# We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

## 6,900
Open access books available

## 185,000
International  authors and editors

## 200M
Downloads

## 154
Countries delivered to

Our authors are among the

## TOP 1%
most cited scientists

## 12.2%
Contributors from top 500 universities

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

# Predicting Block Time:
# An Application of Quantile Regression

Tony Diana

Additional information is available at the end of the chapter

## 1. Introduction

According to Merriam Webster[1], to predict is 'to declare or indicate in advance; foretell on the basis of observation, experience, or scientific reason.' The advent of sophisticated mathematical and statistical techniques has taken 'divination' out of prediction. In the late 19th century, the work of Francis Galton in the areas of regression analysis, correlation and the normal distribution has been instrumental in helping analysts investigate the relationship between dependent and independent variables and, as a result, to be able to improve forecast. More recently, economists such Robert Engle and Clive Granger have made significant contributions to the study of time series that have widespread applications nowadays in economics and especially finance, such as price and interest rate volatility, as well as risk measurement.

Aviation is another industry that faces risk and uncertainty and has greatly benefited by advances in mathematical, statistical and operations research techniques. A flight is an event that can be scheduled up to six months ahead of its execution. However, despite the best preparation, flight performance is subject to many factors beyond human control such as weather, equipment failure, labor actions, security threats, etc. As a main contributor to the economy and global trade, government regulators, airlines and airport authorities have a vested interest in ensuring that the aviation system supports unimpeded movements of goods and people from their origin to their final destination. According to the *Total Delay Impact Study*[2] by a group of Nextor researchers, "the total cost of all US air transportation delays in 2007 was $32.9 billion. The $8.3 billion airline component consists of increased

---

[1] The source is http://www.merriam-webster.com/dictionary/predict.

[2] Ball, M. et al., 2010. *Total delay impact study, a comprehensive assessment of the costs and impacts of flight delay in the United States*, Nextor, vii. The report is available at the following website: http://its.berkeley.edu/sites/default/files/ NEXTOR_TDI_Report_Final_October_2010.pdf.

expenses for crew, fuel, and maintenance, among others. The $16.7 billion passenger component is based on the passenger time lost due to schedule buffer, delayed flights, flight cancellations, and missed connections. The $3.9 billion cost from lost demand is an estimate of the welfare loss incurred by passengers who avoid air travel as the result of delays".

Predictability is all the more difficult to achieve as airlines often face three types of delay. First, delays can be induced: The air traffic control authority can initiate a ground delay program in case of adverse weather conditions or heavy traffic volume on the ground or en-route. Second, delays can be propagated: In a sequence of legs operated by the same tail-numbered aircraft, a flight may accumulate delays that cannot be recovered by the end of the itinerary. Finally, delays can be stochastic because they are the results of random events such as equipment breakdown or extreme weather events.

Predictability represents a key performance area in the aviation industry for several reasons.

- For the International Civil Aviation Organization (ICAO), predictability refers to the "ability of the airspace users and ATM service providers to provide consistent and dependable levels of performance."[3]
- One of the goals of the U.S. Next Generation of Air Transportation System (NextGen) is to foster the transition from an air traffic control to more of an air traffic managed system where pilots have more flexibility to select their routes, utilize performance-based navigation (PBN) with the help of satellites and make decisions based on automated information-sharing.
- According to Rapajic (2009:51), "cutting five minutes of average of 50 per cent of schedules thanks to higher predictability would be worth some €1,000 million per annum, through savings or better use of airlines and airport resources." Unpredictability imposes considerable costs on airlines in the forms of lost revenues, customer dissatisfaction and potential loss of market share.

Recently, much discussion has revolved around the validity of using airlines' schedules as a measure of on-time performance and the variance of block delay as an indicator of predictability. Both airlines' limited control over the three types of delay and airport congestion make it difficult to build robust schedules and to use schedule as a reference for on-time performance. In fact, schedule padding may skew actual airline performance assessment, hence the need for an alternative methodology.

This article proposes a methodology to determine the predictability of block time based on the case study of the Seattle-Oakland city pair. The predictable block time is located at the percentile where the sign and magnitude of the pseudo coefficient of determination is the highest, while all the covariates are significant at a given confidence level. Ordinary-least-square (OLS) regression models enable analysts to evaluate the percentage of variation in actual block time explained by changes in selected operational variables. However, quantile

---

[3] Henk J. Hof, Development of a Performance Framework in support of the Operational Concept, ICAO Mid Region Global ATM Operational Concept Training Seminar, Cairo, Egypt, November 28–December 1, 2005, p. 36.

regression is more robust to outliers than the traditional OLS regression because the latter does not focus on the conditional mean.

This is of importance to aviation practitioners and, especially, airline schedulers who have often resorted to schedule padding in order to make up for ground and en route delays. This research presents a different perspective on the study of predictability with the intent to help aviation analysts achieve the following objectives:

- To assess the impact of selected operational covariates at different locations of the distribution of block time.
- To derive more predictable block times based on the impact of operational covariates at various quantiles.
- To test a model without any assumption about the distribution of errors and homoscedasticity (constant variance of the residuals).

After a brief background, the discussion will proceed with the methodology, an explanation of the outcomes and some final comments.

## 2. Background

A focus group including communication navigation surveillance and air traffic management representatives[4] defined predictability as "a measure of delay variance against a performance dependability target. As the variance of expected delay increases, it becomes a very serious concern for airlines when developing and operating their schedules".

According to Donohue et al. (2001:398), "predictability focuses on the variation in the ATM [Air Traffic Management] system as experienced by the user. Predictability includes both variability in flight times and arrival rates". In this article, the study of predictability is extended beyond wheels-off (takeoff) and wheels-off (landing) times to include any flight operations between gate-out and gate-in times such as taxi-out and taxi-in movements. This approach takes into account passenger experience.

