MODELING PROBABILITY DISTRIBUTIONS OF PRIMARY DELAYS IN THE NATIONAL AIR TRANSPORTATION SYSTEM

Shitong Wang[12] and Vikrant Vaze[2*]

[1]Tianjin Chengjian University
School of Economics and Management, Tianjin Chengjian University
300384, Tianjin, China
Tel: +86-139-2010-3370; Email: wangshitong@tcu.edu.cn

[2]Thayer School of Engineering, Dartmouth College
14 Engineering Drive, Hanover, NH 03755
Tel: 603-646-9147; Fax: 603-646-3856; Email: {shitong.wang, vikrant.s.vaze}@dartmouth.edu

*Corresponding author

Word count: 5,745 words text (excluding references) + 5 tables x 250 words (each) = 6,995 words + 34 references

TRR Paper number: 16-2921

Submitted on November 13th, 2015

Submitted for presentation in the 95th Annual Meeting of the Transportation Research Board and publication in Transportation Research Record: Journal of the Transportation Research Board.

ABSTRACT
Flight delays impose enormous costs and are further exacerbated by propagation effects. Development and evaluation of advanced schedule planning approaches for mitigating propagated delays requires primary delay distribution models with a long prediction horizon of several weeks-to-months. To address this need, this paper presents a hybrid modeling approach combining logistic regression and quantile regression. Using two large datasets, each consisting of approximately six million flights, as the training and testing samples, various types of factors affecting primary delay distributions are identified. Several new insights are obtained into the factors affecting delay distributions. Coefficient estimates are found to have a high statistical significance and intuitive interpretation of their signs and relative magnitudes. Results from model estimation, validation and testing show that fit values are stable across datasets, thus avoiding overfitting. The relative explanatory power of various types of predictive factors – including distance factors, seasonality factors, time-of-day factors, airport factors and airline factors – is also quantified. A new method is provided for quantifying the fit of the joint model in the form of the log-likelihood of the observed data, and the model is shown to provide a considerable improvement in log-likelihood.

Keywords: Flight delays, Primary delays, Delay propagation, Quantile Regression, Logistic Regression, Probability Distribution

1. INTRODUCTION

In recent years, flight delays have become prevalent at many major airports in the world. Flight delays and disruptions is a serious and widespread problem in the U.S. According to the U.S. Department of Transportation (DOT), a flight is considered as delayed if it arrives at the destination gate 15 minutes or more after its scheduled arrival time (*1*). Over the last decade (that is, over the last 10 full calendar years, 2005-2014), out of approximately 60 million flight operations by major U.S. carriers, about 20.35% flights arrived 15 minutes or more after their planned arrival times, 1.78% flights were cancelled and 0.23% flights were diverted, resulting in only 77.64% flights arriving within 15 minutes of their scheduled arrival times.

Flight delays have a significant impact on the U.S. economy. The Total Delay Impact Study commissioned by the Federal Aviation Administration (*2*), estimated the total cost of flight delays to the U.S. economy at \$31.2 billion in 2007. This includes \$8.3 billion in direct costs to the airlines and \$16.7 billion in passenger delay costs. Another study also focusing on year 2007, the U.S. Congress Joint Economic Committee report (*3*), estimated that flight delays resulted in consumption of 740 million additional gallons of jet fuel, costing an additional \$1.6 billion in fuel costs, and releasing an additional 7.1 million metric tons of carbon dioxide into the atmosphere. In addition to the direct costs of delays, airlines also suffer from loss of goodwill and possibly a reduced demand for future travel from passengers who suffered large delays on a flight of that airline.

There are many reasons for flight delays. An important reason is the insufficient airport and airspace capacity, which fails to meet the requirements of the increasing air traffic demand. Other causes of flight delays include extreme weather conditions, mechanical problems, luggage problems, airline operational issues, late-arriving aircraft and crews from previous flights, etc. Bureau of Transportation Statistics (BTS) website classifies delays based on several predefined causes. Over the last 10 full calendar years (2005-2014): 'late-arriving aircraft delay' contributed 34.7%, 'National Aviation System (NAS) delay' contributed 33.3%, 'air carrier delay' (as a result of maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.) contributed 28.2%, 'extreme weather delay' (due to tornadoes, blizzards, hurricanes, etc.) contributed 3.5% and 'security delay' contributed 0.3% of all delays (*1*).

Among these delay causes, some can be considered secondary in nature, in the sense that, once a delay occurs in one part of the aviation system, these lead to additional delays in other parts of the network. This phenomenon is called *delay propagation*. For example, late-arriving aircraft delays result from delay propagation through aircraft connections between flights. Also, some fraction of delays categorized as air carrier delays results from delay propagation through crew connections between flights. Note that these secondary delays result from original (or *primary*) delays that are caused by other reasons listed above. According to the BTS website for the 2005-2014 period, for more than 44% of the late-arriving aircraft delays, the primary delays were categorized as NAS delays (*1*). Additionally, around half of the primary delays themselves were categorized as NAS delays. NAS delays are those that are caused by non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control. Simply put, NAS delays are the delays due to congestion in the air transportation system.

Among the wide variety of approaches proposed to mitigate congestion and delays, a prominent category strives to minimize delay propagation through aircraft, crew and passenger connections by constructing schedules that either do not get disrupted often (called robust scheduling approaches) or are easy to get back on plan once a delay or disruption occurs (called recoverable scheduling approaches) (*4*). Effectiveness of these propagated delay minimization approaches depends on the probability distributions of primary delays. Note that these scheduling decisions need to be made several weeks-to-months ahead of the day-of-operations when the actual weather conditions on the day-of-operations are unknown. Therefore, these approaches can benefit from good models of primary delay distributions that do not rely on the knowledge of actual weather conditions on the day-of-operations. Such models of primary delay distributions are beneficial not only to develop approaches for mitigating delay propagation but also to evaluate and compare multiple candidate approaches with each other.

Most of the existing delay models assume knowledge of weather forecasts to some extent. Additionally, while many studies aim to model delay averages, models of probability distributions of delays are needed. Finally, none of the existing studies explicitly model either overall primary delays or congestion-related primary delays, which are the two major focuses of this paper. This study models probability distributions of primary delays using only the predictive factors that are known several week-to-months in advance, while especially focusing on delays due to congestion. Using two large datasets each consisting of approximately six million flights as the training and testing samples, various types of important factors affecting primary delay distributions are identified. Because of the format of the publicly available data, a hybrid modeling approach is used combining logistic and quantile regression. Coefficient estimates are found to have a high statistical significance and are intuitive in terms of their signs and relative magnitudes. Results from model estimation, validation and testing show that model fit values are stable across datasets, thus avoiding overfitting. The relative explanatory power of various types of predictive factors – including distance factors, seasonality factors (both weekly and annual), time-of-day factors, airport factors and airline factors – is also quantified. Finally, a new method is described for quantifying the fit of the joint (logistic and quantile regression) model in the form of the log-likelihood of observed data, and the results show that the joint model provides a considerable improvement in log-likelihoods in all cases.

The remainder of the paper is organized as follows. Section 2 summarizes prior literature on flight delay models. Section 3 describes the modeling method that combines logistic regression and quantile regression techniques. Section 4 describes the data used to build the models. Section 5 summarizes results and Section 6 concludes.

