Institute of Transportation Studies
University of California at Berkeley

Multivariate Analysis of the Impacts of NAS Investments: A Case Study of a Major Capacity Expansion at Dallas-Fort Worth Airport

Mark M. Hansen
Wenbin Wei

RESEARCH REPORT
UCB-ITS-RR-98-11

NATIONAL CENTER OF EXCELLENCE FOR
NEXTOR
AVIATION OPERATIONS RESEARCH

NEXTOR Research Report RR-98-18

January 1999
ISSN 0192 4095

Preface and Acknowledgments

This report documents research undertaken by the National Center of Excellence for Aviation Operations Research, under Federal Aviation Administration Research Grant Number 96-C-001. This document has not been reviewed by the Federal Aviation Administration (FAA). Any opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation.

The financial support of this research by the FAA is gratefully acknowledged. The authors wish to thank Steve Bradford, Norm Fujisaki, Randy Stevens, and Ronnie Uhlenhaker of the FAA for their guidance and support of this research.

Table of Contents

List of Figures

List of Tables

Multivariate Analysis of the Impacts of NAS Investments: A Case Study of A Major Capacity Expansion at DFW

ABSTRACT

This paper presents a case study of how to empirically estimate the impact of a NAS investment on system performance. While a variety of analytical and simulation tools have been developed to predict benefits of NAS investments prior to implementation, relatively little effort has gone into monitoring and analyzing what actually occurs after implementation. Such a posteriori studies are critical in order to determine whether anticipated benefits are actually realized, and in identifying unintended consequences of investments and their benefit implications.

In our case study, we employ multivariate statistical analysis to estimate the impacts of a major capacity expansion at DFW airport, including a new runway and airspace reconfiguration, which occurred in October, 1996. Our multivariate model relates a measure the daily average flight time for flights arriving at DFW, termed the Daily Flight Time Index (DFTI), to demand, weather, origin airport congestion, and the expansion itself. A quadratic functional form, which allows for non-linear and interaction effects, is employed, and estimated in a manner which avoids the problem of heteroscedasticity in the original data. The model is estimated for the DFTI and for three components of the DFTI—departure delay, airborne time, and taxi time.

The effect of the capacity expansion on DFTI is found to depend strongly on visibility. On most days the DFTI has changed by two minutes or less as a result of the expansion, but on low visibility days we estimate reductions in the 4-10 minute range. On average, the DFTI in the post-expansion period is 1.3 minutes less as a consequence of investment. This overall change includes a larger reduction in departure delay which is offset by an increase in taxi time. Moreover, the DFTI reduction caused by the expansion has been more than offset by increases in DFTI resulting from other factors, notably increased demand and worse weather. The effect of increased demand raises this issue of whether the expansion itself caused demand to increase. We show that the extent of this "induced demand" effect is critical in assessing the overall impact of the capacity expansion on delay.

1. INTRODUCTION

The need for and benefits of increased airport capacity have been the subject of extensive study, both at the broad policy level and in the context of specific proposals to improve the National Airspace System (NAS). Airport capacity can be increased through conventional means such as construction of new runways, or through deployment of advanced technologies. This paper assesses the impact of a major expansion in airfield capacity at Dallas-Fort Worth International Airport (DFW). On October 1, 1996, a new runway, 35R/17L, was commissioned. Nine days later, the DFW Metroplex Plan, which included a major airspace reconfiguration along with new air traffic control procedures, was implemented. These changes allowed triple and quadruple independent arrival streams under IFR and VFR conditions, respectively. Arrival capacities increased from 102 to 146 under VFR, and from 66 to 108 under IFR (Uhlenhaker, 1998).

There are a variety of approaches to assessing the benefits of increased airport capacity. The vast majority of capacity/delay studies endeavor to predict how proposed changes will affect system performance based on queuing models and computer simulation tools (Odoni et al, 1997). Such models are essential for informing investment decisions, but require many simplifying assumptions and are subject to a variety of errors. Another approach is retrospective, statistical, and based directly on empirical observation. Such retrospective studies cannot directly inform decisions, but are able to observe what actually happens in the aftermath of a NAS investment. They therefore provide a useful check on predictive, model-based, studies.

While there have been prior efforts to statistically analyze airport delay data, few if any efforts have attempted to use such analysis to study the impact of a capacity enhancement, or more generally to analyze variation of delay at a given airport over an extended period of time. Geisinger (1989) used Standardized Delay Reporting System (SDRS) data to estimate national trends in the amount and cost of delay from the mid-70s to the mid-80s, and to assess, using cross-sectional airport-level data, the relationship between the incidence of “serious” delays and the average delay per flight. FAA (1987), presented a cross-sectional model, based on data for 32 airports, relating the average delay per operation at an airport to the ratios of air carrier and general aviation operations to capacity under IFR and VFR conditions. Morrison and Winston (1989) used a two stage technique, first analyzing individual flight times to estimate airport-specific, hourly, arrival and departure delay coefficients, and then regressing these coefficients on the ratios of general and air carrier operations per runway in the observed hour. Poldy (1982), also estimated hourly delay coefficients by airport, reporting results for Sydney’s Kingsford Smith Airport. None of these studies attempts to model how delay at a particular airport varies over a multi-day period, or to measure the impact of capacity enhancements at an airport.

Hansen et al. (1998) studied the impact of the capacity expansion at DFW. They considered both operational impacts and ways in which airlines operating at DFW may have adapted to the new capacity. A broad range of impacts and adaptations were observed. However, since their study is primarily descriptive and based on before/after comparisons, it is not able to isolate the impacts of the expansion from other, potentially confounding, influences. This paper adopts a more rigorous statistical approach to assess the impact of the expansion on flight delay. It presents a multivariate analysis relating delay to multiple causal factors simultaneously, thereby allowing a more precise determination of the effect of the expansion.

A central question in empirically evaluating the delay impact of a capacity expansion is how much delay would have occurred if the capacity expansion had not taken place. Since such "counterfactual" delay is not observable, the question cannot be answered by simple empirical observation. Moreover, delay at an airport depends upon many factors, including demand, weather, congestion at "upstream" airports, and airline schedule padding. One must control for these influences in order to isolate the capacity expansion impact. In this paper, we do so by estimating multiple regression models which predict daily average flight times for DFW arrivals. The models have a flexible functional form, and yield statistically robust estimation results. The estimated models give a plausible, and reasonably detailed, picture of how the 1996 expansion influenced arrival delays at DFW. This paper also considers the question of how airlines responded to the new capacity, and the impact of this response on delay. We offer evidence that airlines "traded in" some delay reduction for schedule changes, and some tentative estimates of the extent to which this occurred. These estimates suggest that, in this particular case, scheduling changes resulting from the expansion offset some, but not all, of the delay reduction from the expansion.

Our results suggest that delay reductions from the capacity expansion are well below what was predicted in certain pre-expansion analyses. These studies are not published in the open literature, and it is not our objective to assess the extent of or reasons for the apparent discrepancies. Rather, our aim is to demonstrate that despite the tremendous stochasticity surrounding airport operations, careful statistical analysis can tease out relatively modest impacts from real-world operational data.

The remainder of this paper is organized as follows. Section 2 briefly introduces the methodology. Section 3 specifies the regression model. Numerical results are reported in Section 4. Section 5 concludes the paper.

2. METHODOLOGY

Our study is confined to the impacts of the expansion on delays of DFW-bound flights. We do not consider the propagation of these delays to other flights, nor do we consider impacts on flights departing DFW. We track changes in arrival delay using a metric which measures the time interval from scheduled departure to actual arrival of the average DFW-bound flight on a given day. We refer to this time interval for a particular flight as the Effective Flight Time (EFT), and to the daily average as the Daily Flight Time Index (DFTI). Details of the averaging procedure are discussed below.

Our approach contrasts with many other delay studies, which employ as measures of airline delay the difference between scheduled and actual arrival or in-flight times. We refer to such measures as Delay Against Schedule (DAS). DAS metrics are not appropriate for this study because they can be manipulated by "padding" the schedule to reflect anticipated delays, and therefore may not accurately reflect how the expansion affected actual delays. On the one hand, some airlines may have reduced schedule padding in response to improved conditions at DFW. On the other hand, there is evidence that American Airlines actually increased padding in the months subsequent to the capacity expansion in order to improve its on-time performance. A comparison of American's scheduled flight times between DFW and 20 major important destinations in January 1996 and January 1998 reveals that scheduled flight times increased by an average of approximately 4 minutes. To eliminate spurious effects related to changes in schedule padding, a more refined methodology that controls for such scheduling changes is required.