For Vossen et al. (2011:388), "flexibility can be defined as the amount of operational latitude granted to the carriers in meeting their individual objectives (e.g. on-time arrival, network preservation, profit) when disruptions occur. […] The notion of predictability is closely related, and can be defined as the reduction of uncertainties in the implementation of ATFM [Air Traffic Flow Management] initiatives". Although airlines have to face many events in the course of a flight that cannot thoroughly be anticipated and planned for, "ATFM initiatives should provide the user with time to react, and the provider's intent should be communicated as clearly and as far in advance as possible".

Predictability is sometimes associated with the concept of robust airline scheduling. The latter is the outcome of four sequential tasks as schedule generation, fleet assignment,

---

[4] Report of the Air Traffic Services Performance Focus Group and Communication Navigation Surveillance, February 1999. *Airline Metric Concepts for Evaluating Air Traffic Service Performance.* The website is http://www.boeing.com/ commercial/caft/cwg/ats_perf/ATSP_Feb1_Final.pdf.

aircraft routing and crew pairing/rostering (Wu 2010; Abdelghany and Abdelghany 2009). Fleet assignment models (FAM) are often used to determine how demand for air travel is met by available fleet (see Abara 1989 and Hane et al. 1995). Moreover, the fleet assignment models present two challenges: complexity and size of the problem that the FAM can handle.

Rapajic (2009) identified network structure and fleet composition as sources of flight irregularities. Wu (2010) provided an excellent exposition of issues related to delay management, operating process optimization, and schedule disruption management. Wu explained that "airline schedule planning is deeply rooted in economic principles and market forces, some of which are imposed and constrained by the operating environment of the [airline] industry" (2010:11). He presented a schedule optimization model to improve the robustness of airline scheduling. However, such a model does not consider how selective operational variables are likely to impact scheduling.

Morrisset and Odoni (2011) compared runway system capacity, air traffic delay, scheduling practices, and flight schedule reliability at thirty-four major airports in Europe and the United States from 2007 to 2008. The authors explained that European airports limit air traffic delay through slot control. The other difference is that declared capacity (therefore, the number of available slots) is based mainly on operations under instrument meteorological conditions (IMC). By not placing any restrictions on the number of operations, schedule reliability in the United States depends more on weather conditions than at European airports.

## 3. Methodology

### 3.1. The sample and the assumptions

The sample includes daily data for the month of June to August in 2000, 2004, 2010 and 2011 for the Seattle/Tacoma International (SEA)-Oakland International (OAK) city pair. The summer season is usually characterized by low ceiling and visibility that determine instrument meteorological conditions[5] and weather events such as thunderstorms—all likely to skew the distribution of block times.

Illustration 1 compares the boxplots of actual block times in minutes for the four summers under investigation. The boxplot shows the spread of the distribution, the selected quantile values, the position of the mean and median block times, and the presence of outliers that make it important to consider a regression model at different quantiles. The boxplots reveal an increase in the actual block times between summer 2004 and 2011. Summer 2010 features the largest range as well as the lowest block times at the 5th percentile among the four samples (Illustration 1). It is also characterized by the highest proportion of operations in instrument meteorological conditions compared with the other three samples (Table 1). The

[5] The minimum ceiling and visibility at SEA are respectively 4,000 feet and 3 nautical miles. At OAK, they are 2,500 feet and 8 nautical miles.

skewness coefficients[6] are 0.11, -0.44, 0.37, and 0.19 respectively for summer 2011, 2010, 2004 and 2000. A negative skew indicates that the left tail is longer. While the standard deviation is appropriate to measure the spread of a symmetric distribution, interquartile ranges are more indicative of spread changes in skewed distributions (see Figure 1).
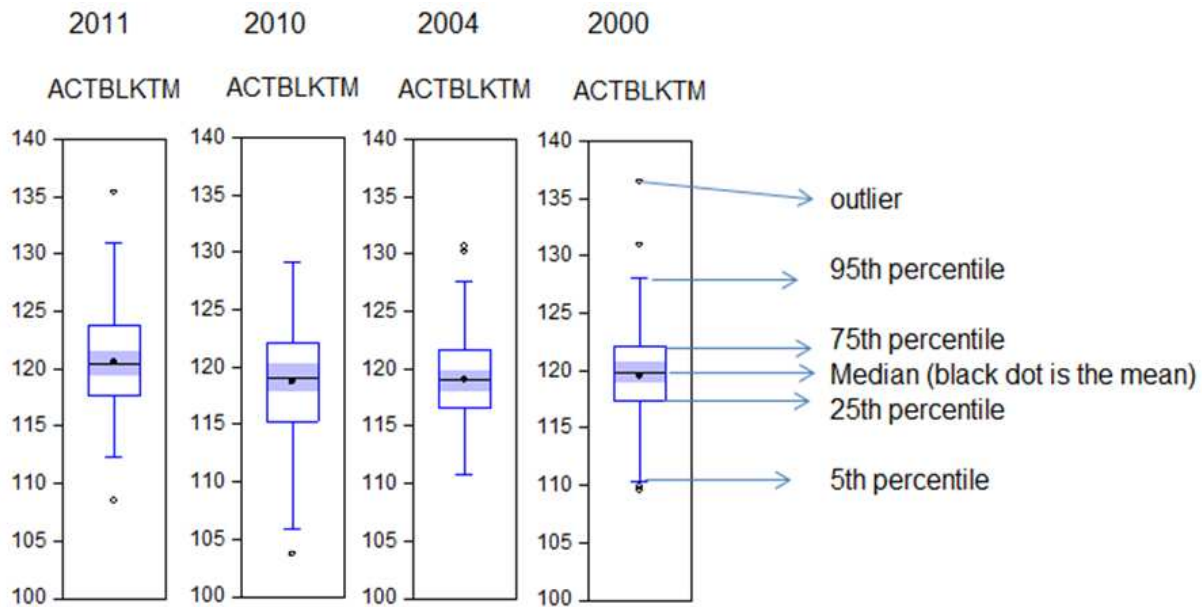


**Figure 1.** Box Plots of Actual Block Times (SEA-OAK)

Secondly, summer is part of the high travel season when demand is usually at its peak. This, in turn, is likely to increase airport congestion and subsequently impact block time. Finally, the years were selected to account for (1) pre- and post-September 11, 2001 traffic, (2) lower traffic demand resulting from the 2008-2009 economic recession, and (3) the introduction of the Green Skies over Seattle[7] after 2010.