2. LITERATURE REVIEW

Several prior studies have focused on the general task of flight delay modeling and prediction. They can be classified based on multiple criteria including 1) modeling method, 2) delay representation, 3) delay types considered, and 4) prediction time horizon.

Similar to the modeling method-based classification provided by (*5*), studies that model flight delays can be divided into five categories: *a) regression/econometric methods, b) Bayesian network methods, c) clustering and classification methods, d) simulation methods* and *e) Data Envelopment Analysis (DEA) methods*. The first category includes studies that estimate the relationship between delay as the dependent variable and one or more explanatory (independent) variables. Specific methods employed by these studies span a wide range including regression (*6-8*), censored regression (*9*), simultaneous equation model (*10*), nested Logit (*11*), Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) model (*12-13*), Multivariate Adaptive Regression Splines (MARS) models (*5;14*), etc. Some past studies (*15-17*) have used Bayesian Networks (BNs) to model the probabilistic relationships among different components of flight delays and various factors affecting delays. Bayesian network methods are especially suitable for modeling the relationships between the probability distributions of different delay components including departure delay, taxi-out delay, en-route delay, taxi-in delay, arrival delay, etc. Another category of past studies has used clustering and classification methods for flight delay prediction (*18-22*). The general idea is to predict delays for a particular day by using historical data to identify historical days when the state of the air transportation system was similar to that day. In Bayesian network methods as well as clustering and classification methods, the state of the system is often described based on weather forecasts, flight volumes, and the planned and executed traffic management initiatives (TMIs) on a particular day. A wide variety of simulation tools have been developed and used for modeling and predicting flight delays (*23-26*). Simulation methods can model flight delay distributions, but suffer from long run-times (often requiring multiple runs) and need a lot of calibration and validation before implementation. While the original classification scheme used by (*5*) does not include the last category (DEA), a number of more recent studies have used DEA methods for the estimation of production frontiers wherein one or more of the *undesirable outputs* of the production process are characterized in terms of delay measures such as delayed flights percentage, average flight delay (*27-28*), etc.

In terms of delay representation, almost all aforementioned studies represent delay in the form of average flight delay or number of delayed flights. Only exceptions are those studies which use either simulation or Bayesian network methods (*15-17;23-26*). Some other studies (e.g., (*29*)) have tried to fit certain predefined probability distributions, e.g., Gaussian, Poisson, etc., to flight delays. However, none of these studies have tackled the problem of modeling delay probability distributions as functions of explanatory variables. This is the problem addressed in this paper.

In terms of the delay type considered, most of the previous studies fall into one of the two categories: they deal either with overall delay per flight (*10;12-13;27-28*) or with specific components of flight delays, such as, departure delay, taxi-out delay, en-route delay, taxi-in delay, arrival delay, etc. (*16;19-20*). Some of the studies (e.g., (*8;15;17*)) focus specifically on propagated delays. However, none of the studies (with the exception of (*21*)) focus on either primary delays or NAS delays. This paper focuses on modeling the primary delays without differentiating between different delay components, beyond the exclusion of propagated delays.

In terms of the prediction horizon being considered, a majority of prior studies focuses on the day-of-operations with the aim of improving the response of airlines and/or air traffic managers to delays and disruptions. Only a few of these studies, including DEA and production frontier approaches (*7;27-28*) and some regression and econometric approaches (*6;8;11*) model delays without the benefit of the specific weather forecasts and detailed data on flight volumes and TMIs. This study falls in this category because over the long-term schedule planning horizon, accurate weather forecasts, detailed flight volume data and actual or forecasted TMI information is not available.

In summary, this study models the probability distributions (rather than average values) of primary delays (rather than propagated delays) to aid long-term scheduling decisions (rather than benefiting short-term decision-making) using a combination of logistic regression and quantile regression. This is the first study that models parameterized probability distributions of non-propagated delays using public data. Also, despite the flexibility afforded by quantile regression in modeling probability distributions, the authors are not aware of any study that uses quantile regression for modeling flight delays of any kind. More details of the modeling approach are provided in Section 4.

3. DATA AND VARIABLES

The objective of this study is to model the probability distributions of primary delays several weeks-to-months ahead of the actual day-of-operations. Therefore, the analysis is based on data from Airline On-Time Performance (AOTP) database from BTS which is the largest public source of flight delay data. It also provides delay causes. It contains planned and actual schedules for flights operated by all U.S.-certified air carriers that account for at least 1% of domestic scheduled passenger revenues. Data is included on all AOTP-reported flights over the two year time horizon 2013-2014. The sample size is 6,369,482 for 2013 and 5,819,811 for 2014. The sample is limited to non-cancelled and non-diverted flights to ensure the availability of delay causes. Observations with missing data for either actual arrival time, or actual departure time, or both, were dropped from the analysis, yielding a sample of 6,271,207 flights for 2013 and 5,690,183 flights for 2014.

AOTP database provides delays categorized as 1) NAS delays, 2) carrier delays, 3) late-arriving aircraft delays, 4) extreme weather delays, and 5) security delays. Late-arriving aircraft delays are not the focus of this study because they are not primary delays. Also, extreme weather and security delays are excluded because together they account for less than 4% of all delays. NAS delays are the primary delays due to congestion, and carrier delays are a combination of primary delays due to carrier operations and delays propagated through crew connections. Modeling only the NAS delays ignores nearly half of the primary delays, while modeling the sum of NAS and carrier (NAS+Carr) delays leads to inclusion of some propagated delays (those which propagate through crew connections). Therefore two different models are developed, one for NAS delays, and another for NAS+Carr delays.

Based on prior literature, and on data availability several weeks-to-months ahead of the day-of-operations, independent variables are categorized into five types: 1) distance factors, 2) weekly and annual seasonality factors, 2) time-of-day factors, 4) airport factors, and 5) airline factors. The first type includes only the distance (transformed as natural logarithm) between the flight's origin and destination airports as a single continuous variable. Seasonality factors include six dummies for day-of-week, Monday through Saturday, with Sunday as reference; and three dummies for season-of-year, namely, Spring (March-May), Summer (June-August), and Fall (September-November), with Winter (December-February) as reference. Time-of-day factors include three departure time dummies and three arrival time dummies. Scheduled local departure time at the origin airport and scheduled local arrival time at the destination airport, is divided into Night (midnight-6am), Morning (6am-noon), and Afternoon (noon-6pm), with Evening (6pm-midnight) as reference.

Airports are divided into four types depending on congestion level as measured over the last decade before the modeling period. The first three types cover the 35 busiest airports in the U.S., often called the OEP35 (Operational Evolution Plan 35) airports. Type 1 airports include the five most congested OEP35 airports, namely, Newark Liberty International (EWR), John F. Kennedy International (JFK), LaGuardia (LGA), Chicago O'Hare (ORD), and San Francisco International (SFO). Type 2 airports include 13 moderately congested OEP35 airports, namely, Hartsfield-Jackson Atlanta International (ATL), Logan International (BOS), Cleveland Hopkins International (CLE), Cincinnati/Northern Kentucky International (CVG), Reagan National (DCA), Denver International (DEN), Dallas/Fort Worth International (DFW), Fort Lauderdale-Hollywood International (FLL), Washington Dulles International (IAD), Houston George Bush Intercontinental (IAH), Chicago Midway International (MDW), Philadelphia International (PHL), and Pittsburgh International (PIT). Type 3 airports include the remaining 17 OEP35 airports and Type 4 airports serve as reference and include all non-OEP35 airports.

