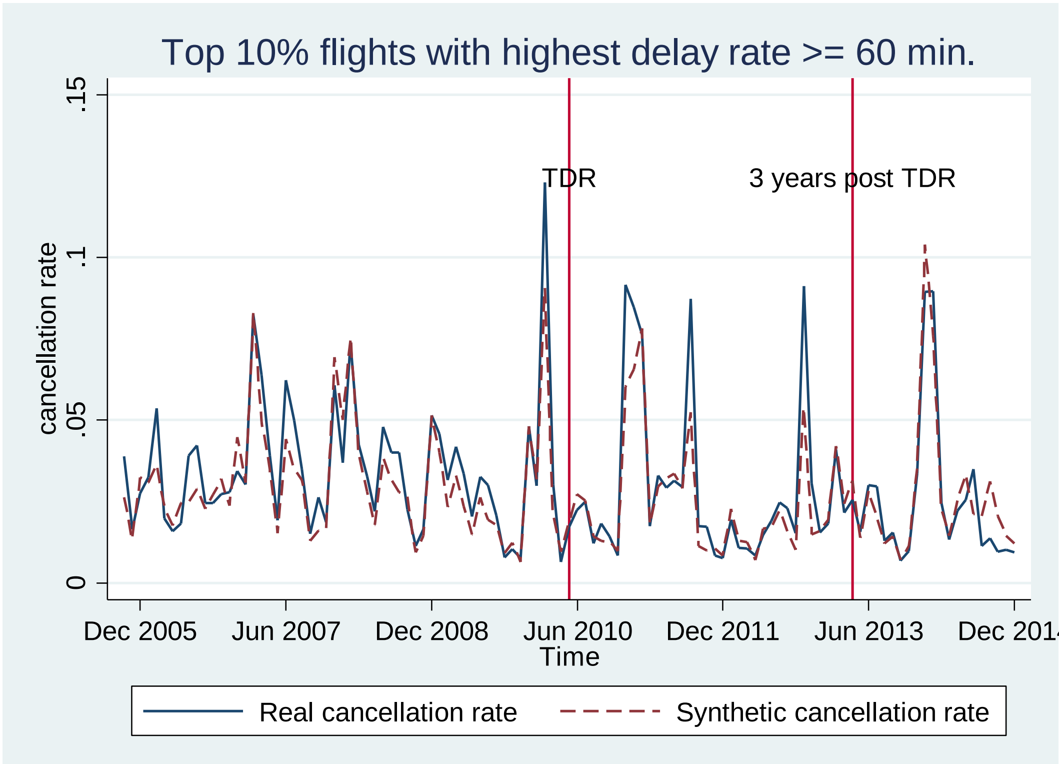


Figure B3. Synthetic Group Results (Exposure: Delay >=90 min)