The key performance indicators of flight performance are summarized in Table 1. Although the number of flights increased between 2000 and 2011 and the average minutes of expected departure clearance times (EDCT) were higher in 2011 than in 2000, the percentage of on-time gate departures and arrivals and other key delay indicators such as taxi-out delay (a measure of ground congestion) improved in 2011. It is interesting to point out that the percentage of flights in IMC did not change significantly at OAK among the four selected summers. IMC operations were, however, much higher in 2010 and 2011 than in 2000 at SEA, which explains the existence of average minutes of EDCT in 2010 and 2011.

The sample does not include a variable that measures performance-based navigation. The available surveillance data such as Traffic Flow Management System (TFMS) do not capture

---

[6] The skewness coefficient is computed as $\gamma = E[(x - \mu)^3/\sigma] = \mu^3/\sigma^3$ where $\mu^3$ is the third moment about the mean $\mu$ and $\sigma$ is the standard deviation and E is the expectation operator.

[7] The Green Skies over Seattle program includes initiatives such as reduced track mileage to minimum possible distance to protect the environment, optimized profile descent, reduction or elimination of low altitude radar vectoring, as well as required navigational performance.

whether a pilot had requested a required navigation performance procedure, whether air traffic control had granted the request, and whether the procedure had actually been implemented. Moreover, it is presently difficult to differentiate flown performance-based navigation procedures from instrument landing system (ILS) approaches in the case of flight track overlay.

| SEA-OAK | Flight Count | % On-Time Gate Departures | % On-Time Airport Departures | % On-Time Gate Arrivals | Arrivals With EDCT* | Average EDCT Where EDCT>0 (min) | Gate Departure Delay (min) | Taxi Out Delay (min) | Average Taxi Out Time (min) | Airport Departure Delay (min) | Airborne Delay (min) | Taxi In Delay (min) | Block Delay (min) | Gate Arrival Delay (min) | Percent IMC** SEA | Percent IMC** OAK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2000 | 1,411 | 76.26 | 67.54 | 77.18 | 0 | 0 | 9.07 | 4.48 | 14.29 | 12.83 | 4.33 | 0.95 | 2.61 | 9.23 | 10.58 | 29.49 |
| 2004 | 1,473 | 73.52 | 68.30 | 79.16 | 0 | 0 | 11.03 | 3.49 | 12.80 | 13.84 | 6.93 | 1.22 | 1.44 | 9.26 | 8.33 | 29.71 |
| 2010 | 1,147 | 91.63 | 85.27 | 92.85 | 1 | 94 | 4.48 | 3.01 | 13.34 | 6.97 | 5.05 | 0.50 | 1.67 | 3.76 | 30.29 | 29.86 |
| 2011 | 1,196 | 93.73 | 90.64 | 94.90 | 1 | 104 | 3.83 | 2.79 | 13.03 | 5.92 | 3.92 | 0.70 | 1.90 | 3.24 | 22.83 | 29.78 |

\* In the event of a ground delay, airlines are issued an expected departure clearance time (EDCT). Flights held by FAA at the departure airport due to problems at the arrival airport.
EDCT hold delay is computed by comparing EDCT wheels-off time to the flight plan's wheels-off time.
\*\* Instrument Meteorological Conditions
Performance is compared with the last flight plan filed before wheels-off.
Source: ASPM

**Table 1.** Performance Metrics for the SEA-OAK City Pair

Secondly, the availability of Q-routes makes it possible for RNAV/RNP capable aircraft to reduce mileage, to minimize conflicts between routes and to maximize high-altitude airspace. Q-routes are available for use by RNAV/RNP capable aircraft between 18,000 feet MSL (mean sea level) and FL (flight level) 450 inclusive. They help minimize mileage and reduce conflicts between routes.

Thirdly, block time as a measure of gate-to-gate performance is sensitive to delays on the ground and en route. To account for this, airborne delay represents a surrogate for enroute congestion, while increases in taxi times imply surface movement congestion.

## 3.2. Sources and definition of the variables

The sources for the variables are ARINC[8]'s Out-Off-On-In times and the U.S. Federal Aviation Administration's Traffic Flow Management System (TFMS). The directional city pair data originated from the 'Enroute' and 'Individual Flights' data marts of the Aviation System Performance Metrics (ASPM) data warehouse[9].

The choice of variables reflects operational and statistical considerations. On the one hand, some model variables represent significant factors in airport congestion (taxi times) and enroute performance (airborne delays). On the other hand, the model with the highest values for the Akaike Information Criterion (AIC)[10] and Bayesian Information Criterion (BIC)[11] was selected in order to prevent overfitting and to reduce the number of covariates.

The dependent (response variable) and independent variables (covariates) are defined as follows:

---

[8] AIRINC stands for Aeronautical Radio, Inc. (http://www.arinc.com).

[9] The TFMS (formerly ETMS) and ARINC data as well as the ASPM delay metrics are available at http://aspm.faa.gov.

[10] The Akaike Information Criterion is defined as $2k - 2 \ln(L)$ where k is the number of parameters and L the maximized value of the likelihood function for the estimated model.

[11] The Bayesian Information Criterion is $-2 \ln(L) + k.\ln(n)$ where n is the number of observations.