Similar to (*30*), airlines are divided into four types: 1) legacy carriers: American Airlines (AA), Delta Air Lines (DL), United Air Lines (UA), and US Airways (US); 2) low-cost carriers: JetBlue Airways (B6), Frontier Airlines (F9), AirTran Airways (FL), Virgin America (VX), and

Southwest Airlines (WN); 3) regional carriers: Endeavor Air (9E), ExpressJet Airlines (EV), Envoy Air (MQ), SkyWest Airlines (OO), and Mesa Airlines (YV); and 4) non-continental carriers: Hawaiian Airlines (HA), and Alaska Airlines (AS). These are modeled using three dummies with non-continental airlines as reference.

Summary statistics of the independent variables for the cleaned data (as described earlier) are shown in Table 1. The 'Variable' column also lists the abbreviated variables names (inside square brackets). These abbreviated names are used in Tables 2, 3 and 4.

4. MODEL DESCRIPTION

In the AOTP database, a flight is reported as *delayed* when it arrives 15 or more minutes late compared with its scheduled arrival time. AOTP reports NAS delay and carrier delay as zero for any flight that is not *delayed*. In the year 2013, over 78% flights arrived on-time (i.e., not *delayed*) and hence had zeros for reported NAS delays and carrier delays. Note that, even among the delayed flights, the NAS delay and/or carrier delay can equal zero if the delay cause is different than these two. As a result, in 2013 data approximately 89% flights had zero NAS delay and 83% had zero NAS+Carr delay. Therefore, the delay probability distributions in the dataset have two distinct parts: a tall peak at zero and a smooth decreasing curve for positive values. Due to this peculiar nature of the distribution, modeling it directly using quantile regression method leads to low statistical significance and severe numerical issues in the estimation process. In order to match the models closely with the structure of the available data and in order to get statistically significant and practically meaningful results, delay distribution model is developed in two phases. In the first phase, using logistic regression, a binary dependent variable is modeled such that it equals 1 when the relevant (NAS/NAS+Carr) delay is positive and 0 otherwise. This model is estimated using all observations in the training dataset. The second phase uses quantile regression to model the Cumulative Distribution Function (CDF) of the relevant delay conditioned on the relevant delay being positive. This model is estimated using only the observations with positive values of the relevant delay.

In logistic regression, the probability of $D_j > 0$, where D_j is the relevant delay in the j^{th} observation, is modeled as CDF ($\pi(x_j)$) of the logistic distribution, given in Equation (1). Here $x_j \epsilon R^n$ is the independent variable vector for the j^{th} observation, $x_{ij} \epsilon R$ is value of the i^{th} independent variable in the j^{th} observation, and $\beta \epsilon R^n$ is the coefficient vector.

$$P\left(D_j > 0 \middle| x_j\right) = \pi(x_j) = \frac{e^{g(x_j)}}{1+e^{g(x_j)}} \quad (1)$$

$$g(x_j) = \ln\left[\frac{\pi(x_j)}{1-\pi(x_j)}\right] = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_n x_{nj} = \beta_0 + \sum_{i=1}^{n} \beta_i x_{ij} \quad (2)$$

As mentioned earlier, independent variables are classified into five types. First type includes a single variable which equals the natural logarithm of the distance between flight's origin and destination airports. Natural logarithm was found to provide better results than using distance directly or using any other transformation. Second type includes six day-of-week dummies and three season-of-year dummies, third includes six time-of-day dummies, fourth includes six airport type dummies, and finally, fifth includes three airline type dummies. The definitions of these types and of individual variables within these types were chosen by trial-and-error to find the best model specification.

Quantile regression (QR) is an extension of the classical ordinary least squares (OLS) regression. While OLS regression is used for modeling the mean, quantile regression is a collection of models for different quantile functions. It was first introduced by Koenker and Bassett, and further developed by Koenker and Hallock (*31-32*). The QR approach is more robust than the OLS approach in the presence of outliers and when the distribution of the dependent variable is different from the Gaussian distribution. Both these properties hold for flight delays data. Quantile regression estimates a range of different quantiles of the dependent variable, and thus provides a more complete characterization of the full effect of the independent variable (*32-33*).

Consider a random variable Y whose CDF is given by $F_Y(y) = P(Y < y)$. Then the τ^{th} quantile of Y is defined as the inverse of the CDF function, i.e., as $Q(\tau) = inf[y: F(y) > \tau]$ where $0 < \tau < 1$. The sample quantile, $\xi(\tau)$, analogous to the population quantile, $Q(\tau)$, is obtained as the solution of the optimization problem,

$$\min_{\xi \in R} \sum_j \rho_\tau(y_j - \xi) \tag{3}$$

where $\rho_\tau(z) = z * (\tau - I(z < 0))$ for $0 < \tau < 1$. $I(\cdot)$ is the indicator function. When the scalar ξ is replaced by a parametric function $\xi(x_j, \gamma) = x_j{}'\gamma$ such that $x_j \epsilon R^n$ is the independent variable vector for the j^{th} observation and $\gamma \epsilon R^n$ is the coefficient vector, an estimate of the linear function of n independent variables approximating the τ^{th} quantile, can be obtained by solving

$$\hat{\gamma}(\tau) = arg \min_{\gamma \epsilon R^n} \sum_j \rho_\tau(y_j - x_j'\gamma) \tag{4}$$

The vector $\hat{\gamma}(\tau)$ contains the τ^{th} quantile regression coefficient estimates.

Quantile regression is used to model the CDF of the relevant (NAS/NAS+Carr) delay variable conditioned on the relevant delay being positive. So the CDF under consideration (denoted earlier by $F_Y(y) = P(Y < y)$) is the conditional CDF $F_{D_j}(y | D_j > 0) = P(D_j < y | D_j > 0)$. As with logistic regression, the set of independent variables is divided into five types, namely, distance, seasonality, time-of-day, airport, and airline factors.

5. RESULTS

Logistic and quantile regression estimation results are described in Sub-sections 5.1 and 5.2 respectively. Sub-section 5.3 describes the results of 10-fold cross validation for both models and out-of-sample testing of both models. Sub-section 5.3 also presents an approach to quantify the overall fit provided by the joint logistic and quantile regression model, and presents the joint model fit results. In general, because the models are estimated to be used over a long-term time horizon of several weeks-to-months before the day-of-operations when only a limited amount of information is available about the actual conditions on the day-of-operations, the models are expected to have a relatively low goodness-of-fit. While the results are consistent with this expectation, the statistical significance of the models is very high. Additionally, the goodness-of-fit for the upper quantile models is found to be higher. This is especially noteworthy given that typically the larger delays lead to much higher costs than smaller delays and it is therefore more critical to model them accurately.

5.1 Logistic Regression Results

Table 2 summarizes logistic regression estimation results obtained by using 2013 data for model training. First four columns provide the coefficient estimates and their statistical significance. Plugging these coefficient estimates into Equations 1 and 2 gives the positive delay probability estimate for any given flight. A positive (negative) value of any coefficient estimate indicates its increasing (decreasing) effect on the delay probability. Except for the Afternoon Departure Time dummy in the NAS+Carr model, the remaining 51 coefficient estimates are statistically significant at 1% level (in fact, 50 of them are significant at 0.00000001% level). The McFadden R^2 is 2.9% for the NAS model and 3.2% for the NAS+Carr model, while the Cox-Snell R^2 value is 2.0% for the NAS model and 2.9% for the NAS+Carr model.