Past efforts to control for the schedule padding problem have generally revolved around identifying a lower bound flight time for a particular segment and defining delay as the difference between the time for a given flight and this lower bound. Odoni (1995) reports that the FAA, in the delay statistics reported in the *Aviation System Capacity Enhancement Plan*, uses as the lower bound the 5th percentile flight time. While such efforts are necessary in order to determine the total amount of delay, they are not required in order to measure the *change* in delay over time. Such changes can be detected from changes in the EFT, if we define delay as the difference between the observed EFT and a hypothetical lower bound EFT which is stable over time. In this case changes in delay and changes in EFT are equivalent, as are changes in average delay and changes in DFTI. We use these terms interchangeably in this paper. In taking this approach, we are including as changes in delay certain phenomena, for example longer taxi routes or more circuitous terminal approaches, that others might label differently. Semantics aside, such changes must clearly be taken into account in assessing the net impact of the expansion.

In order to estimate the impact of airport capacity expansion on the DFTI, we develop a regression model relating it to various explanatory variables, based on daily observation data. We extend the work reported in Hansen et al. (1998) and use the metrics developed there, together with some new metrics defined below, as explanatory variables for the regression model. We incorporate the airport capacity expansion as a 0-1 dummy variable (0 for the days before expansion and 1 for the days after expansion). We then estimate the daily delay reduction effect of the capacity expansion by fixing all variables except the dummy variable at the daily observed values, and then comparing the predicted DFTI when the expansion is "turned on" and "turned off." We also use the results to estimate the contributions of other factors, such as demand and weather, to observed differences in delay before and after the expansion. Furthermore, these contributions as well as the delay impacts are decomposed by flight time component—yielding, for example, the impact of the expansion on departure delay.

3. MODEL DEVELOPMENT

The regression model is built based on the relationship:

$$DFTI_t = f(Demand_t, Weather_t, Origin\_Airport\_Congestion_t, Expansion_t) + e_t$$

This equation states that average flight time of DFW-bound flights on day t depends on airline demand, airport weather conditions, and congestion at origin airports on day t, and whether day t is prior to or after the October, 1996 capacity expansion. In addition, the model includes an additive error term which encompasses a variety of unobserved factors. It is important to note that this model makes predictions on a daily basis. The daily level specification is desirable for a statistical model because it allows us to assume that the stochastic error terms are independent, and also allows use of coarser data for variables such as weather and demand. The increased simplicity of the daily model comes at the price of the loss of ability to predict delays on a more disaggregate basis (e.g., by time of day or type of flight) and to accurately capture certain sub-day-scale phenomena, such as temporary airport closure due to thunderstorms. An important research question is how successful a model with such limitations can be in explaining day-to-day variation in delay.

3.1 Metric Development

To operationalize the above model, it is necessary to develop metrics which capture day-to-day variation in each of the included variables. We discuss these metrics below.

3.1.1 Flight Time

As previously noted, our flight time measure, the DFTI, is a daily average of the Effective Flight Time (EFT), defined as the time interval between *scheduled* departure and *actual* arrival, for completed flights into DFW. This metric includes departure delay and gate-to-gate travel time, and it can also be divided into taxi time (taxi-out time at the origin airport and taxi-in time at DFW), airborne time, and departure delay. Unlike DAS metrics, EFT does not change as a result of changes in schedule padding. The EFT does, however, depend on the origin and destination of a flight. To track overall changes in the EFT for flights going into DFW, it is necessary to average across flights in a way that maintains a fixed "market basket" of flight origins. To do this, we first identified all origins that had at least five completed flights into DFW for every day in the period from January 1, 1995 to March 31, 1998. 53 airports, accounting for about 90 percent of all flights into DFW by major air carriers, satisfied this criteria. For each such origin airport, we calculated the daily average EFT across its flights to DFW, which we term the Daily Average Flight Time (DAFT). Finally, we calculate the Daily Flight Time Index (DFTI) as a weighted average of the origin-specific DAFTs, with the weights based the origins' shares of completed flights into DFW in the first quarter of 1997. The DFTI is analogous to a consumer price index, where the prices are flight times and the commodities are origins. (In contrast to the unit-less price index, we express the DFTI in time units.) Like the EFT, the DFTI can be decomposed into departure delay (gate time), taxi-out time at origin airport and taxi-in time at DFW (taxi time), and airborne time.

To formalize the above description, let I be the set of origins which had at least five completed flights into DFW on every day from January 1, 1995 to March 31, 1998. Let EFT_{ift} be the Effective Flight Time of flight f from origin i into DFW on day t, and let F_{it} be the number of completed flights into DFW from origin i on day t. The Daily Average Flight Time is given by:

$$DAFT_{it} = \frac{\sum_{f=1}^{F_{it}} EFT_{ift}}{F_{it}} \tag{1}$$

The Daily Flight Time Index is calculated as:

$$DFTI_t = \frac{\sum_{i \in I} W_i \, DAFT_{it}}{\sum_{i \in I} W_i} \tag{2}$$

where the weight W_i is the number of completed flights into DFW from origin i in the first quarter of 1997. The development of daily metrics for the components of effective flight time, including departure delay, taxi time, and airborne time, proceeds in similar fashion.

To measure the DFTI, we used the Airline Service Quality Performance (ASQP) database. For each flight included in the database, the ASQP reports the scheduled departure and arrival times, the actual departure and arrival times, and the actual wheels-off and wheels-on times. ASQP contains data for all domestic flights of the ten major air carriers, which account for 60 percent of all aircraft operations into DFW. Thus our estimates apply most directly to impacts experienced by major airlines, although they are suggestive for other users as well.

Figure 1 plots the DFTI time series. The lower bound is slightly less than 150 minutes. Numerous spikes, exceeding 200 minutes in several cases, are evident. Thus there are days when the average flight delay approaches 1 hour. There is no discernable change in DFTI after the capacity expansion, except perhaps for a modest reduction in the incidence of spikes.

3.1.2 Demand

Airport demand can be captured by a variety of metrics. The number of daily operations is a general measure of the magnitude of demand, but this measure doesn't capture the schedule of these operations, which is very critical to delay at an airport, particularly one with large connecting flight banks such as DFW. The peak rate of scheduled arrivals is another potential metric. However, it is based entirely on a single period of the day when the peak occurs, and is insensitive to the duration and number of peaks. To overcome these deficiencies, Hansen et al. (1998) proposed a measure called Hypothetical Deterministic Delay (HDD) as a gauge of the concentration of an airline's daily schedule and the degree of banking. This *hypothetical* delay is a function of the daily flight schedule and an *assumed* capacity level (90 operations per hour in our study). The flight schedule can be for arrivals only, departures only, or both arrivals and departures; here we use the arrival HDD. Figure 2 illustrates how it is calculated. The airport is treated as a single-server queuing system. The top curve in the figure represents the cumulative "customer arrivals" based on the scheduled arrival times at DFW, while the bottom curve, whose maximum slope corresponds to the capacity, represents cumulative "customer departures." HDD is the area between the two curves, divided by the number of flights. It captures the average queuing delay per flight that would be expected given the assumed capacity and flight schedule. The HDD used in this study is based on scheduled arrival times of flights in the ASQP data base.

Figure 3 plots the evolution of the seven-day moving average arrival HDD demand metric from January 1995 through March 1998. Substantial fluctuations are evident. There was a sharp drop during early 1995, followed by an increase beginning in late 1995, including a sharp jump in the summer of 1996. Further increases, albeit with occasional downward fluctuations, are observed after the expansion.

