[INCOMPLETE DRAFT]

The Flight Operations Decision Process

Nicholas G. Rupp*
Department of Economics
East Carolina University
Greenville, NC 27858-4353

George M. Holmes
University of North Carolina
Cecil G. Sheps Center
Chapel Hill, NC 27599-7590

April 16, 2004

Abstract

The purpose of this paper is to determine how carriers make flight operations decisions. The paper examines flight operation decisions (cancel, delay, or on time) by using a sample of domestic flights from the U.S. Bureau of Transportation Statistics (BTS) data in 2002. Each day carriers are faced with the problem of how best to allocate their scarce resources in order to maximize profits. We model flight outcomes as a function of economic, route competition, airport competition, logistical, and weather variables. Previous research on the service quality in the airline industry has examined delays (Mayer and Sinai, 2003a & 2003b; Mazzeo, 2003), cancellations (Rupp and Holmes, 2004), and irregular operations (Rupp, Holmes, and DeSimone, 2004). This paper is novel in that it is the first to examine flight outcomes during normal operations.

---

*Contact at ruppn@mail.ecu.edu or (252) 328-6821.

1 Introduction

The purpose of this paper is to determine how carriers make flight operations decisions. The paper examines flight operation decisions (cancel, delay, or on time) by using a sample of domestic flights from the U.S. Bureau of Transportation Statistics (BTS) data in 2002. Each day carriers are faced with the problem of how best to allocate their scarce resources in order to maximize profits. We model flight outcomes as a function of economic, route competition, airport competition, logistical, and weather variables. Previous research on the service quality in the airline industry has examined delays (Mayer and Sinai, 2003a & 2003b; Mazzeo, 2003), cancellations (Rupp and Holmes, 2004), and irregular operations (Rupp, Holmes, and DeSimone, 2004). This paper is novel in that it is the first to examine flight outcomes during normal operations.

Using a nested logit specification, we model the decision process as a sequential decision where the carrier first decides whether or not the plane departs on time, and for planes which do not leave on time, whether to delay or cancel the flight. While the Rupp et al. (2004) paper also uses a nested logit model, their sample of the day in which an airport closed due to security concerns, may not reflect normal daily operation decisions made by carriers. This analysis of flight outcomes will consider various economic variables, including average and potential flight revenue, yield, route load capacity, and seating capacity of the aircraft. We also investigate the link between service quality and competition at both the route level and airport level. All estimations control for logistical factors at the aircraft, airline, and route level. These important logistical factors include departure time, number of daily scheduled flights on route, last flight of day, and the amount of time between scheduled departures. Finally, since most airports also serve as weather reporting stations, we control for severe weather by using rain, temperature, and frozen precipitation.

The remainder of the paper is organized as follows. The next section discusses the data that we analyze. Section 3 outlines our econometric model, and section 4 presents the results of estimating the model. Section 5 concludes the paper.

2 Data

These data consist primarily of individual domestic flights from the Bureau of Transportation Statistics (BTS) TranStats database.[1] All U.S. carriers with revenues from domestic passenger flights of at least one percent of total industry revenues are required to report flight operation decisions for each flight. We obtain every domestic flight in 2002 for the ten largest[2] U.S. carriers. These ten carriers accounted for more than 90 percent of 2001 domestic revenues.

All variables are constructed from the original data set of 5,267,770 flights in 2002. Table 1 indicates that during the year following the September 11th terrorist attacks there was a 12 percent reduction in scheduled flights (700,000 fewer scheduled flights from 2001) and a noticeable improvement in airline flight operations performance. Specifically, 2002 marked the lowest yearly cancellation rate (1.2 percent) and the highest on-time departure rate (83.9 percent) since the BTS began compiling on-time performance statistics in 1995.

Due to the computational constraints presented by the nested logit estimation, we randomly select a 1 percent subsample resulting in 52,656 flights. Incomplete weather reporting data in addition to missing/incorrect aircraft tail numbers reduces the sample by 3,793 flights or 7.2 percent. Table II provides summary statistics for the subsample of 2002 flights.

This paper examines the determinants of flight outcomes: cancellations, delays, or on time. For each flight, exactly one of these indicators (canceled, delayed, or on time) equals one, while the

---

[1] http://www.transtats.bts.gov.