All else equal, probability of having positive primary delays increases with flight distance. Monday, Thursday and Friday are the days with highest delay probability which is consistent with the fact that, as shown in Table 1, these are the three days with the highest number of flights and hence the highest congestion in 2013. Saturday has the lowest delay probability consistent with its lowest flight volume. Summer and Winter are found to have higher delay probability than Fall and Spring (similar to (*30*)). All else equal, afternoon departures and afternoon and evening arrivals (in other words, flights later in the day) have higher delay probability. Probability of NAS+Carr delays peaks later in the day than that for NAS delays possibly because of the inclusion of crew-based delay propagation within NAS+Carr delays. Consistent with the airport type definitions, Type 1 airports have the highest delay probability, followed by Type 2 airports. The lowest delay probability is for airport Types 3 and 4.

The effects of distance, seasonality, time-of-day and airport factors are similar in NAS and NAS+Carr models. However, the effects of airline factors are different because only NAS+Carr delays include carrier delays. All else equal (and specifically since the airport effects are controlled for) NAS delay probability is the lowest for legacy carriers followed closely by low-cost carriers, and is highest for the non-continental carriers. NAS+Carr delay probability, however, is highest for low-cost carriers, followed by regional carriers, and is the lowest for the non-continental carriers.

Two different methods are developed to quantify the relative explanatory power of different types of factors.

- Additive method: First, for each type of factors (e.g., time-of-day, seasonality, etc.), R^2 value is calculated using only the intercept and the variables of that type. This process is repeated for each of the five types. Then the five R^2 values are normalized to sum up to 100%.
- Subtractive method: First, for each type of factors, R^2 value is calculated using intercept and all the variables of types *other than* that type. Then this is subtracted from the overall model R^2 to get a measure of that type's contribution. This process is repeated for each of the five types. Finally the five contributions are normalized to sum up to 100%.

The last eight columns of Table 2 show the fit contributions for each type of factors. Note that the results are very similar for McFadden and Cox-Snell R^2 values and also for the additive and subtractive methods. As expected, distance being a single variable, contributes very little to the overall fit with contributions under 4%. Seasonality factors, consisting of nine dummies, contribute between 20% and 23% across all cases. Airline factors, consisting of three dummies, contribute between 9% and 14% across all cases, with larger contributions to the NAS+Carr model than the NAS model in all cases as expected. Time-of-day factors, consisting of six dummies, have the largest contribution (42%-44%) to the NAS+Carr model but somewhat lower (29%-30%) contribution to the NAS model, while airport factors, also consisting of six dummies, have the largest contribution (32%-36%) to the NAS model but somewhat lower (20%-23%) to the NAS+Carr model.

5.2 Quantile Regression Results

Tables 3 and 4 summarize quantile regression results for NAS delay model and NAS+Carr delay model respectively. They are obtained using that subset of 2013 data for which the respective delays are positive. In either table, the first 10 columns summarize the coefficient estimates (the underlined ones are statistically not significant at 1% level) while the last four summarize the fit contributions. An inner product of this coefficient estimate vector with the independent variable vector gives the estimated conditional quantiles for any flight. These, together with the positive delay probability estimates obtained as described in Sub-section 5.1, provide the estimated CDFs of primary delays. A positive (negative) value of any coefficient estimate indicates its increasing (decreasing) effect on the corresponding quantile. Though QR models are estimated for all quantiles from 5% to 95% in steps of 5%, due to space limitations only those in steps of 10% are reported. The unreported quantile (e.g., 10%, 20%, etc.) results are similar to those reported. Also, fit contribution results for only the 85% and 95% quantiles are reported due to a) space limitations, and b) higher quantiles being more important from a delay-cost perspective. Of the 520 coefficient estimates in these two tables, only 42 are not significant at 1% level, out of which 26 are for day-of-week dummies. The overall pseudo R^2 (*34*) fit value varies between 0.5% and 12.7% for NAS model and between 0.9% and 4.2% for NAS+Carr model with a general trend of increasing fit with increasing quantiles.

All quantiles (except 95% quantile of NAS delay) increase with increasing flight distance. Seasonality factors contribute less than 10% to the overall fit in all but one of the reported cases. Day-of-week effects are weak for all quantiles below 85% with magnitudes less than a minute for NAS and less than two minutes for NAS+Carr model. The 85% and 95% quantiles are the lowest for Tuesday and Saturday, and they are the highest for Sunday indicating that even though Sunday has fewer positive delays (as reflected by logistic regression results), when the delays do occur, upper delay quantiles on Sundays are higher than those on other days. Also, most delay quantiles for Summer and Spring are higher than those for Winter and Fall. Time-of-day factors contribute little to the NAS model fit (<6%), with the highest delay quantiles obtained for morning and afternoon departures and evening arrivals. However, the fit contribution from time-of-day factors to the NAS+Carr model is higher (9%-20%) with the highest delay quantile value obtained for night departures and morning arrivals.

Airport factors account for a remarkably large fraction of the NAS model fit (78%-93%) indicating that variations in NAS delays are explained to a great extent by airport factors. Consistent with logistic regression results, NAS delay quantiles are the highest for Type 1 airports, followed by Type 2 airports and lowest for Type 4 airports, in almost all cases. Airline factors contribute less than 5% to the NAS model fit. In terms of the NAS+Carr model, a large fraction of fit contribution comes from airport (39%-47%) and airline factors (33%-39%), which is intuitive. For both NAS and NAS+Carr models, regional carriers have higher delay quantiles than legacy and low-cost carriers in most cases.

5.3 Ten-fold Cross Validation and Model Testing

First 12 rows and four columns of Table 5 summarize the ten-fold cross validation results. These are obtained by dividing 2013 data randomly into 10 equal parts (folds). Each time nine of those 10 folds are used for model training and the remaining fold is used for validation. This process is repeated 10 times and average R^2 values thus obtained for training and validation datasets are reported. For NAS and NAS+Carr delays, for logistic and quantile regression, validation fit values are similar to the corresponding training values, thus eliminating any concerns of model overfitting.