3.1.3 Weather

Weather is obviously an important determinant of delay. For a statistical delay model to be credible and successful, it must control for variation in weather. We obtained daily weather information for DFW from the National Weather Service Global Surface

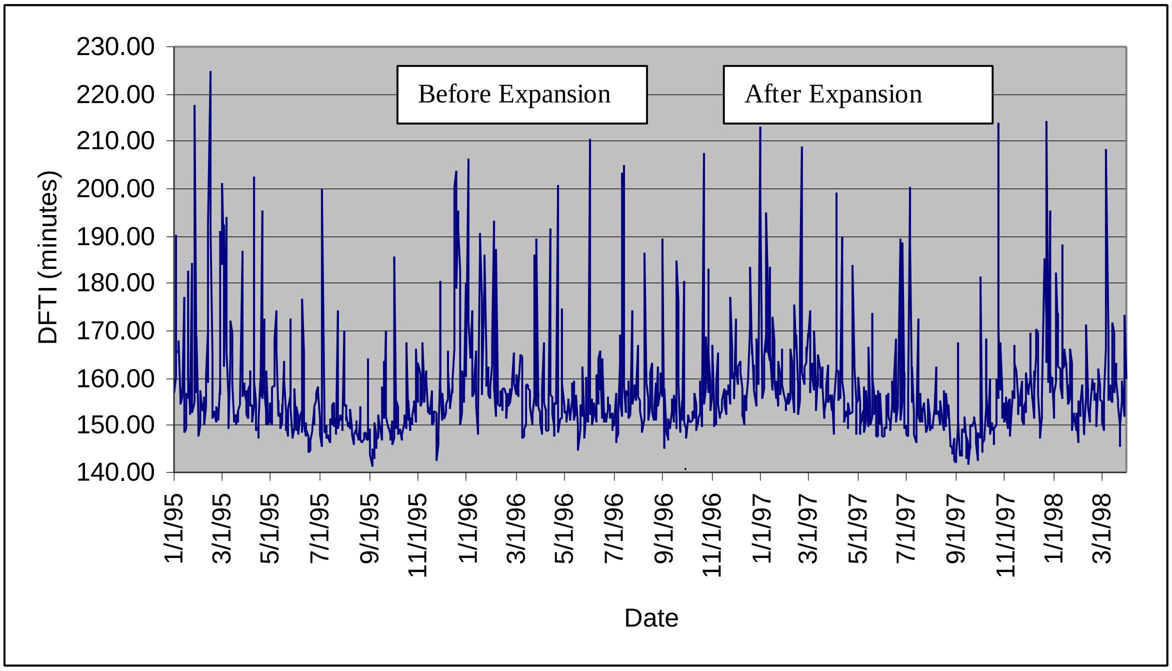

Figure 1: Daily Flight Time Index (DFTI) Time Series

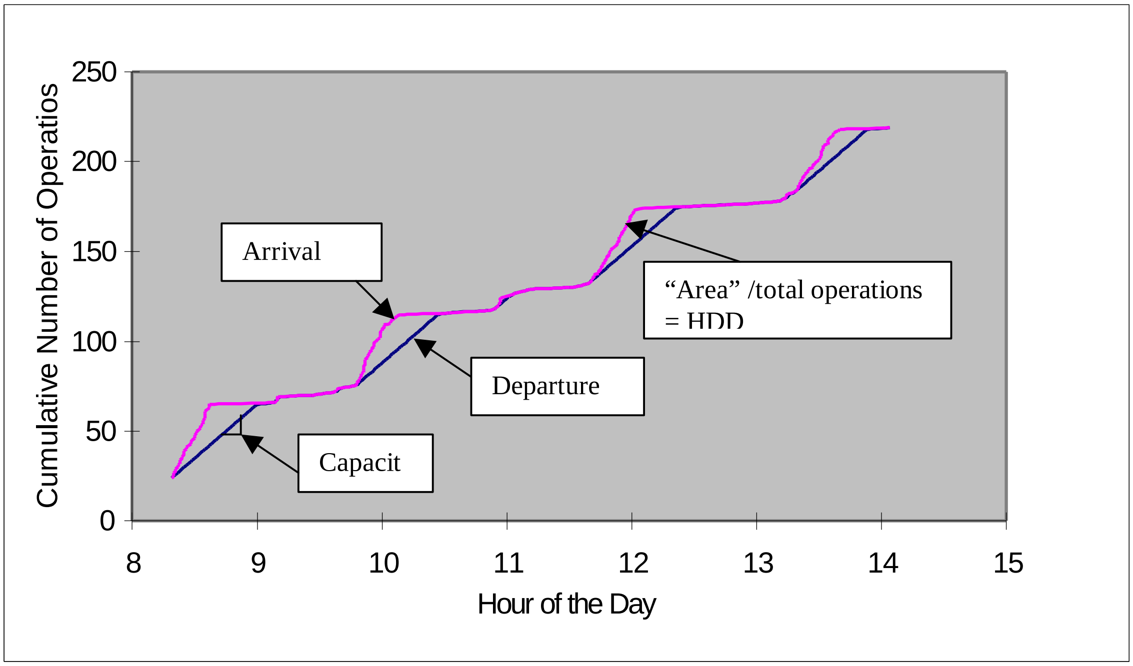

Figure 2: Method for Computing Hypothetical Deterministic Delay (HDD)

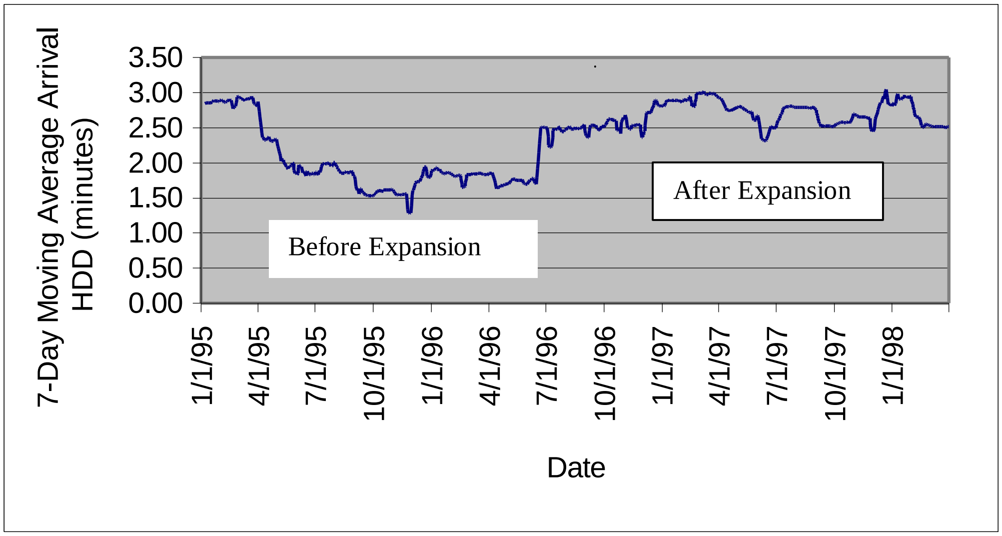

Figure 3: Daily Arrival HDD Before and After the Expansion

Summary (GSS). The variables reported in the weather data are described in Table 1. In general, the data elements pertain to temperature, wind, precipitation, visibility, and barometric pressure. Since the data are daily, they do not fully capture weather effects that influence delay. For example, the daily visibility does not directly measure the occurrence of IFR conditions, and the 1-0 thunderstorm variable does not differentiate between cases where there was a single, inconsequential storm and those when a chain of storms shut down the airport. These limitations notwithstanding, we chose to use the daily data for several reasons. First, it is readily available and easy to work with. Second, since the model requires daily-level data, it is necessary to employ weather variables defined at that temporal scale. Although hourly level weather data might be aggregated to create a richer set of daily weather variables, this would require substantial effort for success, and it is not clear that it would result in a better model. As discussed below, we found that using the GSS data to control for weather effects in our model worked quite well, although there is clearly room for improvement.

Based on these variables describing the weather situation at DFW, a statistical technique, factor analysis, was used to collapse the original 15 variables into a more manageable number. Four orthogonal factors, which are linear combination of all these original variables, were derived. The linear coefficients relating the four factors to the fifteen variables are shown in Table 2. From the coefficients, it is apparent that FACTOR1 has high correlation to temperature; FACTOR2 describes the wind speed; FACTOR3 relates to precipitation; and FACTOR4 corresponds to visibility. The factors are all standardized variables with mean 0 and standard deviation 1. Thus a day with positive scores for all factors would be warmer, windier, wetter, and clearer than average. All four factors were employed in the regression model.