[2] These are Alaska, America West, American, American Eagle, Continental, Delta, Northwest, Southwest, United and US Airways.

other two equal zero. This paper adopts the Department of Transportation's (DOT) convention that a flight is considered on time if it departs no more than 15 minutes after its scheduled departure.[3] The DOT on time measure is the industry standard for on-time performance and is commonly cited in company advertisements and public relations media. In the subsample, 86.1 percent of scheduled flights depart on-time, while 12.6% experience a delayed departure and 1.3 percent are canceled.

We model flight outcomes as a function of economic, route competition, airport competition, logistical, and weather variables. While a majority of the variables are constructed from the BTS TranStats data base, some variables are obtained from other sources. Aircraft tail numbers provided by the BTS are matched to the FAA Aircraft Registry database in order to determine the *seating capacity* of the aircraft.[4] The average aircraft has seating capacity for 156 passengers.

We also match individual flights to quarterly passenger fare data from the *Airline Origin and Destination Survey.* This survey by the U.S. Office of Airline Information of the BTS is a 10% sample of domestic airline tickets from reporting carriers and is also located at the TranStats database. These fare data enable us to estimate flight revenue since actual flight revenue is unobservable. The \$157.48 average one-way *air fare* is the 2002 quarterly average for carrier j on route r. Round-trip itineraries are divided by two in order to obtain the one-way *air fare.* The average revenue passenger mile or route *yield* is defined as the average one-way *air fare* divided by the flight *distance* (miles between airports).

The average one-way *air fare* is multiplied by *seating capacity* to obtain *potential revenue* per flight. In the sample, the average *potential revenue* is \$25,172 per flight. *Potential revenue*

---

[3] We have recently learned that flights which depart 15 minutes after their scheduled departure time are considered late by the DOT. In our sample, 521 of the 52,656 depart exactly 15 minutes after their scheduled time, hence, in future versions of this paper, these flights will be counted as "late". In this version of the paper, however, such flights are considered "on-time".

[4] In situations where the tail number is unknown, seating capacity if found by substituting the median value of seats on comparable flights (i.e., same carrier, route, flight number, and month).

provides an accurate depiction of flight revenue if every seat is occupied on the aircraft. Hence *potential revenue* provides an upper bound of flight revenue since flights are typically not 100 percent full. A more realistic representation of flight revenue may be provided by *average revenue* per flight which is found by multiplying the average one-way *air fare* by the monthly average number of occupied seats (*seating capacity* multiplied by average *load capacity*) for carrier j on route r. *Load capacity* is the proportion of total seats that are occupied and is obtained from T-100 domestic market data at the TranStats database. During 2002 the average *load capacity* was 66.1 percent and the mean *average revenue* was $16,883 per flight.

To determine if carriers provide better service to and from an airline's hub airport, we include indicator variables for both *airline hub origination* and *airline hub destination* flights.[5] Hub flights are especially important for carriers given that consumer demand is higher for airlines with large operations from an origin city (Morrison and Winston, 1989). Morrison and Winston (1995, p. 44) also report that at a typical hub a majority of passengers are making connections. In addition, some of these passengers may be making international connections. A flight delay or cancellation for a *hub destination* flight can cause considerable passenger inconvenience, especially so for those attempting to make connections. Hence we create an interaction variable (*seating capacity*hub destination*) which multiplies the *seating capacity* of the aircraft by one for *hub destinations* and by zero otherwise.

Given that 40 percent of flights originate from an airline's hub airport and 40 percent of the flights are destined for an airline hub in the sample, we investigate the flight operation decisions at various airline hub sizes. Using the convention of Mayer and Sinai (2003a), we too define airline hub size using the number of connecting flights at the origination airport with airlines

---

[5] We use the airline hub definition of Mayer and Sinai (2003a) that an airport is considered a hub for the airline if the carrier has 26+ connecting flights.

that offer 71+ connections being *large airline hub origination*, between 46 and 70 connections termed *medium airline hub origination*, and *small airline hub origination* having between 26 and 45 connections. Similar size categories are used for *hub destination* flights. We also create three more interaction variables by interacting *seating capacity* with *airline hub destination* size. *Airport concentration origination (destination)* is the sum of the squared carrier shares as a percentage of all daily flights at the origination (destination) airport.