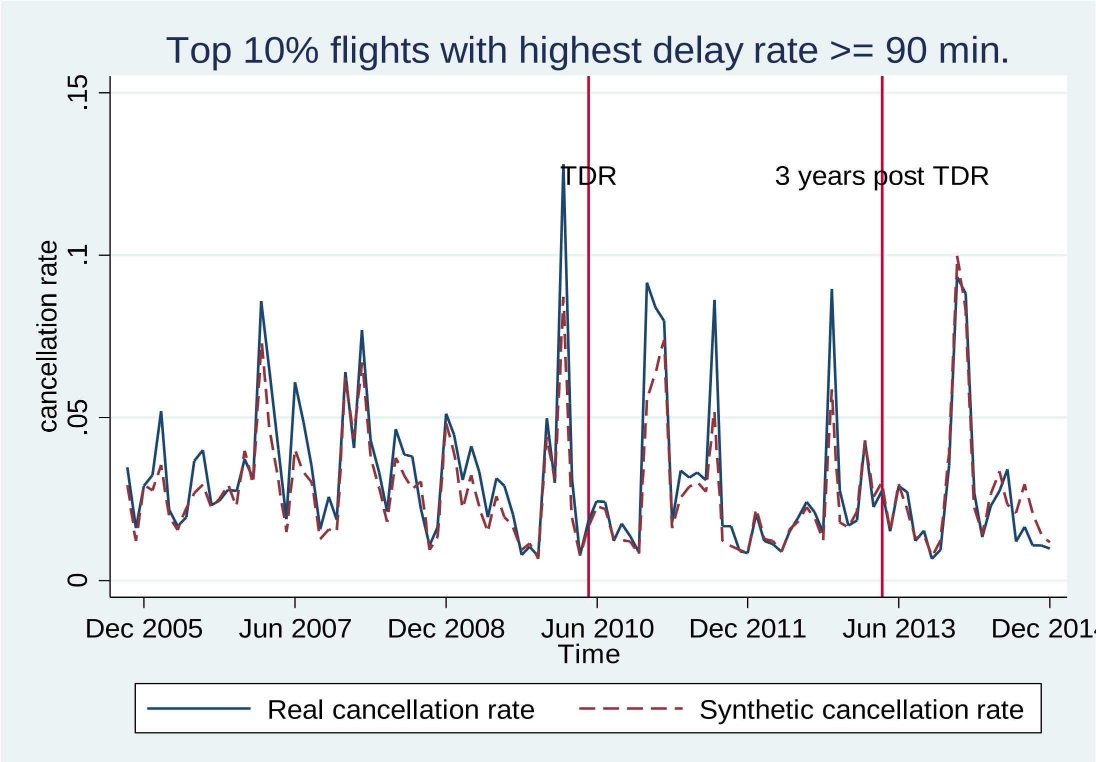


Figure B4. Synthetic Group Results (Exposure: Delay >=120 min)

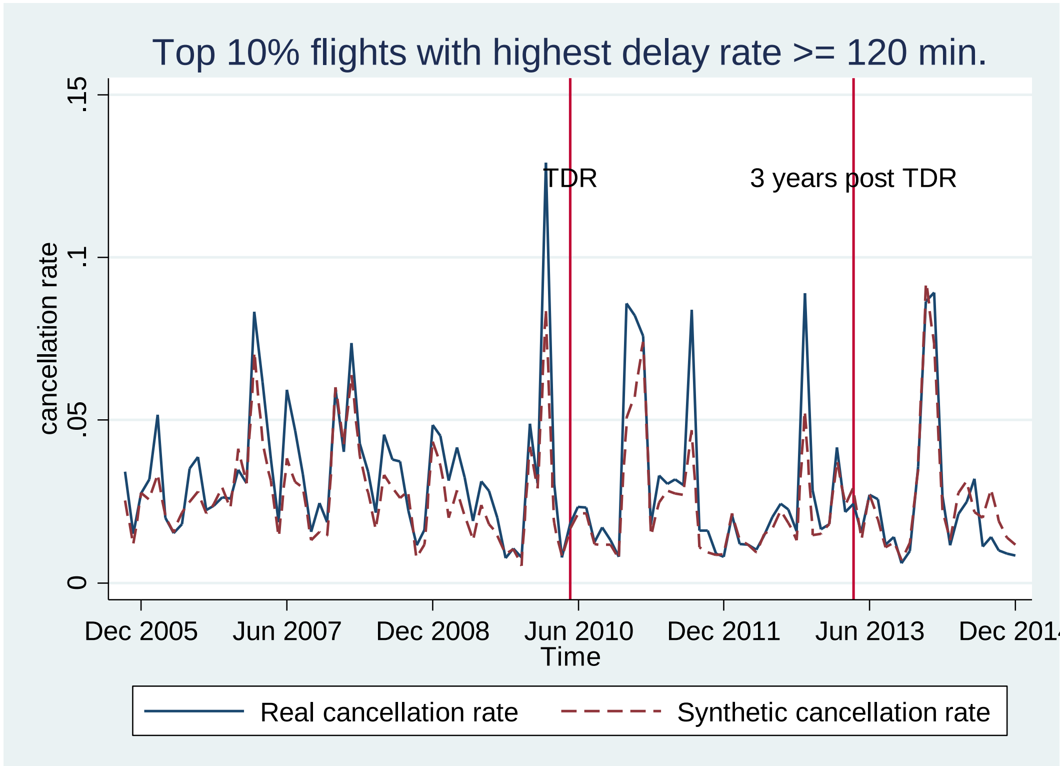


EXHIBIT C. MAJOR CONTRIBUTORS TO THIS REPORT

Name	Title
Betty Krier	Supervisory Economist/ Program Director
Jerrod Sharpe	Senior Economist
Evan Rogers	Economist
Joao Macieira	Economist
Audre Azuolas	Writer/Editor