First 12 rows and last four columns of Table 5 summarize the effectiveness of the models, estimated using all of 2013 data, over a test dataset consisting of all of 2014 data. As observed from these values, the prediction accuracy (as reflected by the R^2 fit values) for the 2014 testing dataset is very similar to that for the 2013 training dataset, thus highlighting the predictive power of each model separately. However, the true predictive ability of the joint model cannot be fully quantified through these separate fit values for the individual components. Therefore, a new method is developed to assess the joint model's predictive ability using the concept of likelihood. The likelihood of a sample of size m is the joint probability of all its data points ($d_j, j \in \{1, \dots, m\}$).

$$P(D_j = d_j) = \begin{cases} P(D_j = 0) = \frac{1}{1+e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_{ij}}} \dots\dots\dots\dots\dots & \text{if } d_j = 0 \\ P(D_j > 0).P(D_j = d_j | D_j > 0) = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_{ij}}}{1+e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_{ij}}}.P(D_j = d_j | D_j > 0) \dots & o.w. \end{cases} \quad (5)$$

where x_{ij} is the value of i^{th} independent variable in j^{th} observation and β_i is the logistic regression coefficient of i^{th} independent variable. Quantile regression model gives ξ values such that $\tau = P(D_j < \xi | D_j > 0)$ for $\tau = 5\%, 10\%, 15\%, \dots, 95\%$. Obviously, $P(D_j < 0 | D_j > 0) = 0$. Consider an observation j with dependent variable value $d_j > 0$. Also, let $\xi(\tau_k) < d_j \leq \xi(\tau_{k+1})$ for some pair $(\xi(\tau_k), \xi(\tau_{k+1}))$ of consecutive estimated quantiles. All reported delays in the AOTP database are integer minutes. Let l be the number of integers between $\xi(\tau_k)$ and $\xi(\tau_{k+1})$. If the likelihood of each of those l integers is assumed to be equal, then $P(D_j = d_j | D_j > 0) = \frac{\tau_{k+1} - \tau_k}{l}$. In addition to $\tau = 0\%, 5\%, 10\%, 15\%, \dots, 95\%$, a model for $\tau = 99\%$ is also estimated. For $d_j > \xi(99\%)$ (which is true for less than 0.25% observations), it is assumed that $P(D_j = d_j | D_j > 0) = P(D_j = d_j' | D_j > 0)\ s.t.\ d_j' \in (\xi(95\%), \xi(99\%)) \cap \mathbb{Z}$.

Thus, $P(D_j = d_j | D_j > 0)$ values are calculated and substituted in Equation (5) to get $P(D_j = d_j)$ for the j^{th} observation. Overall likelihood of the sample is given by $\prod_{j=1}^{m} P(D_j = d_j)$. Third-last row of Table 5 reports the natural logarithm of the overall sample likelihood (called the sample log-likelihood). Same calculation is repeated for a naïve model and the log-likelihood is reported in second-last row of Table 5. Naïve model contains no independent variables except for an intercept term in both logistic and quantile regressions. The last row provides the differences between the log-likelihoods of the full and naïve models. Log-likelihood of training and testing samples is significantly higher for the full model than for the naïve model suggesting that the explanatory variables significantly improve likelihood of the observed data. In all cases, the observed data is more likely under the full model than under the naïve model by a multiplier ranging between $exp(239{,}955)$ and $exp(282{,}923)$. Thus the full model improves the log-likelihood by 3.5% to 4.9% compared with the naïve model.

6. CONCLUSIONS

This study modeled the probability distributions of primary delays using public data by employing a model combining logistic and quantile regression methods. This is the first study to model parameterized probability distributions of non-propagated delays. Also, this is the first study to use the quantile regression method for modeling flight delays. Two types of primary delays are modeled, namely, NAS delays and the sum of NAS and carrier (NAS+Carr) delays. Several of the results are as expected and intuitively understandable. When controlled for other factors, primary delays are found to be higher for longer (compared with shorter) distance flights, higher during summer months of June through August (compared with rest of the year), higher during later (compared with earlier) times of the day, lower during Tuesdays and Saturdays (compared with rest of the week) and especially higher at EWR, JFK, LGA, ORD and SFO (compared with the other) airports.

Additionally, several new insights are obtained for the first time due to the explicit modeling of delay probability distributions rather than delay averages. For example, when controlled for other factors, though the positive delay probability is found to be similar for low-cost and legacy carriers, the longer delays are considerably less probable for the low-cost carriers compared with the legacy carriers. Also, while seasonality and time-of-day factors explain more than half of the variation in the positive delay probabilities, they explain very little of the variation in higher delay quantiles. Airport factors, on the other hand, explain probabilities of longer delays far better than they explain the positive delay probabilities. Airline factors' contribution to explaining upper quantiles of NAS delays is negligible, but they explain the upper quantiles of NAS+Carr delays significantly better. Moreover, airline factors explain a lot larger proportion of the variability in upper NAS+Carr delay quantiles than that in positive delay probabilities. None of these insights could have been obtained without explicit parameterized modeling of delay quantiles.

These models are motivated by the need to accurately assess advanced schedule planning approaches for reducing propagated delays. Therefore as next steps in this research, these models need to be evaluated in that context. Integration of these primary delay distributions into airline operations simulation environments will enable airlines to develop and implement effective strategies to reduce delay propagation, in turn leading to improved airline profits, reduced passenger inconvenience, more efficient airport operations, and reduction in adverse environmental effects. Also, while this study focused on long-term time horizon of several weeks-to-months, it would also be interesting to extend these models to the day-of-operations context to enable assessment of the effectiveness of delay recovery methods.

REFERENCES

1. Bureau of Transportation Statistics. *Airline On-Time Statistics and Delay Causes*. U.S. Department of Transportation, Washington, D.C. www.rita.dot.gov/bts/help/aviation/html/understanding.html. Accessed Mar. 25, 2015.
2. Ball, M., C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, B. Zou, R. Britto, D. Fearing, P. Swaroop, N. Umang, V. Vaze, and A. Voltes. *Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States*. The National Center of Excellence for Aviation Operations Research (NEXTOR), 2010.
3. Schumer, C. E., and C. B. Maloney. *Your Flight Has Been Delayed Again: Flight Delays Cost Passengers, Airlines, and the U.S. Economy Billions*. The U.S. Senate Joint Economic Committee, 2008.
4. Barnhart, C., and V. Vaze. Irregular Operations: Schedule Recovery and Robustness. In *The Global Airline Industry, 2nd Edition* (P. Belobaba, A. Odoni, and C. Barnhart, ed.), John Wiley & Sons, Inc., New York, N.Y., 2015, pp. 263-286.
5. Xu, N. *Method for Deriving Multi-Factor Models for Predicting Airport Delays*. PhD Dissertation. George Mason University, Fairfax, VA, 2007.
6. Mayer, C., and T. Sinai. Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why All Delays Are Not Evil. *American Economic Review,* Vol. 93, No. 4, 2003, pp. 1194-1215.
7. Ramdas, K., and J. Williams. *An Empirical Investigation into the Tradeoffs that Impact On-Time Performance in the Airline Industry*. Working Paper. Graduate School of Business, Stanford University, Stanford, CA, 2008.
8. Vigneau, W. *Flight Delay Propagation, Synthesis of the Study*. EEC Note No 18/03. EUROCONTROL Agency M3 Systems, Brussels, Belgium, 2003.
9. Hansen, M., and D. Peterman. Throughput Effect of Time-Based Metering at Los Angeles International Airport. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1888,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 59-65.
10. Hansen, M., and Y. Zhang. Operational Consequences of Alternative Airport Demand Management Policies - Case of LaGuardia Airport, New York. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1915*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 95-104.
11. Rupp, N. G. *Flight Delays and Cancellations*. Working Paper ECU0502. Department of Economics, East Carolina University, Greenville, NC, 2005.
12. Hansen, M., and C. Y. Hsiao. Going south?: Econometric Analysis of U.S. Airline Flight Delays from 2000 to 2004. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1915*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 85–94.
13. Hsiao, C. Y., and M. Hansen. Econometric Analysis of U. S. Airline Flight Delays with Time-of Day Effects. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1951*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 104-112.
14. Xu, N., L. Sherry, and K. B. Laskey. Multifactor Model for Predicting Delays at U.S. Airports. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2052*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp.