We also consider route competition measures including *monopoly* routes which comprise 55.3 percent of the sample. One-third of the sample are for routes served by two carriers with the *large duopoly carrier* providing more daily scheduled on a route compared to it counterpart the *small duopoly carrier*.[6]

We include a variety of logistical variables including the number of *hours until next flight* by carrier j on route r. For the final flight of the day, we count the number of hours until the first flight the following day. The average wait between flights is 6 hours. On routes served by one daily flight, the hours of waiting time equals the maximum of 24. The *last flight of day* is the final scheduled flight by carrier j on route r. Red-eye flights (mostly from the West to the East coast) that are scheduled to depart the following day between midnight and 2:30 a.m. are considered as the "same day" for the *last flight of day* designation. To track the occurrence of cascading delays (Mazzeo, 2003) later in the day, *time01* represents the renormalized scheduled departure time between 0 (midnight) and 1 (11:59 p.m.).

We also count the *daily total number of carrier flights* on route r by carrier j to examine if carriers make flight operation decisions which penalize routes with frequent daily service. For example, passengers on a sparsely occupied plane on a route with frequent daily service might

---

[6] These monopoly and duopoly definitions are based solely on the number of major carriers, hence low-cost carriers such as JetBlue, AirTran, and regional airlines like Comair, ASA, and Continental Express are excluded from these measures of route competition.

be subject to more flight cancellations. Carriers provide an average of 6.3 flights per day on a route. Finally, as a proxy for airport congestion we use the total number of daily operations (i.e., take-offs and landings) at the origination airport.

Given that most airports also serve as active weather reporting stations,[7] we obtain daily weather data at origination and destination airports from the U.S. National Oceanic & Atmospheric Administration (NOAA). We include rain, temperature, and frozen precipitation.[8] *Rain origination (destination)* is the amount of daily precipitation at the origination (destination) airport measured in hundredths of an inch (e.g., 50 equals half an inch of rain). *Minimum temperature origination (destination)* is the daily coldest temperature reading in Fahrenheit at the origination (destination) airport. Finally, *frozen precipitation origination (destination)* is an interaction term since it multiplies *rain origination (destination)* by one if daily *minimum temperature origination (destination)* < 33, otherwise zero.

3 Econometric Model

Each day airlines must make flight operation decisions in an effort to maximize profits. In addition, carriers must reallocate resources due to equipment failures, crew shortages, severe weather, and periodic security-related airport closures.[9] For each flight the agent managing air traffic for a representative airline must decide between three possible departure outcomes: cancellation, delay, or on-time.

A choice set consisting of three discrete outcomes suggests the use of a discrete choice econometric model. Suppose that the (net future discounted) profit from flight i with outcome j

---

[7] In cases of missing weather data, we substituted the nearest weather reporting station within a twenty-five mile radius.

[8] We would have liked to include both wind and snow, however, many weather stations fail to report these data.

[9] See Rupp, Holmes, and DeSimone (2004) for an analysis of flight operation decisions following security-related airport closures.

which incorporates both short-term (e.g., rebooking costs) and long-term (e.g., service quality reputation) effects is represented as

$$\pi_i(j) = \pi_j(X_i) + \varepsilon_{ij} \tag{1}$$

where for outcome j, $\pi_j(X_i)$ is a deterministic function of profits from the vector of observable characteristics X_i of flight i. We assume that $\pi_j(X_i)$ can be approximated by a linear function of X_i, hence the profit function becomes

$$\pi_i(j) = X_i\beta_j + \varepsilon_{ij} \tag{2}$$

where ε_{ij} represents unobserved factors that influence profit. For example, the profit from flight i being canceled is

$$\pi_i(CANCEL) = X_i\beta_{CANCEL} + \varepsilon_{i,CANCEL} \tag{3}$$

Assume (for the moment) that each ε_{ij} is independent and drawn from an identical Weibull distribution. Then the choice of which outcome j maximizes profit for flight i, as represented in equation (2), is equivalent to the conventional multinomial logit model (Domencich and McFadden, 1975),