3.1.4 Origin Airport Congestion

The EFT of a DFW-bound flight can be affected by conditions at the origin airport, or elsewhere in the National Airspace System, which result in a departure delay. The methodological challenge is to control for these effects without masking the effect of ground-hold programs resulting from congestion at DFW, which also cause departure delay. In order to do this, we calculate a metric based on the average departure delay at origin airports for flights whose destinations are not in the DFW area. As with the DFTI flight time index, the average is weighted to reflect each origin airport's share of flights going into DFW. In mathematical terms, define E_{it} as the set of flights originating at point i on day t whose destination is not in the DFW region, and let N_{it} be the number of flights in E_{it}. Let D_f be the departure delay (actual departure time – schedule departure time) for flight f. Then the Origin Airport Congestion for day t is given by:

$$OAC_t = \frac{\sum_{i \in I} \left(W_i \frac{\sum_{f \in E_{it}} D_f}{N_{it}} \right)}{\sum_{i \in I} W_i} \qquad (3)$$

Weather Data	Description
TEMP	Mean temperature for the day in degrees Fahrenheit to tenths
DEWP	Mean dew point for the day in degrees Fahrenheit to tenths
SLP	Mean sea level pressure for the day in millibars to tenths
STP	Mean station pressure for the day in millibars to tenths
VISIB	Mean visibility for the day in miles to tenth
WDSP	Mean wind speed for the day in knots to tenths
MXSPD	Maximum sustained wind speed reported for the day in knots to tenths
MAX	Maximum temperature reported during the day in Fahrenheit to tenths
MIN	Minimum temperature reported during the day in Fahrenheit to tenths
PRCP	Total precipitation reported during the day in inches and hundredths
FOG	Fog (1=yes, 0=no/not reported)
RAIN	Rain (1=yes, 0=no/not reported)
SNOW	Snow (1=yes, 0=no/not reported)
HAIL	Hail (1=yes, 0=no/not reported)
THUNDER	Thunder (1=yes, 0=no/not reported)

Table 1: Description of Weather Data

	FACTOR1 (Temperature Factor)	FACTOR2 **(Wind Factor)**	**FACTOR3 (Precipitation Factor)**	**FACTOR4 (Visibility Factor)**
TEMP	0.96	-0.07	-0.01	0.17
DEWP	0.94	-0.07	0.11	-0.03
SLP	-0.74	-0.44	-0.04	0.28
STP	-0.70	-0.46	-0.04	0.31
VISIB	0.03	0.18	-0.17	0.79
WDSP	-0.03	0.92	-0.02	0.04
MXSPD	-0.03	0.90	0.19	0.05
MAX	0.92	-0.04	-0.05	0.22
MIN	0.94	-0.08	0.07	0.11
PRCP	-0.01	0.13	0.71	-0.12
FOG	-0.02	0.05	0.04	-0.73
RAIN	-0.01	0.07	0.74	-0.37
SNOW	-0.29	0.11	-0.07	-0.32
HAIL	0.00	-0.05	0.32	0.12
THUNDER	0.15	0.07	0.83	-0.04

Table 2: Correlation Coefficients between Weather Factors and Weather Variables

3.1.5 Capacity

As explained previously, the goal of our analysis is to isolate the impact of the new runway opening and Metroplex Plan implementation on delay. Since these two events occurred within ten days of each other, we treat them as a single event, and specify a 0-1 dummy variable to indicate the time periods before and after it. Furthermore, we exclude the entire month of October, 1996 from our data set, to avoid effects related to the initial transition to the new runway/airspace configuration.

An additional facility improvement beyond the runway and airspace changes may also have affected arrival delays at DFW. During the 1995-98 period, the Final Approach Spacing Tool (FAST), a component of NASA's Center Terminal Automation System (CTAS), was being tested. We did not control for the possible effect of FAST, since it was used on a relatively small number of days, and for a just a few hours on these days (Davis, 1997).

3.2 Model Specification and Estimation

For our initial model specification, we chose a generalized quadratic function:

$$Y_t = b_0 + \sum_{i=1}^{N} b_i X_{it} + \sum_{i=1}^{N}\sum_{j=1}^{N} l_{ij} X_{it} X_{jt} + a_0 E_t + \sum_{i=1}^{N} a_i E_t X_{it} + e_t \qquad (4)$$

where Y is the flight time index (DFTI) or one of its components (departure delay, taxi time, or airborne time), the $X_i's$ are the explanatory variables (including demand, weather, and origin airport congestion), E is a dummy variable for the capacity expansion, e is a stochastic error term, and the $b's$, $l's$, and $a's$ are coefficients to be estimated. The t subscript on all variables is used to indicate their value of day t. The $X_{it} X_{jt}$ are second order terms which capture interaction effects, for example how visibility influences the impact of origin airport delay, as well as quadratic effects, such as the impact of the origin airport delay squared. The EX_i terms capture interactions between the impact of the capacity expansion and other variables, for example how the impact of the expansion was affected by precipitation.

Estimation was performed initially on the total DFTI model, using ordinary least squares (OLS) and a specification including all interaction terms. Once the initial model (termed Model 1) was estimated, a series of statistical tests were performed to determine whether particular sets of second order terms could be eliminated from the model. The tests revealed that the null hypothesis that coefficients on interaction terms involving the demand variable are 0 could not be rejected. The model was therefore re-estimated without these coefficients set to 0, yielding Model 2. Models 1 and 2 were found to have heteroscedastic errors, with predictions less accurate on days with high predicted delays. This is evident in the residual plot for Model 2, shown in Figure 4. Further investigation revealed that error magnitudes correlated with the visibility weather factor. To eliminate the influence of heteroscedasticity on our estimates, we regressed the absolute value of the OLS error against the visibility factor. The delay equation was then re-estimated as:

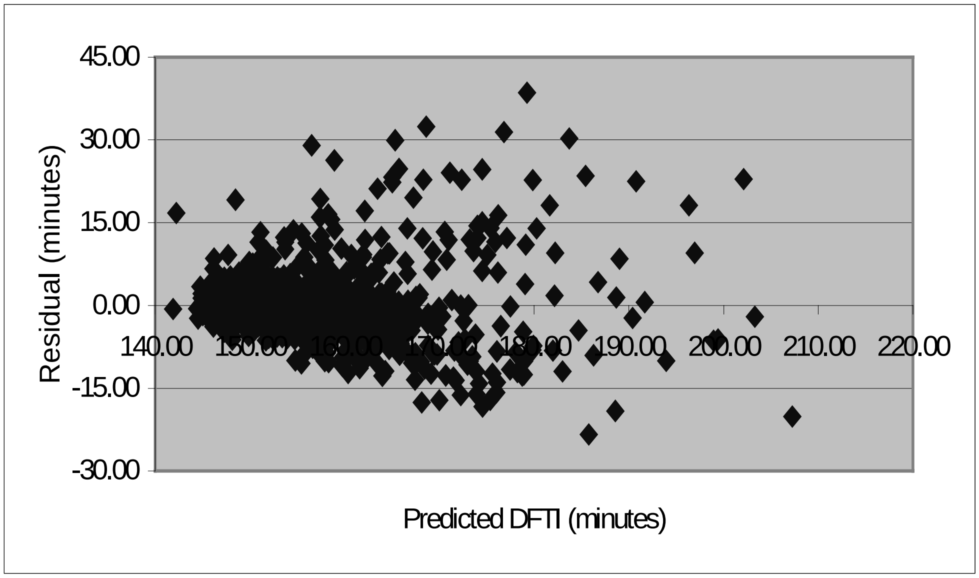

Figure 4: Residual Plot for Model 2

$$Y_t / C_t = b_0 / C_t + \sum_{i=1}^{N} b_i X_{it} / C_t + \sum_{i=1}^{N} \sum_{j=1}^{N} l_{ij} X_{it} X_{jt} / C_t + \\ a_0 E_t / C_t + \sum_{i=1}^{N} a_i E_t X_{it} / C_t + m_t \tag{5}$$

where C_t is the predicted value of the error in the original model on day *t*, as derived from its regression against visibility. Statistical tests again showed that demand interaction terms could be removed. We refer to the estimated equation (5) with the parameters for demand interaction terms set to 0 as Model 3. Inspection of the residuals from this model shows that the heteroscedasticity problem has been eliminated, as shown in Figure 5. Figure 5 also reveals that the errors are non-normally distributed, with the largest positive errors of greater magnitude than the largest negative ones. As discussed in Johnson (1984), non-normal errors do not strongly affect regression results, with point estimates for parameters remaining unbiased and efficient and standard error estimates asymptotically valid for large samples.