- **Actual Block Time** (ACTBLKTM) is the dependent variable. It refers to the time in minutes from actual gate departure to actual gate arrival.
- **Block Buffer** (BLKBUFFER) represents the difference between planned and optimal block time. The latter is the sum of unimpeded taxi-out times and filed estimated time enroute. Block buffer is the additional minutes included in planned block time in order to take into account potential induced, propagated and stochastic delays. It has also been defined as "the additional time built into the schedule specifically to absorb delay whilst the aircraft is on the ground and to allow recovery between the rotations of aircraft" (Cook, 2007:105). Donohue et al. (2001:113) explained that "to obtain their desired on-time performance, airlines will add padding into a schedule to reflect an amount above average block times to allow for delay and seasonally experienced variations in block times."
- **Departure Delay** (DEPDEL) corresponds to difference between the actual and planned gate departure time at the departure airport in a city pair.
- **Arrival Delay** (ARRDEL) represents to the difference between the actual and planned gate arrival time at the arrival airport in a city pair.
- **Airborne Delay** (AIRBNDEL) accounts for the total minutes of airborne delay. It is the difference between the actual airborne times (landing minus takeoff times) minus the filed estimated time enroute.
- **Taxi-Out Time** (TXOUTTM) refers to the duration in minutes from gate departure to wheels-off times (gate-out to wheels-off).

### 3.3. Quantile regression

Readers interested in quantile regression are referred to Hao and Naiman (2007), Koenker (2005), Koenker and Hallock (2001) and Koenker and Bassett (1998). Quantile regression provides several advantages compared with the ordinary-least-square (OLS) regression in assessing the influence of selected operational factors on the variations of block time at various locations of its distribution:

- Quantile regression specifies the conditional quantile function and, therefore, a way to assess the probability of achieving a certain level of performance. It permits the analysis of the full conditional distributional properties of block time as opposed to ordinary-least-square (OLS) regression models that focus on the mean.
- It defines functional relations between variables for all portions of a probability distribution. Quantile regression can improve the predictive relationship between block times and selected variables by focusing on quantiles instead of the mean. As Hao and Naiman (2007:4) pointed out, "While the linear regression model specifies the changes in the conditional mean of the dependent variable associated with a change in the covariates, the quantile regression model specifies changes in the conditional quantile."
- It determines the effect of explanatory variables on the central or non-central location, scale, and shape of the distribution of block times.
- It is distribution-free, which allows the study of extreme quantiles. Outliers influence the length of the right tail and make average block time irrelevant as a standard for

identifying the best-possible block time. A single rate of change characterized by the slope of the OLS regression line cannot be representative of the relationship between an independent variable or covariate and the entire distribution of block time, the response variable. In the quantile regression, the estimates represent the rates of change conditional on adjusting for the effects of the other model variables at a specified percentile. Therefore, the skewed distribution of block times calls for a more robust regression method that takes into account outliers or the lack of sufficient data at a particular percentile (especially at the extremes of the distribution) and generates different slopes for different quantiles.

The difference between OLS and quantile regression characteristics are summarized in the table below:

| Linear Regression | Quantile Regression |
|---|---|
| • Estimates the mean of a response variable conditional on the values of the explanatory variables (specifies the conditional mean function)<br>• Determines the rate of change in the mean of the response variable | • Specifies the conditional quantile function (focus on quantiles).<br>• Defines functional relations between variables for all portions of a probability distribution |
| • Provides a measure of the impact of explanatory variables on the central location of the distribution of the response variable.<br>• Does not account for full conditional distribution properties of the response variable | • Determines the effect of explanatory variables on the central or non-central location, scale, and shape of the distribution of the response variable<br>• Permits the analysis of the full conditional distributional properties of the dependent variable |
| Normal distribution (sensitive to outliers) | Distribution-free (allows study of extreme quantiles) |
| Determines best fitting line for all data | Different estimates for different quantiles |
| Normal distribution of errors | No assumption about the distribution of errors |
| Assumption of constant variance in errors (homoscedasticity) | Does not assume homoscedasticity |

**Table 2.** The Assumptions of Linear and Quantile Regression

## 4. Outcomes and implications

Appendix 1 provides the estimates for the OLS models. The intercept that represents the predicted value of actual block time when the covariates are equal to zero is not significant

at a 95% confidence level in the 2011 and 2010 samples. However, since the intercept is necessary to provide more accurate predictions, it was left in the model.

Among the dependent variables, gate arrival and departure delays are not significant at a 95% confidence level in the 2010 sample. This implies that airlines can make up for ground delays once en route or ground delays are likely to be more significant in a few extreme cases. The F statistics suggest that there is a zero percent chance that the dependent variable estimates are equal to zero. A value of the Durbin Watson statistic close to 2.00 suggests that there is little statistical evidence that the error terms are positively auto-correlated. The values of the coefficients of determination ($R^2$) imply that the model covariates explain a high proportion of the variation in block times.

In the quantile regression models, the covariate estimates, as well as the key regression statistics at the 5th, 25th, median, 75th and 95th percentile, are summarized in the appendix 2 table. The 50th quantile estimates can be used to track location changes. According to Hao and Naiman (2007: 55), the 5th and 95th percentiles "can be used to assess how a covariate predicts the conditional off-central locations as well as shape shifts of the response." In the case of the 50th percentile in summer 2011, the quantile regression model for at τ (tau) = 0.50 (50th percentile or median) is as follows:

$$\text{Block Time}_{\tau = 0.50} = -14.8091 - 0.8909 * X_{\text{BLKBUFFER}} + 1.0038 * X_{\text{SCHEDBLKTM}} - 0.3329 * X_{\text{DEPDEL}} + 0.2936 * X_{\text{ARRDEL}} + 1.0957 * X_{\text{AIRBNDEL}} + 1.1606 * X_{\text{TXOUTTM}} + \varepsilon \tag{1}$$

In equation (1), 1.1606 represents the change in the median of block time between SEA and OAK corresponding to a one minute change in taxi-out time at SEA. Since the p value is zero, we reject the null hypothesis, at a 95 percent confidence level, that taxi-out times at SEA has no effect on the median block time between SEA and OAK in summer 2011. The pseudo coefficient of determination is a goodness-of-fit measure[12]. In the case of summer 2011, 80.21% of the variation in block time is explained by the model covariates at the 50th percentile of block time (appendix 1).