$$\Pr(i \text{ chooses outcome } j) = \frac{e^{X_i\beta_j}}{\sum_{k=1\ldots3} e^{X_i\beta_k}} \tag{4}$$

where identification requires $\beta_k \equiv 0$ for one of the three outcomes. It is well-known that the multinomial logit model assumes the independence of irrelevant alternatives (IIA): the ratio of any two outcome choice probabilities is independent of whether the third option is available. For instance, if for a particular flight the probability of each outcome is 1/3, the elimination of one option (e.g., on time departure) implies that the probability of each of the other two outcomes (e.g., delay and cancellation) is 1/2, so that the ratio of these probabilities remains equal to one.

An alternative discrete choice specification that does not impose the restrictive IIA assumption is the nested logit model. Moreover, the nested logit model in our setting seems natural given that the decision between these three flight outcomes may likely occur sequentially as two binary choices: first, the carrier decides whether the flight departs on time or not on time, and second, for flights which are not departing on time, the carrier must decide to either delay or cancel the flight.[10] The flight decision process is displayed in Figure 1.[11]

Next we outline the econometric method involved in estimating the nested logit model. Define the inclusive value I_i as the natural log of the sum of exponentiated (expected) profits from not canceling:

$$I_i = \ln \sum_{k=CANCEL,DELAY} e^{X_i \beta_k} \tag{5}$$

Calculating the probability of choosing each outcome j is now a three-step process which is outlined in Greene (2000):

1. Conditional on not departing on time, the probability that a flight is canceled (rather than delayed) is estimated equivalently to a standard (binary choice) logit using only the flights which were not on time:

$$\Pr_i(\text{ CANCEL} \mid \text{ON TIME=0}) = \frac{e^{X_i \beta_{CANCEL}}}{1 + e^{X_i \beta_{CANCEL}}} \tag{6}$$

---

[10] The nested logit model does not require a sequential decision process: an econometrically equivalent interpretation is that the decision between the three outcomes occurs at one time, but the errors are heteroskedastic. The sequential decision interpretation, however, is natural in this context.

[11] Of course, an alternative plausible decision ordering is that the carrier first chooses to cancel or not cancel a flight with the second decision being whether the flight departs on time or is delayed. We plan to implement a Vuong test of to determine the prefered model specification.

The first term in the denominator is simplified by the normalization that $\beta_{DELAYED}$ equals 0.

2. Using the same normalization, the inclusive value is

$$I_i = \ln\left(1 + e^{X_i\beta_{CANCEL}}\right) \tag{7}$$

3. The probability that the flight is on time (rather than not on time) is estimated equivalently to a standard logit model for the decision between on time and not on time (either delaying or cancelling the flight), augmented by an additive inclusive value term in the exponential expressions in both the numerator and denominator:

$$\Pr_i(\text{ON TIME}) = \frac{e^{X_i\beta_{ONTIME}+\tau I_i}}{1 + e^{X_i\beta_{ONTIME}+\tau I_i}} \tag{8}$$

This is the standard logit model for on time vs. "other" with the additional inclusive value term.

The unconditional probabilities of cancel and delayed departure are straightforward to compute. Because the estimated β and τ parameters are difficult to interpret, we also report marginal effects

$$\text{me}_j(x) \equiv \frac{1}{N} \sum_{i=1\ldots N} \frac{\partial \Pr_i(j)}{\partial x_i}. \tag{9}$$

In sum, we have modeled the flight operation decision process as one in which the carrier first attempts to maximize the number of on-time departures. For flights not departing on time are delayed with the intent of eventually departing, if possible.

4 Results

We estimate four nested logit models. The only difference among the first three models is the selection of the economic variable: *average revenue* (model 1), *potential revenue* (model 2), and *yield* (model 3). Model (4) examines how airline hub size affects flight operation decisions, hence this model includes *large, medium*, and *small airline hub origination* and *destination* indicator variables instead of the more general hub dummy specification. Model (4) also interacts the size of the hub destination with *seating capacity* to determine if the flight operations decisions are different for larger hub destination flights. Given that the first three models use the same set of explanatory variables and just one different economic variable, we begin our analysis by discussing the similarities across these three specifications.