62–71.
15. Xu, N., G. Donohue, K. B. Laskey, and C. H. Chen. Estimation of Delay Propagation in the National Aviation System using Bayesian Networks. Presented at 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD, 2005.
16. Xu, N., K. B. Laskey, C. H. Chen, S. C. Williams, and L. Sherry. Bayesian Network Analysis of Flight Delays. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
17. Laskey, K. B., N. Xu, and C. H. Chen. Propagation of Delays in the National Airspace System. Appears in Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, 2006.
18. Callaham, M. B., J. S. DeArmon, A. M. Cooper, J. H. Goodfriend, D. Moch-Mooney, and G. H. Solomos. Assessing NAS Performance: Normalizing for the Effects of Weather. Presented at 4th USA/Europe Air Traffic Management R&D Symposium, Santa Fe, NM, 2001.
19. Hoffman, B., J. Krozel, S. Penny, A. Roy and K. Roth. A Cluster Analysis to Classify Days in the National System. Presented at AIAA Guidance, Navigation, and Control Conference, Austin, TX, 2003.
20. Penny, S., B. Hoffman, J. Krozel, and A. Roy. Classification of Days in the National Airspace System using Cluster Analysis. *Air Traffic Control Quarterly,* Vol. 13, No. 1, 2005, pp. 29-54.
21. Chatterji, G. B., and B. Sridhar. National Airspace System Delay Estimation using Weather Weighted Traffic Counts. Presented at AIAA Guidance, Navigation, and Control Conference, San Francisco, CA, 2005.
22. Grabbe, S., B. Sridhar, and A. Mukherjee. Clustering Days with Similar Airport Weather Conditions. Presented at AIAA Aviation Technology, Integration, and Operations Conference, Atlanta, GA, 2014.
23. Odoni, A., J. Bowman, D. Delahaye, J. J. Deyst, E. Feron, R. J. Hansman, K. Khan, J. K. Kuchar, N. Pujet, and R. W. Simpson. *Existing and Required Modeling Capabilities for Evaluating ATM Systems and Concepts*. International Center for Air Transportation, Massachusetts Institute of Technology, 1997.
24. Bilimoria, K., B. Sridhar, G. Chatterji, K. Sheth, and S. Grabbe. FACET: Future ATM Concepts Evaluation Tool. Presented at 3rd USA/Europe Air Traffic Management R&D Symposium, Napoli, Italy, 2000.
25. Long, D., and S. Hasan. Improved Prediction of Flight Delays using the LMINET2 System-Wide Simulation Model. Presented at AIAA Aviation Technology, Integration, and Operations Conference, Hilton Head, South Carolina, 2009.
26. Pyrgiotis, N., K. M. Malone, and A. Odoni. Modelling Delay Propagation within an Airport Network. *Transportation Research Part C: Emerging Technologies,* Vol. 27, 2013, pp. 60-75.
27. Lozano, S., and E. Gutiérrez. Slacks-Based Measure of Efficiency of Airports with Airplanes Delays as Undesirable Outputs. *Computers & Operations Research,* Vol. 38, No. 1, 2011, pp. 131-139.
28. Lozano, S., E. Gutiérrez, and P. Moreno. Network DEA Approach to Airports Performance Assessment Considering Undesirable Outputs. *Applied Mathematical Modelling,* Vol. 37, No. 4, 2013, pp. 1665-1676.
29. Mueller, E. R., and G. B. Chatterji. Analysis of Aircraft Arrival and Departure Delay Characteristics. Presented at AIAA Aircraft Technology, Integration and Operations Conference,

Los Angeles, CA, 2002.
30. Barnhart, C., D. Fearing, and V. Vaze. Modeling Passenger Travel and Delays in the National Air Transportation System. *Operations Research*, Vol. 62, No. 3, 2014, pp. 580-601.
31. Koenker, R., and G. J. Bassett. Regression Quantiles. *Econometrica,* Vol. 46, No. 1, 1978, pp. 33-50.
32. Koenker, R. and K. F. Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, Vol. 15, No. 4, 2001, pp. 43-56.
33. Mueller, J. M., and J. B. Loomis. Does the Estimated Impact of Wildfires Vary with the Housing Price Distribution? A Quantile Regression Approach. *Land Use Policy,* Vol. 41, 2014, pp. 121-127.
34. Koenker, R. and J. A. F. Machado. Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association,* Vol. 94, No. 448, 1999, pp. 1296-1310.

LIST OF TABLE TITLES AND FIGURE CAPTIONS

TABLE 1 Data Description and Summary Statistics

Type	Sub-Type	Variable	2013	2014
Distance	Distance	ln (Distance) [ln (Dist)]	na	na
Seasonality	Day-of-Week	Monday [Mon]	949,139 (14.90%)	846,414 (14.87%)
		Tuesday [Tue]	929,959 (14.60%)	821,775 (14.44%)
		Wednesday [Wed]	931,240 (14.62%)	841,922 (14.80%)
		Thursday [Thu]	944,819 (14.83%)	840,238 (14.77%)
		Friday [Fri]	952,033 (14.95%)	845,619 (14.86%)
		Saturday [Sat]	767,525 (12.05%)	690,033 (12.13%)
		Sunday [Reference]	894,767 (14.05%)	804,182 (14.13%)
	Season-of-Year	Spring [Spr]	1,612,212 (25.71%)	1,461,041 (25.68%)
		Summer [Sum]	1,661,182 (26.49%)	1,506,069 (26.47%)
		Fall [Fal]	1,536,234 (24.50%)	1,405,930 (24.71%)
		Winter [Reference]	1,461,579 (23.31%)	1,317,143 (23.15%)
Time–of–Day	Departure Time Period	Night (midnight-6am) [Night Dep]	91,008 (1.45%)	86,560 (1.52%)
		Morning (6am-noon) [Morn Dep]	2,488,481 (39.68%)	2,274,748 (39.98%)
		Afternoon (noon-6pm) [Aft Dep]	2,404,754 (38.35%)	2,149,523 (37.78%)
		Evening (6pm-modnight) [Ref]	1,286,964 (20.52%)	1,179,352 (20.73%)
	Arrival Time Period	Night (midnight-6am) [Night Arr]	99,010 (1.58%)	93,453 (1.64%)
		Morning (6am-noon) [Morn Arr]	1,720,154 (27.43%)	1,560,265 (27.42%)
		Afternoon (noon-6pm) [Aft Arr]	2,389,925 (38.11%)	2,153,566 (37.85%)
		Evening (6pm-modnight) [Ref]	2,062,118 (32.88%)	1,882,899 (33.09%)
Airport	Origin Airport Type	Type 1 Departure Airport [Dep Type 1]	789,857 (12.60%)	741,863 (13.04%)
		Type 2 Departure Airport [Dep Type 2]	1,685,090 (26.87%)	1,565,435 (27.51%)
		Type 3 Departure Airport [Dep Type 3]	1,833,785 (29.24%)	1,634,272 (28.72%)
		Type 4 Departure Airport [Ref]	1,962,475 (31.29%)	1,748,613 (30.73%)
	Destination Airport Type	Type 1 Arrival Airport [Arr Type 1]	788,497 (12.57%)	740,408 (13.01%)
		Type 2 Arrival Airport [Arr Type 2]	1,684,126 (26.86%)	1,563,899 (27.48%)
		Type 3 Arrival Airport [Arr Type 3]	1,832,631 (29.22%)	1,633,791 (28.71%)
		Type 4 Arrival Airport [Ref]	1,965,953 (31.35%)	1,752,085 (30.79%)
Airline	Type of Airline	Legacy Carriers [Legacy]	2,189,183 (34.91%)	2,216,688 (38.96%)
		Low-Cost Carriers [Low-Cost]	1,666,231 (26.57%)	1,622,640 (28.52%)
		Regional Carriers [Regional]	2,189,714 (34.92%)	1,617,025 (28.42%)
		Non-continental Carriers [Ref]	226,079 (3.61%)	233,830 (4.11%)
Total	na	na	6,271,207	5,690,183