No sample includes covariates that are significant at a 95% confidence level at all quantiles. Gate departure and arrival delays are significant only at the 95th percentile across the four samples. This means that departure and arrival delays are more likely to affect consistently block times in the upper percentiles—in case of severe airport congestion, for instance. Moreover, the magnitude of block buffers and gate departure delays have a negative impact on the conditional quantile of block time at all samples' selected percentiles. The size of the buffer and the time an aircraft will spend on the tarmac before take-off are conditions likely to affect block times. As a result, there is a need for analysts to decompose and to measure the different operations between gate-out and wheels-off times including gate departure, push-back, taxi-out and queuing times before wheels off. Airport Surface Detection

---

[12] See Koenker and Machado 1999 for further explanations. According to Fitzenberger et al. (2010: 234), "the pseudo $R^2$ equals one minus the sum of weighted deviations about estimated quantile over the sum of weighted deviations around raw quantile".

Equipment, Model X or ASDE-X data should help do so as the system relying on a combination of surface movement radar and transponder multi-lateration sensors becomes more widespread.

Taking the example of summer 2011, 95 percent of the distribution of block times between SEA and OAK was below 129.14 minutes compared with a mean of 120.62 minutes (appendix 2). In other words, there is a 95 percent chance that actual block time will be lower than 129.14 minutes—based on a quantile regression model that explains 85.34 percent of the variation in block times. One benefit of quantile regression is that it facilitates the evaluation of scale and magnitude changes across samples and percentiles.

The quantile regression estimates in appendix 1 imply that block times increased in between summer 2000 and 2011 at all quantiles. In a comparison of summer 2000 with summer 2011, there had been an increase of 2.21 minutes in block times at the 95th percentile, for instance. The SEA-OAK city pair has been mainly operated by Southwest Airlines (SWA) and Alaska Airlines (ASA) with a predominant fleet of Boeing 737s. The total number of ASA arrivals and departures declined to 356 in summer 2011 from 693 in summer 2000—with 91 ASA flights operated by Horizon's Bombardier Q400[13]. Nevertheless, ASA operated larger capacity models such as the dash 400, 800 and 900 series, while SWA utilized a combination of dash 300, 500 and 700 models. The reason for the increase in block time may be attributed to airlines' operations policy to slow aircraft speed in order to save on fuel costs[14]. Weather conditions characterized by the percentage of operations in instrument meteorological conditions (IMC) did not vary substantially at OAK compared with SEA (see Table 1).

In appendix 3, the graphs illustrate the 95% confidence bands in the case of summer 2000. The estimates show a positive relationship between the quantile value and the estimated coefficients for scheduled block times, taxi out times and airborne delay, with a stronger effect in the upper tail. The effect of gate departure and arrival delays is not relatively constant, especially at the 50th percentile as implied by the wider bands around the 50th percentile value. These graphs are important for the analysts in identifying the quantiles where quantile value is likely to be close to the estimated coefficients and, therefore, to improve the accuracy of predicted block time.

## 5. Final comments

Based on the analysis of the SEA-OAK city pair case study, this research showed how quantile regression can help aviation practitioners develop more robust schedules. Originally proposed by Koenker and Bassett (1978), quantile regression is a rather novel approach to the analysis of airlines' on-time performance. Although it is more widely used

---

[13] The sources for schedules and aircraft mix are the Official Airline Guide (http://www.oag.com) and Innovata (http://www.innovata-llc.com).

[14] Associated Press. *Airlines slow down flights to save on fuel: JetBlue adds 2 minutes to each flight, saves $13.6 million a year in jet fuel*, May 1, 2008. The article is available at the following website: http://www.msnbc.msn.com/id/24410809/ ns/business-us_business/t/airlines-slow-down-flights-save-fuel/#.T01rmPES2Ag

in ecology and biology than in the transportation industry, quantile regression is seldom featured in econometric textbooks. Nevertheless, it presents several advantages.

First, it enables aviation analysts to consider the impact of selected covariates on different locations of the distribution of block times. Secondly, the significance and the strength of the impact of selected covariates on block times make it possible to assess the probability that gate-to-gate operations is likely to reach a specific duration. This is made possible by looking at the conditional quantile in the case of quantile regression as opposed to the conditional mean of the distribution of block times in the case of OLS models. Thirdly, quantile regression makes it easier to evaluate the scale and magnitude of change across specific percentiles over a sample. Finally, quantile regression can help analysts study the impact of covariates from different perspectives. For instance, in summer 2011, the data analysis suggests that 95% of the block time distribution will be below the quantile dependent variable value of 129.14 minutes as a result of the impact of the covariates' impact. Quantile regression enables the identification of more realistic threshold times based on quantiles and it allows airline practitioners to simulate and to evaluate various scenarios linked to changes in the models' covariates.

Predictability is a key performance area identified by the International Civil Aviation Organization. Moreover, it is a corner stone of the Next Generation of Air Transport System (NextGen) initiatives in the U.S. and the Single European Sky ATM Research program (SESAR) to ensure the transition from an air traffic controlled to a more air traffic managed environment. As air transportation regulators are under public pressure to crack down on tarmac and other types of delays, it has become imperative for airline schedulers to evaluate models that reflect the predictable influence of key operational variables on actual on-time performance. The complexity of the air traffic system, the inability for airline schedulers to fully anticipate both airport and en route congestion, and delays all make it more significant for aviation practitioners to assess the impact of some key operational variables at different locations of the distribution of block times that usually tends to be skewed due to outliers.

The imbalance between air travel demand and airport capacity usually results in delays. As block times become more predictable, it is more possible for airline and airport operators to optimize airport capacity— especially at large congested airports. This is all the more significant in the U.S. where arrival and departure flows are not slot- constrained as in Europe. Block time predictability does not only affect how airports and airlines operate, but also the capability of air traffic control authorities to anticipate staff workload, as well as the ability of ground handlers to minimize aircraft turn times by allocating resources where and when needed.