We then estimated models for the DFTI components: departure delay, airborne time, and taxi time. The component models were all specified identically to Model 3, and estimated using the same C_t values. This ensures additive consistency between the component models and the total DFTI model, so that effects of the expansion observed in the latter can be decomposed in effects on individual flight components.

4. RESULTS

4.1 Total DFTI

Table 3 summarizes our estimation results for the three models of total DFTI described above. Estimates appearing in bold in Table 3 are statistically significant at the 5 percent level; those appearing in bold italic are significant at the 1 percent level. Standard errors appear in parentheses below the estimates themselves. Model 1 includes all interaction terms, and is estimated using OLS. Model 2 eliminates second order terms involving demand, since these were found to be collectively insignificant, and is again estimated using OLS. The third model also excludes the second order demand terms, and is estimated using the transformed version of the original model (equation 5) to avoid the heteroscedasticity problem. Model 3 is therefore our preferred model. Note that the standard errors reported for Models 1 and 2 are those output from OLS, which are known to be biased when heteroscedasticity is present. The biased standard error estimates bias the reported significance levels as well.

In the preferred model, all first order terms are significant at the 1 percent level, with the exception of the expansion dummy, which is significant at the 5 percent level. Lower DFTI values are associated with the capacity expansion, high temperature, and high visibility. Higher DFTI values result from high demand, high delay at origin airports, high wind, and precipitation. Models 2 and 3 includes many more significant estimates for the first order terms than Model 1, suggesting that including the second order demand terms in the model causes serious multicollinearity problems. Model 3 yields a larger and more significant estimate for the first order expansion term than Model 2. In general,

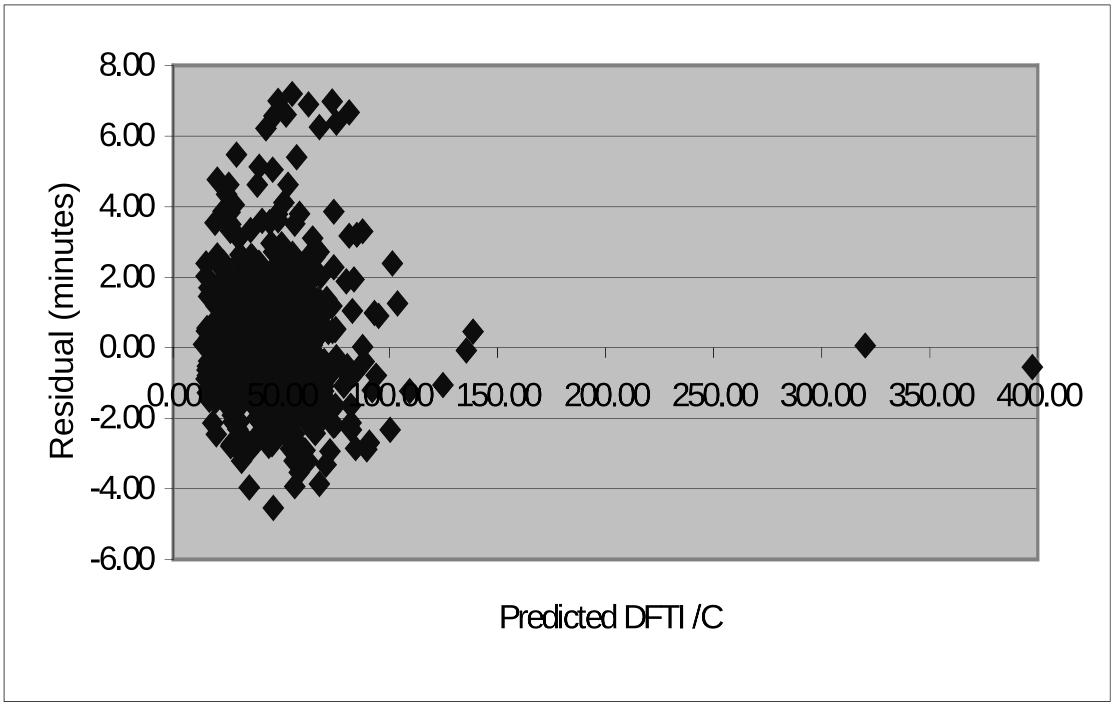

Figure 5: Residual Plot for Model 3

Table 3: Total DFTI Estimation Results

Variable	Estimate (Standard Error) Model 1	Model 2	Model 3
INTERCEPT	***146.118***	***144.903***	***146.274***
	(6.1705)	***(1.4507)***	***(1.0804)***
EXPANSION	0.4264	-0.7496	**-1.8336**
	(5.2887)	(0.9627)	**(0.7366)**
DEMAND	1.0134	***2.1756***	***1.9537***
	(5.8071)	***(0.5571)***	***(0.3933)***
ORIGIN CONGESTION	**0.7217**	***0.7052***	***0.5675***
	(0.3501)	***(0.1677)***	***(0.1283)***
TEMPERATURE	-0.9751	***-1.9439***	***-2.0023***
	(1.3035)	***(0.5526)***	***(0.4025)***
WIND	**2.8919**	***1.5096***	***1.401***
	(1.3337)	***(0.4889)***	***(0.3913)***
PRECIPITATION	1.1351	***1.5666***	***1.8841***
	(1.4499)	***(0.6060)***	***(0.4246)***
VISIBILITY	-2.2591	***-2.3321***	***-2.6448***
	(1.2335)	***(0.5221)***	***(0.6620)***
EXPANSION*DEMAND	-0.3733	--	--
	(1.9973)	--	--
EXPANSION*ORIGIN CONGESTION	-0.0749	-0.0673	0.0527
	(0.1485)	(0.0991)	(0.0724)
EXPANSION*TEMPERATURE	0.1075	-0.1466	-0.414
	(0.5710)	(0.4140)	(0.2755)
EXPANSION*WIND	0.4436	-0.028	-0.2011
	(0.5933)	(0.4198)	(0.3045)
EXPANSION*PRECIPITATION	0.0135	0.136	0.4007
	(0.5206)	(0.4221)	(0.3356)
EXPANSION*VISIBILITY	***2.3016***	***2.3018***	***2.2473***
	(0.5016)	***(0.4208)***	***(0.5375)***
DEMAND*DEMAND	0.2399	--	--
	(1.3451)	--	--
DEMAND*ORIGIN CONGESTION	-0.0006	--	--
	(0.1534)	--	--
DEMAND*TEMPERATUE	-0.4518	--	--
	(0.5816)	--	--
DEMAND*WIND	-0.6872	--	--
	(0.6231)	--	--

Variable	Estimate (Standard Error) Model 1	Model 2	Model 3
DEMAND*PRECIPITATION	0.2117	--	--
	(0.6147)	--	--
DEMAND*VISIBILITY	-0.0299	--	--
	(0.5232)	--	--
ORIGIN CONGESTION*ORIGIN CONGESTION	**0.0151**	**0.0153**	***0.0177***
	(0.0067)	**(0.0067)**	***(0.0049)***
ORIGIN CONGESTION*TEMPERATURE	0.0244	0.0251	0.022
	(0.0538)	(0.0534)	(0.0353)
ORIGIN CONGESTION*WIND	-0.0642	-0.0643	-0.0341
	(0.0483)	(0.0478)	(0.0358)
ORIGIN CONGESTION*PRECIPITATION	***0.3053***	***0.3079***	***0.1995***
	(0.0535)	***(0.0531)***	***(0.0400)***
ORIGIN CONGESTION*VISIBILITY	***-0.1776***	***-0.1742***	***-0.1953***
	(0.0503)	***(0.0500)***	***(0.0659)***
TEMPERATURE*TEMPERATURE	0.1811	0.1863	**0.3257**
	(0.1942)	(0.1928)	**(0.1322)**
TEMPERATURE*WIND	***0.664***	***0.7082***	0.2656
	(0.2267)	***(0.2221)***	(0.1411)
TEMPERATURE*PRECIPITATION	**0.6571**	**0.6066**	0.0005
	(0.3221)	**(0.2990)**	(0.0007)
TEMPERATURE*VISIBILITY	***1.0827***	***1.1123***	***0.9652***
	(0.2053)	***(0.1982)***	***(0.2571)***
WIND*WIND	0.1894	0.2187	0.1519
	(0.1748)	(0.1716)	(0.1235)
WIND*PRECIPATION	-0.3785	**-0.4182**	-0.1456
	(0.1979)	**(0.1941)**	(0.1153)
WIND*VISIBILITY	***0.6259***	***0.6504***	0.1469
	(0.2278)	***(0.2214)***	(0.2704)
PRECIPITATION*PRECIPITATION	***-0.3444***	***-0.3377***	***-0.2315***
	(0.1209)	***(0.1185)***	***(0.0743)***
PRECIPITATION*VISIBILITY	-0.1921	-0.1898	-0.4437
	(0.2092)	(0.2064)	(0.2414)
VISIBILITY*VISIBILITY	***0.4285***	***0.4535***	0.2397
	(0.1607)	***(0.1573)***	(0.2070)