Note: 'na'= not applicable

TABLE 2 Logistic Regression Results

Type	Variable	Estimate		$Pr(<\|z\|)$		NAS				NAS+Carr			
						McFadden		Cox-Snell		McFadden		Cox-Snell	
		NAS	NAS+Carr	NAS	NAS+Carr	Add.	Sub.	Add.	Sub.	Add.	Sub.	Add.	Sub.
Intercept	Intercept	-3.27	-2.55	<2E-16	<2E-16	na	na	na	na	na	na	na	na
Distance	ln (Dist)	0.10	0.09	<2E-16	<2E-16	2.78%	2.12%	2.79%	2.12%	3.34%	1.43%	3.35%	1.42%
Seasonality	Mon	0.11	0.09	<2E-16	<2E-16	22.39%	22.74%	22.40%	22.73%	20.30%	21.63%	20.33%	21.60%
	Tue	-0.03	-0.07	6E-11	<2E-16								
	Wed	0.05	-0.01	<2E-16	0.0031								
	Thu	0.16	0.15	<2E-16	<2E-16								
	Fri	0.11	0.11	<2E-16	<2E-16								
	Sat	-0.18	-0.15	<2E-16	<2E-16								
	Spr	-0.08	-0.14	<2E-16	<2E-16								
	Sum	0.09	0.10	<2E-16	<2E-16								
	Fal	-0.45	-0.42	<2E-16	<2E-16								
Time-of-Day	Night Dep	0.23	-0.58	<2E-16	<2E-16	29.93%	29.77%	29.92%	29.78%	43.95%	42.59%	43.86%	42.67%
	Morn Dep	0.50	-0.31	<2E-16	<2E-16								
	Aft Dep	0.62	0.01	<2E-16	0.0731								
	Night Arr	-0.54	-0.23	<2E-16	<2E-16								
	Morn Arr	-0.17	-0.44	<2E-16	<2E-16								
	Aft Arr	0.09	-0.19	<2E-16	<2E-16								
Airport	Dep Type 1	0.48	0.49	<2E-16	<2E-16	35.71%	32.91%	35.68%	32.94%	22.01%	20.66%	22.03%	20.64%
	Dep Type 2	0.29	0.32	<2E-16	<2E-16								
	Dep Type 3	0.09	0.12	<2E-16	<2E-16								
	Arr Type 1	0.67	0.38	<2E-16	<2E-16								
	Arr Type 2	0.30	0.08	<2E-16	<2E-16								
	Arr Type 3	0.02	-0.09	2E-10	<2E-16								
Airline	Legacy	-0.41	0.37	<2E-16	<2E-16	9.18%	12.46%	9.20%	12.44%	10.40%	13.70%	10.43%	13.67%
	Low-Cost	-0.39	0.76	<2E-16	<2E-16								
	Regional	-0.17	0.58	<2E-16	<2E-16								

Note: 'na'= not applicable

TABLE 3 Quantile Regression Results for the Conditional CDF of NAS Delays

Type	Variable	Coefficient Estimates for Different Quantiles										Fit Contrib. (85%)		Fit Contrib. (95%)	
		5%	15%	25%	35%	45%	55%	65%	75%	85%	95%	Add.	Sub.	Add.	Sub.
Intercept	Intercept	0.10	-1.11	-1.94	-1.45	0.61	3.80	7.15	10.58	18.26	45.79	na	na	na	na
Distance	ln (Dist)	0.23	0.75	1.18	1.44	1.45	1.31	1.13	0.98	0.46	-1.80	0.03%	0.05%	0.16%	0.17%
Seasonality	Mon	-0.05	-0.14	-0.25	-0.26	-0.24	-0.20	-0.19	-0.22	-0.63	-2.56	7.99%	4.50%	12.38%	6.00%
	Tue	-0.01	-0.05	-0.08	-0.10	-0.11	-0.15	-0.32	-0.60	-1.80	-7.58				
	Wed	0.08	0.25	0.42	0.52	0.53	0.54	0.54	0.62	0.44	-1.09				
	Thu	-0.06	-0.13	-0.20	-0.23	-0.21	-0.19	-0.14	-0.13	-0.54	-2.01				
	Fri	-0.07	-0.23	-0.38	-0.52	-0.55	-0.56	-0.59	-0.85	-1.32	-4.06				
	Sat	-0.08	-0.22	-0.30	-0.35	-0.39	-0.41	-0.51	-0.72	-1.78	-6.02				
	Spr	0.01	0.08	0.23	0.34	0.49	0.69	1.00	1.69	3.58	10.40				
	Sum	-0.10	-0.32	-0.43	-0.43	-0.13	0.28	0.88	2.03	5.18	14.43				
	Fal	-0.11	-0.36	-0.55	-0.74	-0.92	-1.04	-1.27	-1.77	-2.34	-0.75				
Time-of-Day	Night Dep	0.34	0.82	1.59	1.95	1.95	1.57	1.23	1.16	1.31	-1.64	4.08%	2.11%	5.38%	2.20%
	Morn Dep	0.45	1.14	1.91	2.72	3.11	3.00	3.14	3.54	4.14	3.00				
	Aft Dep	0.23	0.62	1.03	1.57	1.99	2.22	2.53	3.35	4.91	8.89				
	Night Arr	-0.30	-1.08	-1.76	-2.44	-2.95	-3.54	-3.91	-4.97	-7.14	-11.61				
	Morn Arr	0.22	0.97	1.75	1.90	1.40	0.90	0.35	-0.38	-1.77	-3.75				
	Aft Arr	-0.11	-0.16	-0.27	-0.54	-0.82	-1.05	-1.46	-2.31	-3.54	-4.94				
Airport	Dep Type 1	0.33	1.02	1.71	2.38	2.78	3.25	4.16	5.90	9.01	14.28	83.19%	92.18%	78.75%	90.27%
	Dep Type 2	0.12	0.49	0.83	1.18	1.51	1.82	2.33	3.26	5.31	10.10				
	Dep Type 3	0.13	0.48	0.83	1.03	1.09	1.15	1.22	1.38	1.67	1.46				
	Arr Type 1	1.12	3.66	6.85	8.36	10.11	13.10	19.44	30.30	48.55	84.24				
	Arr Type 2	0.44	1.26	2.32	3.62	4.14	4.20	4.88	6.56	10.42	24.47				
	Arr Type 3	0.22	0.74	1.31	2.02	2.35	2.17	2.20	2.36	3.17	6.26				
Airline	Legacy	-0.29	-0.76	-0.96	-1.22	-1.13	-0.69	-0.20	0.47	1.04	4.12	4.72%	1.16%	3.32%	1.36%
	Low-Cost	-0.62	-1.78	-2.76	-3.59	-3.55	-2.93	-2.22	-1.37	-0.50	3.19				
	Regional	-0.43	-1.04	-1.42	-1.59	-1.13	-0.33	0.34	1.51	3.47	11.22				
Pseudo R^2		0.5%	1.4%	2.7%	3.7%	3.4%	3.3%	4.1%	5.9%	8.8%	12.7%	na	na	na	na