## Author details

Tony Diana[15]
*Division Manager, NextGen Performance, Federal Aviation Administration, Office of NextGen Performance and Outreach, ANG-F1, SW, Washington DC, USA*

---

[15] Note: This article does not represent the opinion of the Federal Aviation Administration.
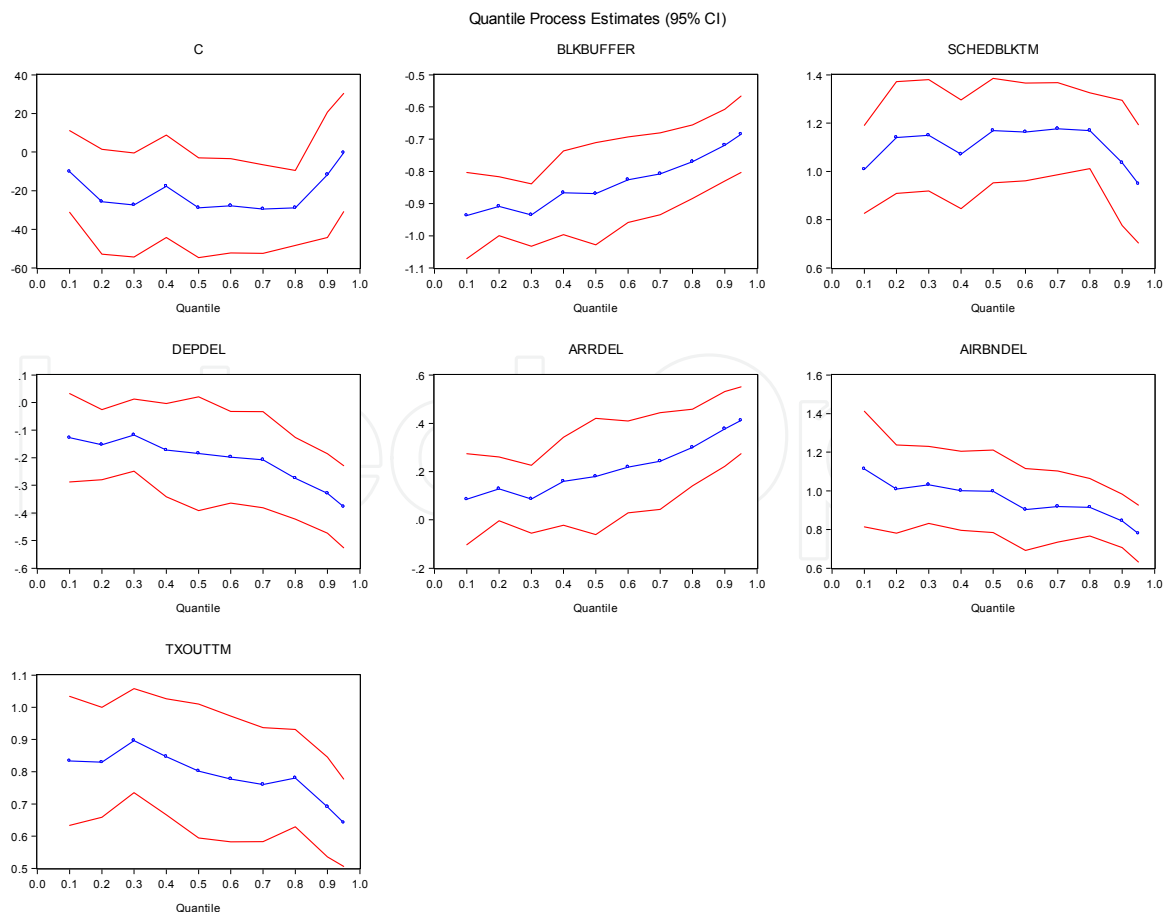
# Appendix

## The ordinary-least square regression outputs

| Alpha = .95 | 2011 | | 2010 | | 2004 | | 2000 | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Probability | Coefficient | Probability | Coefficient | Probability | Coefficient | Probability |
| ACTBLKTM (dep var) | | | | | | | | |
| INTERCEPT | -3.3081 | 0.7037 | -6.0440 | 0.3358 | -14.3526 | 0.0154 | -19.6283 | 0.0249 |
| BLKBUFFER | -0.8383 | 0.0000 | -0.9862 | 0.0000 | -0.9051 | 0.0000 | -0.8087 | 0.0000 |
| SCHEDBLKTM | 0.9291 | 0.0000 | 0.9552 | 0.0000 | 1.0531 | 0.0000 | 1.0940 | 0.0000 |
| DEPDEL | -0.4482 | 0.0000 | -0.0878 | 0.1046 | -0.1263 | 0.0159 | -0.2544 | 0.0000 |
| ARRDEL | 0.4319 | 0.0002 | 0.0726 | 0.2134 | 0.1399 | 0.0144 | 0.2554 | 0.0000 |
| AIRBNDEL | 0.9596 | 0.0000 | 1.0619 | 0.0000 | 0.8741 | 0.0000 | 0.9249 | 0.0000 |
| TXOUTTM | 0.9831 | 0.0000 | 1.0326 | 0.0000 | 0.8850 | 0.0000 | 0.7643 | 0.0000 |
| | | | | | | | | |
| R-squared | 0.9472 | | 0.9879 | | 0.9819 | | 0.9853 | |
| Adjusted R-squared | 0.9435 | | 0.9871 | | 0.9806 | | 0.9843 | |
| S.E. of regression | 1.1578 | | 0.5894 | | 0.5681 | | 0.6043 | |
| Sum squared resid | 113.9466 | | 29.5253 | | 27.4320 | | 31.0360 | |
| Log likelihood | -140.3837 | | -78.2615 | | -74.8788 | | -80.5568 | |
| F-statistic | 254.2718 | | 1158.3370 | | 767.2133 | | 951.7874 | |
| Prob(F-statistic) | 0.0000 | | 0.0000 | | 0.0000 | | 0.0000 | |
| Mean dep var | 120.6251 | | 118.6857 | | 119.1140 | | 119.4768 | |
| Durbin Watson | 1.9670 | | 1.8550 | | 1.9048 | | 2.0154 | |