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

Model 3 has lower standard errors on most parameter estimates, reflecting the improved efficiency from using the transformed model.

Many second order terms are significant in Model 3. Of particular interest is the large estimate on the expansion-visibility term, whose negative sign implies that the expansion reduced the impact of low visibility on airport operations. This is consistent with the fact that airport delays are more sensitive to IFR capacity than VFR capacity. Several second order terms involving origin airport congestion are also significant. These results indicate that the impact of origin congestion is non-linear, and is also exacerbated by poor weather conditions, including precipitation and low visibility. Finally two weather interactions are significant in Model 3. High temperature appears to weaken the relationship between visibility and delay, while there is diminishing marginal impact from precipitation. These effects are probably artifacts of the daily level weather data on which the model is based. For example, a relatively low daily average visibility on a higher temperature day may connote haziness throughout the day which does not affect operations, while the same visibility score on a cooler day may indicate periods of IFR conditions.

The strong interaction between expansion and visibility implies that the overall impact of the expansion on DFTI varies from day to day. To further explore that impact, we used the estimation results to predict, for every day after the expansion represented in our data set, the change in DFTI resulting from the expansion. In other words, for each day, we calculated the value:

$$\Delta\hat{Y}_t = \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i X_{it} \qquad (6)$$

where $\Delta\hat{Y}_t$ is the estimated impact of the expansion on DFTI on day t, and the $\hat{a}'s$ are estimates for the parameters in Equation (5). The results presented here are for the preferred Model 3. (Models 1 and 2 yielded very similar results.) Averaging over all post-expansion days, the effect of the expansion was to reduce DFTI 1.3 minutes. Figures 6 and 7 depict the distribution of the estimated DFTI impact. For the vast majority of days the impact is modest—just 1 or 2 minutes in either direction. On roughly 20 percent of the days, the DFTI is estimated to have increased slightly as a result of the expansion. Such increases are probably the result of increased taxi times for flights once they land at DFW, since the added runway is more distant from the terminal area than the pre-existing ones. On about 60 percent of the days, a DFTI reduction greater than one minute is estimated, and for about 10 percent of days, the expansion resulted in a sizable drop in flight times—from five to ten minutes. These high impact days are the ones with low visibility, and generally high predicted values for DFTI. Thus, as shown in Figure 8, there is a strong correlation between predicted DFTI and estimated impact of the expansion. In sum, the primary impact of the expansion was to make a small proportion of bad days considerably better. It follows that the modest reduction in average flight times resulting from the expansion was accompanied by a more noticeable increase in reliability.

4.2 DFTI Components

Estimation results for the three components of DFTI appear in Table 4. The first order impact of the expansion was to reduce departure delay and increase taxi time. The

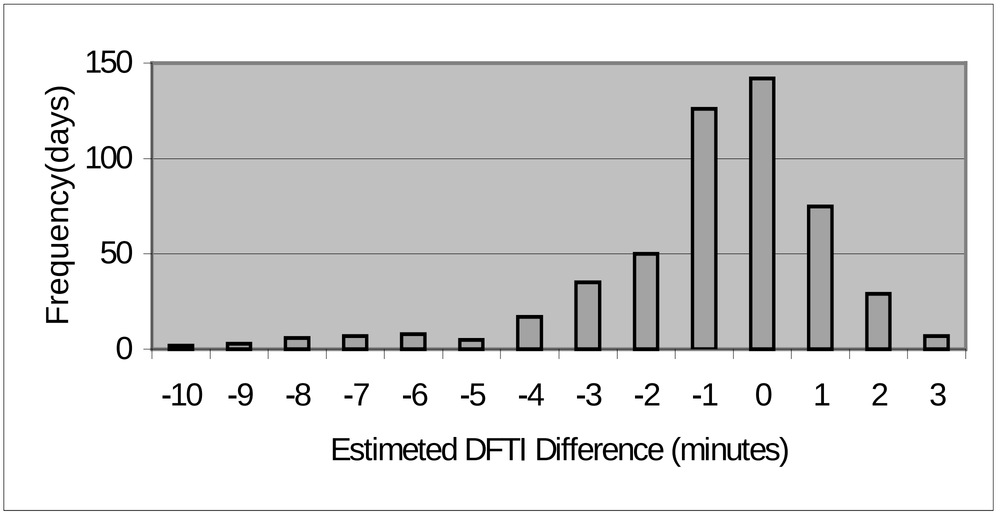

Figure 6: Distribution of Estimated Expansion Impact on DFTI, Post-Expansion Days

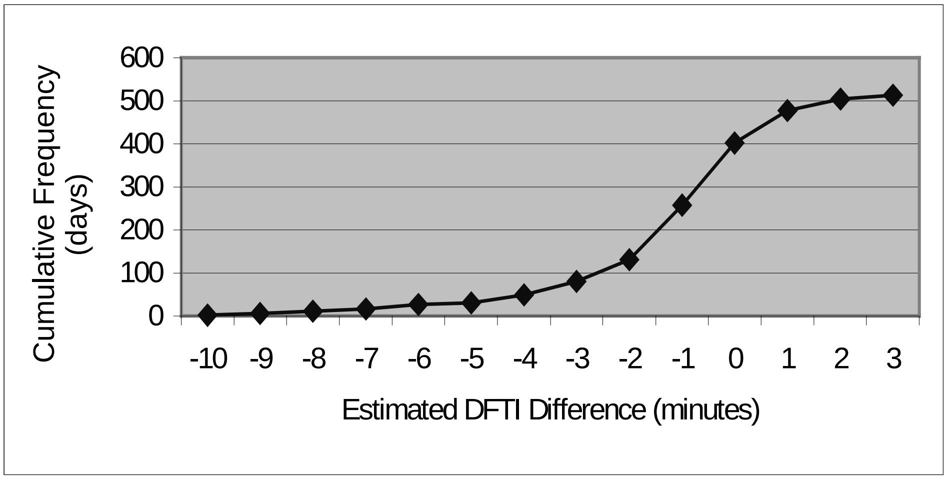

Figure 7: Cumulative Distribution of Estimated Expansion Impact on DFTI, Post-Expansion Days