Note: 'na'= not applicable. Underlined coefficient estimates are statistically not significant at 1% level

TABLE 4 Quantile Regression Results for the Conditional CDF of NAS+Carr Delays

Type	Variable	Coefficient Estimates for Different Quantiles										Fit Contrib. (85%)		Fit Contrib. (95%)	
		5%	15%	25%	35%	45%	55%	65%	75%	85%	95%	Add.	Sub.	Add.	Sub.
Intercept	Intercept	-1.22	-3.01	-1.35	3.19	6.03	8.58	11.72	17.21	29.87	63.29	na	na	na	na
Distance	ln (Dist)	0.59	1.43	1.69	1.43	1.37	1.41	1.43	1.40	1.16	2.77	0.37%	0.26%	0.42%	0.47%
Seasonality	Mon	-0.30	-0.68	-0.61	-0.51	-0.47	-0.58	-0.62	-0.94	-1.90	-5.95	7.64%	8.62%	7.20%	9.33%
	Tue	-0.16	-0.28	-0.20	-0.25	-0.32	-0.49	-0.81	-1.55	-3.29	-10.64				
	Wed	-0.06	-0.04	0.03	0.01	-0.05	-0.14	-0.23	-0.56	-1.45	-4.89				
	Thu	-0.28	-0.58	-0.68	-0.65	-0.69	-0.83	-1.10	-1.74	-2.88	-6.96				
	Fri	-0.29	-0.69	-0.83	-0.77	-0.85	-1.04	-1.29	-1.85	-3.08	-7.55				
	Sat	-0.12	-0.06	0.06	0.05	-0.02	-0.22	-0.39	-0.84	-2.24	-6.97				
	Spr	-0.17	-0.21	-0.12	0.12	0.37	0.77	1.43	2.74	5.43	11.24				
	Sum	-0.30	-0.54	-0.46	-0.02	0.49	1.30	2.58	4.81	9.21	16.89				
	Fal	-0.15	-0.21	-0.36	-0.50	-0.70	-0.91	-1.20	-1.36	-0.93	0.20				
Time-of-Day	Night Dep	2.94	4.13	3.85	3.25	3.09	3.40	4.22	6.51	11.86	45.05	9.65%	9.50%	19.45%	14.32%
	Morn Dep	0.95	3.17	3.22	2.70	2.56	2.62	3.07	3.75	5.25	11.11				
	Aft Dep	0.31	0.86	1.27	1.08	1.05	1.08	1.36	1.81	3.01	8.28				
	Night Arr	-0.42	-0.81	-0.82	-0.64	-0.87	-1.44	-1.95	-2.77	-3.76	-7.00				
	Morn Arr	2.55	5.36	3.73	3.28	3.24	3.48	3.91	4.77	7.09	17.58				
	Aft Arr	0.12	0.29	0.24	-0.04	-0.34	-0.64	-1.23	-2.12	-3.42	-4.04				
Airport	Dep Type 1	0.74	1.85	2.82	3.28	3.87	4.83	6.21	7.87	8.93	3.20	44.33%	46.49%	39.56%	40.82%
	Dep Type 2	0.89	2.18	3.28	3.63	4.01	4.66	5.82	7.19	8.23	1.44				
	Dep Type 3	0.68	1.68	2.44	2.58	2.63	2.75	3.35	4.15	4.88	0.13				
	Arr Type 1	0.55	1.79	2.57	3.57	5.50	8.54	13.38	20.81	31.60	48.65				
	Arr Type 2	0.06	0.23	0.46	0.78	0.99	1.17	1.71	2.96	6.33	18.21				
	Arr Type 3	0.02	-0.09	-0.24	-0.27	-0.36	-0.68	-0.91	-0.92	-0.45	5.89				
Airline	Legacy	-0.49	-0.72	-0.30	-0.04	0.45	1.06	2.01	3.51	5.69	8.22	38.01%	35.13%	33.38%	35.06%
	Low-Cost	-0.96	-2.62	-3.53	-3.39	-2.98	-2.94	-3.02	-3.53	-5.29	-12.22				
	Regional	-0.44	-0.39	0.60	1.37	2.48	4.13	6.99	11.37	18.27	28.83				
Pseudo R^2		0.9%	2.7%	1.9%	3.0%	2.0%	3.3%	3.7%	2.2%	2.5%	4.2%	na	na	na	na

Note: 'na'= not applicable. Underlined coefficient estimates are statistically not significant at 1% level

TABLE 5 Ten-Fold Cross Validation Results and Out-of-Sample Testing Results

Results Type		Ten-Fold Cross Validation Results				Out-of-Sample Testing Results			
Delay Type		NAS Delay		NAS+Carr Delay		NAS Delay		NAS+Carr Delay	
Dataset		2013 Training	2013 Validation	2013 Training	2013 Validation	All 2013	All 2014	All 2013	All 2014
Logistic Regression R^2	McFadden	2.85%	2.87%	3.23%	3.22%	2.88%	3.01%	3.23%	3.42%
	Cox-Snell	2.03%	2.06%	2.82%	2.79%	2.01%	2.20%	2.89%	3.24%
Quantile Regression Pseudo R^2	$\tau = 5\%$	0.46%	0.44%	0.89%	0.88%	0.47%	0.61%	0.90%	1.02%
	$\tau = 15\%$	1.44%	1.42%	2.84%	2.83%	1.44%	1.67%	2.84%	3.09%
	$\tau = 25\%$	2.70%	2.68%	3.21%	3.20%	2.70%	3.08%	3.22%	3.52%
	$\tau = 35\%$	3.66%	3.64%	2.19%	2.18%	3.66%	4.07%	2.19%	2.48%
	$\tau = 45\%$	3.35%	3.34%	2.05%	2.04%	3.36%	3.99%	2.05%	2.20%
	$\tau = 55\%$	3.28%	3.26%	2.23%	2.22%	3.28%	3.65%	2.23%	2.24%
	$\tau = 65\%$	4.10%	4.09%	2.68%	2.66%	4.11%	4.20%	2.68%	2.44%
	$\tau = 75\%$	5.86%	5.84%	3.33%	3.31%	5.87%	5.47%	3.34%	2.74%
	$\tau = 85\%$	8.81%	8.78%	4.00%	3.97%	8.82%	7.55%	4.00%	3.01%
	$\tau = 95\%$	12.65%	12.56%	4.11%	4.05%	12.66%	11.05%	4.13%	3.74%
Joint Model Log-Likelihood	Full Model	na	na	na	na	-5,097,164	-4,918,529	-7,573,460	-7,446,381
	Naïve Model	na	na	na	na	-5,357,231	-5,158,484	-7,856,383	-7,717,994
	Difference	na	na	na	na	260,068	239,955	282,923	271,613

Note: 'na'= not applicable