Not significant at α = .95

## Summer 2000: Quantile process estimates (95% confidence level)



Quantile Process Estimates (95% CI)

## The quantile regression outputs

| Alpha = .95 | 2011 | | 2010 | | 2004 | | 2000 | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Probability | Coefficient | Probability | Coefficient | Probability | Coefficient | Probability |
| **5th Percentile** | | | | | | | | |
| INTERCEPT | 12.5331 | 0.2650 | 4.2506 | 0.7100 | -4.6586 | 0.3253 | -10.3640 | 0.2707 |
| BLKBUFFER | -0.7264 | 0.0000 | -1.0038 | 0.0000 | -0.9720 | 0.0000 | -0.9798 | 0.0000 |
| SCHEDBLKTM | 0.8110 | 0.0000 | 0.8646 | 0.0000 | 0.9664 | 0.0000 | 1.0065 | 0.0000 |
| DEPDEL | -0.7312 | 0.0000 | -0.0540 | 0.2859 | -0.0339 | 0.6064 | -0.0931 | 0.2397 |
| ARRDEL | 0.8353 | 0.0000 | 0.1076 | 0.0530 | 0.0493 | 0.5027 | 0.0357 | 0.7014 |
| AIRBNDEL | 0.8453 | 0.0000 | 1.0346 | 0.0000 | 0.9057 | 0.0000 | 1.1799 | 0.0000 |
| TXOUTTM | 0.6581 | 0.0074 | 1.0355 | 0.0000 | 0.9542 | 0.0000 | 0.8833 | 0.0000 |
| | | | | | | | | |
| Pseudo R-squared | **0.7082** | | **0.8896** | | **0.8900** | | **0.8791** | |
| Adjusted R-squared | 0.6876 | | 0.8818 | | 0.8822 | | 0.8706 | |
| S.E. of regression | 2.7301 | | 1.1423 | | 0.9829 | | 1.1722 | |
| Quantile dependent var | 112.8600 | | 107.5600 | | 113.0000 | | 110.9400 | |
| Mean dependent var | 120.6251 | | 118.6857 | | 119.114 | | 119.4768 | |
| **25th Percentile** | | | | | | | | |
| INTERCEPT | 0.7081 | 0.9534 | -10.3697 | 0.2900 | -7.3948 | 0.1315 | -22.7559 | 0.1164 |
| BLKBUFFER | -0.8557 | 0.0000 | -1.0066 | 0.0000 | -0.9266 | 0.0000 | -0.9058 | 0.0000 |
| SCHEDBLKTM | 0.8868 | 0.0000 | 0.9861 | 0.0000 | 0.9915 | 0.0000 | 1.1157 | 0.0000 |
| DEPDEL | -0.4349 | 0.0003 | -0.0351 | 0.5664 | -0.1217 | 0.1107 | -0.1446 | 0.0394 |
| ARRDEL | 0.3957 | 0.0004 | 0.0367 | 0.6397 | 0.1257 | 0.1233 | 0.1260 | 0.0862 |
| AIRBNDEL | 1.1110 | 0.0000 | 1.0671 | 0.0000 | 0.9263 | 0.0000 | 1.0099 | 0.0000 |
| TXOUTTM | 1.0340 | 0.0000 | 1.0522 | 0.0000 | 0.9112 | 0.0000 | 0.8284 | 0.0000 |
| | | | | | | | | |
| Pseudo R-squared | **0.7694** | | **0.8972** | | **0.8799** | | **0.8878** | |
| Adjusted R-squared | 0.7532 | | 0.8899 | | 0.8714 | | 0.8799 | |
| S.E. of regression | 1.2129 | | 0.7105 | | 0.6992 | | 0.7683 | |
| Quantile dependent var | 117.7100 | | 115.0000 | | 116.5300 | | 117.2100 | |
| Mean dependent var | 120.6251 | | 118.6857 | | 119.114 | | 119.4768 | |
| **50th Percentile** | | | | | | | | |
| INTERCEPT | -14.8091 | 0.5037 | -12.1850 | 0.1575 | -12.3493 | 0.0406 | -28.8452 | 0.0316 |
| BLKBUFFER | -0.8909 | 0.0000 | -0.9754 | 0.0000 | -0.9289 | 0.0000 | -0.8694 | 0.0000 |
| SCHEDBLKTM | 1.0038 | 0.0000 | 0.9990 | 0.0000 | 1.0427 | 0.0000 | 1.1690 | 0.0000 |
| DEPDEL | -0.3329 | 0.0366 | -0.0560 | 0.3936 | -0.1077 | 0.1146 | -0.1849 | 0.0828 |
| ARRDEL | 0.2936 | 0.0333 | 0.0207 | 0.7928 | 0.1070 | 0.1486 | 0.1794 | 0.1486 |
| AIRBNDEL | 1.0957 | 0.0000 | 1.0409 | 0.0000 | 0.8574 | 0.0000 | 0.9972 | 0.0000 |
| TXOUTTM | 1.1606 | 0.0000 | 1.0893 | 0.0000 | 0.8686 | 0.0000 | 0.8021 | 0.0000 |
| | | | | | | | | |
| Pseudo R-squared | **0.8021** | | **0.8946** | | **0.8674** | | **0.8736** | |
| Adjusted R-squared | 0.7881 | | 0.8871 | | 0.8580 | | 0.8647 | |
| S.E. of regression | 1.1988 | | 0.6096 | | 0.5835 | | 0.6142 | |
| Quantile dependent var | 120.4300 | | 119.0000 | | 118.9300 | | 119.8600 | |
| Mean dependent var | 120.6251 | | 118.6857 | | 119.114 | | 119.4768 | |
| **75th Percentile** | | | | | | | | |
| INTERCEPT | -31.1458 | 0.0001 | -2.2100 | 0.7869 | -20.9044 | 0.0037 | -25.7729 | 0.0169 |
| BLKBUFFER | -0.8502 | 0.0000 | -0.9943 | 0.0000 | -0.8452 | 0.0000 | -0.7926 | 0.0000 |
| SCHEDBLKTM | 1.1481 | 0.0000 | 0.9310 | 0.0000 | 1.1098 | 0.0000 | 1.1441 | 0.0000 |
| DEPDEL | -0.3207 | 0.0512 | -0.0846 | 0.3845 | -0.1746 | 0.1296 | -0.2313 | 0.0122 |
| ARRDEL | 0.3281 | 0.0642 | 0.0536 | 0.6036 | 0.1805 | 0.1081 | 0.2584 | 0.0127 |
| AIRBNDEL | 0.9996 | 0.0000 | 0.9994 | 0.0000 | 0.8917 | 0.0000 | 0.9390 | 0.0000 |
| TXOUTTM | 1.0485 | 0.0000 | 1.0311 | 0.0000 | 0.8113 | 0.0000 | 0.7669 | 0.0000 |
| | | | | | | | | |
| Pseudo R-squared | **0.8269** | | **0.8958** | | **0.8639** | | **0.8781** | |
| Adjusted R-squared | 0.8147 | | 0.8884 | | 0.8543 | | 0.8695 | |
| S.E. of regression | 1.4266 | | 0.7151 | | 0.7214 | | 0.7646 | |
| Quantile dependent var | 123.8000 | | 121.9200 | | 121.6000 | | 122.1500 | |
| Mean dependent var | 120.6251 | | 118.6857 | | 119.114 | | 119.4768 | |
| **95th Percentile** | | | | | | | | |
| INTERCEPT | -17.6714 | 0.2819 | -17.4339 | 0.0149 | -31.0322 | 0.0000 | -0.2942 | 0.9850 |
| BLKBUFFER | -0.9532 | 0.0000 | -0.9099 | 0.0000 | -0.7827 | 0.0000 | -0.6847 | 0.0000 |
| SCHEDBLKTM | 1.0483 | 0.0000 | 1.0580 | 0.0000 | 1.1988 | 0.0000 | 0.9482 | 0.0000 |
| DEPDEL | -0.2748 | 0.0332 | -0.2929 | 0.0002 | -0.3229 | 0.0003 | -0.3772 | 0.0000 |
| ARRDEL | 0.3277 | 0.0174 | 0.2663 | 0.0018 | 0.3269 | 0.0006 | 0.4122 | 0.0000 |
| AIRBNDEL | 0.7805 | 0.0000 | 1.0384 | 0.0000 | 0.7939 | 0.0000 | 0.7791 | 0.0000 |
| TXOUTTM | 1.1530 | 0.0000 | 0.9544 | 0.0000 | 0.7826 | 0.0000 | 0.6418 | 0.0000 |
| | | | | | | | | |
| Pseudo R-squared | **0.8534** | | **0.8997** | | **0.8859** | | **0.9103** | |
| Adjusted R-squared | 0.8431 | | 0.8926 | | 0.8778 | | 0.9040 | |
| S.E. of regression | 1.8510 | | 1.0918 | | 1.1257 | | 1.1419 | |
| Quantile dependent var | 129.1400 | | 127.4300 | | 126.0600 | | 126.9300 | |
| Mean dependent var | 120.6251 | | 118.6857 | | 119.114 | | 119.4768 | |