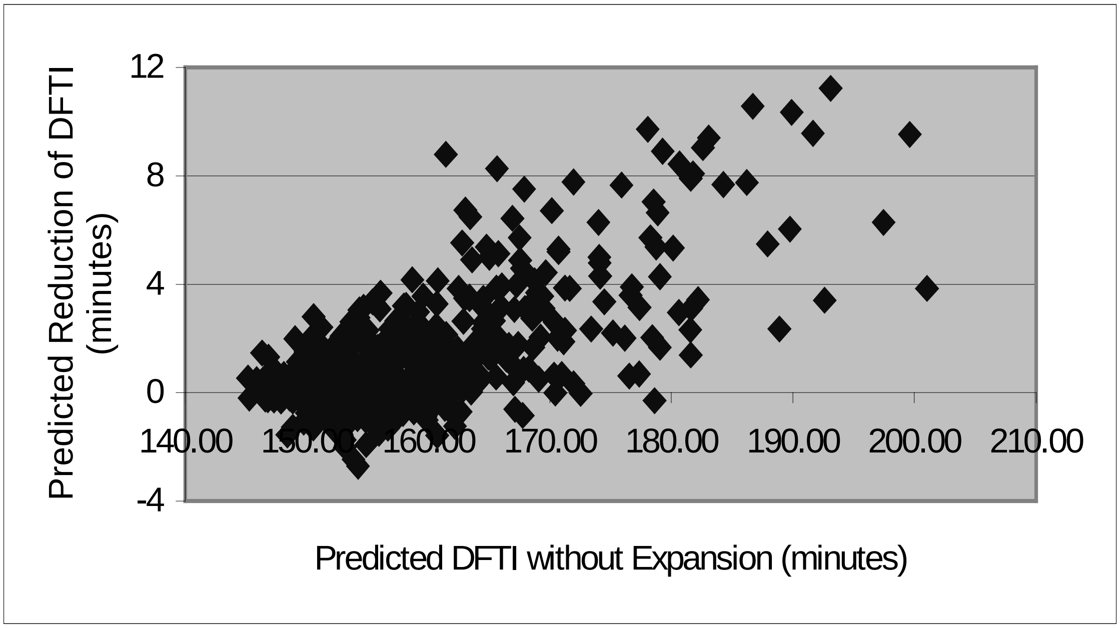

Figure 8: Estimated DFTI Impact of Expansion vs. Predicted DFTI without Expansion, Post-Expansion Days

Table 4: DFTI Component Estimation Results

Variable	Estimate (Standard Error)		
	Departure	Taxi	Airborn
INTERCEPT	0.9287	***20.5135***	***124.8318***
	(0.5742)	***(0.2834)***	***(0.6515)***
EXPANSION	***-1.7519***	***0.6618***	-0.7435
	(0.3915)	***(0.1933)***	(0.4442)
DEMAND	0.1213	***0.5789***	***1.2534***
	(0.2090)	***(0.1032)***	***(0.2372)***
ORIGIN CONGESTION	***0.3835***	***0.2401***	-0.0561
	(0.0682)	***(0.0337)***	(0.0774)
TEMPERATURE	**-0.4630**	0.1935	***-1.7328***
	(0.2139)	(0.1056)	***(0.2427)***
WIND	0.0714	-0.1289	***1.4585***
	(0.2080)	(0.1027)	***(0.2360)***
PRECIPITATION	***0.6581***	0.0427	***1.1833***
	(0.2257)	(0.1114)	***(0.2561)***
VISIBILITY	-0.4478	-0.2353	***-1.9616***
	(0.3518)	(0.1737)	***(0.3992)***
EXPANSION*DEMAND	--	--	--
	--	--	--
EXPANSION*ORIGIN CONGESTION	-0.0039	0.0102	0.0463
	(0.0385)	(0.0190)	(0.0436)
EXPANSION*TEMPERATURE	***0.4601***	-0.0837	***-0.7904***
	(0.1464)	(0.0723)	***(0.1661)***
EXPANSION*WIND	-0.0956	0.1077	-0.2132
	(0.1619)	(0.0799)	(0.1837)
EXPANSION*PRECIPITATION	-0.0662	0.1238	0.3431
	(0.1784)	(0.0880)	(0.2024)
EXPANSION*VISIBILITY	***1.7819***	0.2582	0.2072
	(0.2857)	(0.1410)	(0.3242)
DEMAND*DEMAND	--	--	--
	--	--	--
DEMAND*ORIGIN CONGESTION	--	--	--
	--	--	--
DEMAND*TEMPERATUE	--	--	--
	--	--	--

Variable	Estimate (Standard Error)		
	Departure	Taxi	Airborn
DEMAND*PRECIPITATION	--	--	--
	--	--	--
DEMAND*VISIBILITY	--	--	--
	--	--	--
ORIGIN CONGESTION*ORIGIN CONGESTION	***0.0135***	0.0007	0.0036
	(0.0026)	(0.0013)	(0.0029)
ORIGIN CONGESTION*TEMPERATURE	-0.0034	0.0113	0.0141
	(0.0188)	(0.0093)	(0.0213)
ORIGIN CONGESTION*WIND	-0.0038	-0.0106	-0.0198
	(0.0190)	(0.0094)	(0.0216)
ORIGIN CONGESTION*PRECIPITATION	***0.0772***	***0.0716***	**0.0507**
	(0.0213)	***(0.0105)***	**(0.0241)**
ORIGIN CONGESTION*VISIBILITY	***-0.1471***	-0.0181	-0.0301
	(0.0350)	(0.0173)	(0.0398)
TEMPERATURE*TEMPERATURE	***0.4256***	***0.2738***	***-0.3737***
	(0.0703)	***(0.0347)***	***(0.0797)***
TEMPERATURE*WIND	***0.1962***	0.0380	0.0314
	(0.0750)	(0.0370)	(0.0851)
TEMPERATURE*PRECIPITATION	0.0001	0.0001	0.0003
	(0.0004)	(0.0002)	(0.0004)
TEMPERATURE*VISIBILITY	***0.8939***	***0.3098***	-0.2385
	(0.1366)	***(0.0674)***	(0.1550)
WIND*WIND	0.0687	0.0583	0.0248
	(0.0656)	(0.0324)	(0.0745)
WIND*PRECIPATION	**0.1513**	**-0.0755**	***-0.2215***
	(0.0613)	**(0.0302)**	***(0.0695)***
WIND*VISIBILITY	***0.3870***	0.0927	**-0.3327**
	(0.1437)	(0.0709)	**(0.1630)**
PRECIPITATION*PRECIPITATION	0.0525	***-0.0719***	***-0.2121***
	(0.0395)	***(0.0195)***	***(0.0448)***
PRECIPITATION*VISIBILITY	***-0.5422***	-0.0615	0.1599
	(0.1283)	(0.0633)	(0.1456)
VISIBILITY*VISIBILITY	***0.3332***	0.0872	-0.1807
	(0.1100)	(0.0543)	(0.1248)

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

strong interaction between the expansion and visibility factors is found to involve the departure delay component of the DFTI. Two other, weaker but still statistically significant, interactions between expansion and temperature are also observed. On days with high temperature, the expansion appears to have had a weaker influence on departure delay, but a stronger effect on airborne time. These effects cancel in the overall DFTI model, where the expansion-temperature term is statistically insignificant.

From the component models, we estimated the impact of the expansion on each flight time component on each of the post-expansion days. We again used equation (6), but this time employing the component-specific parameter estimates. Averaging among all of these days, we find that the 1.3 minute reduction in the total DFTI included a 1.9 minute reduction in departure delay, a 0.2 minute reduction in airborne time, and a 0.8 minute increase in taxi time. The departure delay reduction suggests that the major impact of the expansion was to reduce the need for ground delay programs for DFW-bound flights. Since ground delay programs occur primarily when visibility at the destination airport is low, this is consistent with the finding that the expansion had the greatest delay impacts on low visibility days. The increase in taxi time reflects the greater distance of the new runway from the DFW terminal.

These findings contradict the “conventional wisdom” related in several informal conversations that, although the expansion had little effect on delay, it shifted that delay from the air to the ground, where it is less costly. In contrast, we find that the overall shift was from the gate at the origin to the taxiway at the destination. Unlike airborne delay, gate delay is probably less expensive than taxi time. Thus the observed shift implies a smaller economic benefit than an air-to-ground shift. The reduced benefit is a consequence of the FAA ground-hold program, and points to the interplay between benefits from capacity enhancements and air traffic management strategies.

4.3 Sources and Components of DFTI Change After the Expansion

Although the focus of this effort is on assessing the impact of the capacity expansion, it is of interest to consider that impact in the context of other factors that caused the DFTI (and its components) to change after the expansion. We therefore used the estimated models, in conjunction with before-after comparisons of each explanatory variable, to quantify all sources of DFTI change between the pre- and post-expansion periods. Table 5 summarizes the results. The table reveals that the 1.3 minute reduction in total DFTI resulting from the expansion was more than offset by several other factors. First, higher demand during the post expansion period caused a 1.2 minute increase in DFTI, primarily through its effect on the airborne time component. Next, worse weather resulted in a DFTI increase of 0.9 minutes, again primarily affecting the airborne component. Overall, changes not directly related to the expansion caused a 2.5 minute increase in the DFTI. The expansion reduced this increase to 1.2 minutes. Thus, it is simultaneously true that the expansion reduced delay, and that delay was higher after the expansion.