Not significant at α = .95

## 6. References

Abara, J., 1989. Applying integer linear programming to the fleet assignment problem. Interfaces 19(4), 20-28.

Abdelghany, F. and Abdelghany, K., 2009. Modeling applications in the airline industry. Ashgate Publishing Company: Burlington, Vermont.

Cook, A., 2007. European air traffic management: principles, practices, and research. Ashgate Publishing Company: Burlington, Vermont.

Donohue, G.L., Zellweger, A., Rediess, H., and Pusch, C., 2001. Air transportation system engineering: progress in astronautics and aeronautics. American Institute of Aeronautics and Astronautics: Danvers, Massachusetts.

Fitzenberger, B., Koenker, R., and Machado, J., 2010. Economic applications of quantile regression (Studies in empirical economics). Physica Verlag: Heidelberg.

Hane, C.A., Barnhart, C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L., and Sigismondi, G., 1995. The fleet assignment problem: solving a large scale integer programming. Mathematical Programming, 70(2), 211-232.

Hao, L. and Naiman, D. Q., 2007. Quantile Regression. Sage Publications: Thousand Oaks, CA.

Koenker, R., 2005. Quantile Regression. Cambridge, UK: Cambridge University Press.

Koenker, R. and Hallock K., 2001. Quantile Regression. Journal of Economic Perspectives, 15(4), 143-156.

Koenker, R. and Machado, J., 1999. Goodness of fit and related inference processes for quantile regression, Journal of the American Statistical Association, 94(448), 1296-1310.

Morissett, T. and Odoni, A., 2011. Capacity, delay, and schedule reliability at major airports in Europe and the United States. Transportation Research Record: Journal of the Transportation Research Board, 2214, 85-93.

Rapajic, J., 2009. Beyond airline disruptions. Ashgate Publishing Company: Burlington, Vermont.

Vossen, T., Hoffman, R., and Mukherjee, A., 2011. Air traffic flow management in Quantitative problem solving methods in the airline industry: a modeling methodology handbook, Barnhart, C. and Smith, B., editors. Springer: New York.

Wu, C.-L., 2010. Airline operations and delay management: insights from airline economics, networks and strategic schedule planning. Ashgate Publishing Company: Burlington, Vermont.