4.4 Induced Demand and Indirect Impacts of the Expansion

In assessing the impacts of the expansion in the previous sections, we have implicitly assumed away indirect impacts arising from any changes it may have caused in

Source/Components	**Impact on Departure Delay (minutes)**	**Impact on Taxi Time (minutes)**	**Impact on Airborne Time (minutes)**	**Total Impact (minutes)**
Demand	0.08	0.37	0.80	1.24
Origin Airport Delay	0.03	0.00	0.01	0.04
Total Weather	0.17	-0.12	0.84	0.89
Origin Airport Delay-Weather Interactions	0.20	0.08	0.06	0.35
Capacity Expansion	-1.88	0.75	-0.21	-1.34
TOTAL CHANGE	-1.40	1.09	1.50	1.19

Table 5: DFTI Change from Before to After Capacity Expansion, by Source and Component

the other explanatory variables. While this assumption is appropriate for most explanators, it is questionable in the case of demand. As seen from Figure 2, the demand metric used in our analysis is, on average, higher in the post-expansion period. As observed in the previous section, the higher post-expansion value caused a 1.2 minute increase in DFTI. Insofar as the increased demand resulted from the capacity expansion, the accompanying increase in DFTI should be counted against the direct impact in order to assess the total delay impact of the capacity expansion.

A rigorous effort to estimate the degree of "induced demand" from the expansion is beyond the scope of this research. Other factors, most notably conflicts between American Airlines and its pilot union during 1996 and 1997, may have been equally or more influential. However, inspection of Figure 2 permits some informed guesses which bound the range of possibilities. One possibility is that the higher demand after the expansion was not causally related to it in any way. We call this "no inducement" scenario. As a "moderate inducement" scenario, we compare the average demand in the after period with the average during the third quarter of 1996, noting that demand was quite constant over this period before jumping upward about two months after the expansion. Finally, for our "high inducement" scenario, we compare the average demand in the period after the expansion with the average over the first half of 1996. This assumes that the jump beginning the third quarter derived from the expectation of the expansion coming "on-line" a short time later.

Table 6 presents the estimated total impact of the capacity expansion on DFTI under each of the three scenarios. In this table, "Direct Impact" refers to the estimate documented in Section 4.1, while the "Indirect Impact" reflects the estimated increase in DFTI arising from demand inducement. In the "moderate inducement" scenario, airlines "traded in" about one third of the direct delay benefits by adjusting their schedules. In the "high inducement" scenario, higher delay stemming from induced demand more than offset the direct delay reduction from the expansion.

While Table 6 offers no conclusive results about the overall impact of the capacity expansion on airline delay, it suggests a large portion of the direct delay reduction benefit may have been offset by changes in airline schedules. This underscores the need for further research on the demand inducement effect of airline capacity expansion. In addition to quantifying the effect, the question of how to assess the benefit realized by airlines and air travelers as a result of their adaptations to new capacity should be considered. Present NAS investment analysis methods, in focusing exclusively on changes in delay, may miss benefit mechanisms of equal and greater importance related to schedule changes.

5. CONCLUSION

Based on metrics developed in our previous work, new metrics reported earlier in this paper, and multivariate statistical techniques, we have developed a statistically valid multiple regression model that quantifies the relationship between average daily flight time into DFW—as measured by the Daily Flight Time Index (DFTI)--and many of its important determinants. We used the model to estimate the delay impact resulting from the 1996 DFW capacity expansion. Our model controls for weather, demand, congestion

Scenario	Direct DFTI Impact (minutes)	Indirect DFTI Impact (minutes)	Total Impact (minutes)
No Inducement	-1.34	0	-1.34
Moderate Inducement	-1.34	0.48	-0.86
High Inducement	-1.34	1.67	0.33

Table 6: Estimated Total Impact of Expansion on DFTI, by Traffic Inducement Scenario

at origin airports, the capacity expansion itself, and interactions between these factors. It can be estimated for the total DFTI, as well as for its various components.

With all other conditions being equal, the capacity expansion reduced delay by approximately 1.3 minutes per flight. On 60 percent of days the delay reduction was greater than one minute, but it exceeds two minutes on only 25 percent. However, for about 10 percent of the days, the expansion resulted in a sizable drop in the flight time index—from five to ten minutes. These are the days with low visibility and high levels of delay. Thus the primary impact of the expansion was to make a small proportion of bad days considerably better. This implies an increase in reliability. On the other hand, it may have diminished the benefit perceived by airlines, since, despite the improvement, "bad" days remain "bad" and there are a relatively small number of them.

We also studied the three components of the DFTI: departure delay at the origin airport, taxi time and airborne time. Our results reveal a sizable reduction of 1.9 minutes in departure delay due to reduced ground holding. Taxi time increased by 0.8 minute on average probably because the new runway is farther from the gate area. The estimated airborne time change is much smaller—0.2 minute.

Changes in delay in the aftermath of the capacity expansion depend on many factors, such as user adaptation and weather. Our regression model is able to isolate the sources of before-after differences in the DFTI, and distinguish differences caused the expansion from those that result from other, potentially confounding, factors. Demand and weather differences between the pre- and post-expansion periods caused increases in DFTI that more than offset the reduction from the capacity expansion.

A critical uncertainty in assessing the impact of the DFW capacity expansion is whether the expansion contributed to the increase in demand—as measured by our hypothetical deterministic delay metric—observed in the post-expansion period. Depending upon what contribution is assumed, the direct reduction in flight times resulting from the expansion may have been wholly or partly counteracted by the induced demand. Resolving this issue—and the related one of the economic value of the demand changes—are critical research needs if the economic benefits from NAS investments are to be correctly understood.

Underlying these factual conclusions is a set of contributions to the state-of-the-art in assessing the impacts of NAS investments on NAS performance. We have demonstrated that comparatively modest impacts can be measured through careful measurement of daily average flight times, controlling for confounding factors such as weather, congestion at origin airports, and demand, using flexible functional forms, and applying appropriate estimation techniques. Through these methods, it is possible to "explain" in a statistical sense most day-to-day variation in flight delay at an airport. The reduced statistical noise enables a "signal" associated with a NAS improvement to be detected and analyzed, even when it is relatively weak. The results of such retrospective studies can and should be used to validate and refine present methods for predicting benefits. By closing this loop, NAS modernization can be a process of learning and adaptation leading to documented improvement, rather than one governed by untested expectations and achieving success by proclamation.

REFERENCES

Davis, T. J. et al., "Operational Test Results of the Passive Final Approach Spacing Tool," Published at the IFAC 8th Symposium on Transportation Systems 97, Chania, Greece, June 1997.

Federal Aviation Administration, Airport Capacity Enhancement Plan, DOT/FAA/CP-87-3, U.S. Department of Transportation, Washington, D.C., 1987.

Geisinger, Kenneth, "Airline Delay: 1976 – 1986," U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Policy and Plans, March, 1989.

Hansen, Mark, H.-S. Jacob Tsao, S-C. Alex Huang, and Wenbin Wei, "Empirical Analysis of Airport Capacity Enhancement Impacts: A Case Study of DFW Airport," NEXTOR Research Report RR-98-17, December 1998.

Johnson, J., Econometric Methods (Third Edition), McGraw-Hill, 1984.

Morrison, Steve, and Clifford Winston, "Enhancing the Performance of the Deregulated Air Transportation System," Brookings Papers: Microeconomics, Brookings Institute, 1989.

Odoni, Amedeo R., "Research Directions for Improving Air Traffic Management Efficiency," Argo Research Corporation, October 1995.

Odoni, Amedeo R. et al., "Existing and Required Modeling Capabilities for Evaluating ATM Systems and Concepts," final report, International Center for Air Transportation, Massachusetts Institute of Technology, March 1997.

Poldy, F., Airport Runway Capacity and Delay: Some Models for Planners and Managers, Bureau of Transportation Economics, 1982.

Uhlenhaker, Ronnie, "DFW Metroplex Air Traffic System Plan," Unpublished Briefing Slides, Federal Aviation Administration, Southwest Region, 1998.