Overall, the results changed vary little between models 1-3. Only three of the forty-eight estimated parameters experienced a change in statistical significance. Specifically, turning our attention to the economic variables we find that for all three models routes with higher *load capacity* have significantly fewer flight cancellations. It is not surprising to learn that carriers attempt to avoid canceling full flights. Interpreting the marginal effects estimate for route *load capacity* -0.0582 in model (1) as follows, a one-standard deviation increase (0.127) in mean *load capacity* results in a 0.74 percentage point reduction in flight cancellations. While this might seem small, recall that the average flight cancellation rate in the sample is just 1.3 percent, hence this change is a 57% reduction in flight cancellations for routes with fuller flights. The tables do not indicate the variables which obtain marginal statistical significance (10% level), however, it is interesting to note that routes with higher *load capacity* are also marginally less likely to depart on time. This result is consistent with the claim that it takes longer to load a full flight.

Models 1-3 also indicate that aircraft with more *seating capacity* have significantly fewer flight cancellations. The *seating capacity* marginal effect in model (1) indicates that 50 additional seats (approximately a one-standard deviation increase) from the mean *seating capacity* reduces the cancellation rate by 0.335 percentage points (or about a 25 percent reduction). are also subject to fewer flight cancellations. One minor difference in statistical significance among these three models is found for the *seating capacity*hub capacity* interaction term since aircraft with larger *seating capacity* and destined for a hub are less like to depart on time. This interaction variable changes from marginal significance in the first two models to standard significance levels in model (3).

Each of the first three models consider a different measure of flight revenue. The first model uses *average revenue* which is the *average route fare* multiplied by the average number of occupied seats on the aircraft. Given that Rupp, Holmes, and DeSimone report that higher potential revenue flights are less likely to be delayed and more likely to be on time following a security-related airport closure, our expectations were that profit maximizing behavior should dictate a reduction in the cancellation rate and potentially higher on time departure rates to keep the high revenue generating passengers happy. We find evidence that suggests otherwise. Higher *average revenue* flights are more likely to be canceled, while there is an indeterminate effect of *average revenue* for on-time departures. Model (2) estimations report a similar contrarian result for *potential revenue* flights. Recall that *potential revenue* is *average fare* multiplied by aircraft seating capacity and hence implicitly assumes that all seats are occupied. flights also contradict Model (3) which uses route *yield*, however, provides some evidence that supports our expectation that carriers provide better service quality on higher revenue flights. Routes with higher *yield*, which is the average revenue per passenger mile, are significantly more likely to depart on time while the link between *yield* and cancellation rates is ambiguous. Better on time performance for routes with higher *yield* is consistent with a theoretical passenger switching model proposed by Suzuki (2000) which suggests that passengers whom experience poor service quality will switch carriers. Next we turn to the route competition variables.

We find a consistent result across all three models that less competitive routes have higher on-time departure rates. Specifically, monopoly routes are more likely to push back from the gate on time. A similar finding has also been reported by Mayer and Sinai (2003b). The on-time departure rates of duopolists is not significantly different from routes served by three+ carriers. Additionally, we find lower cancellation rates for routes serve by monopoly and duopoly carriers compared to more competitive routes, however, these cancellation rates are significantly lower for only the small duopoly carriers. In addition to route competition, we also consider whether airport competition influences flight operation decisions.

In the first three models we use both airport concentration and an indicator variable for hub airlines as measures of airport competition. Brueckner (2002) presents a theoretical model and some empirical evidence that concentrated airports have fewer flight delays since the dominant carrier fully internalizes costs of airport congestion. Our results in models 1-3 indicate that flights which originate at highly concentrated airports have significantly lower on-time departure rates. The magnitude of this airport concentration affect, however, is rather small. For example, the marginal effect of -0.0303 for *airport concentration at origination* suggests that a one-standard deviation increase (0.22 points) reduces the on-time departure rate by 0.67 percentage points (which corresponds to a change of less than 1 percent). Flight operation decisions appear to be made independent of *airport concentration at destination*.

We also find significantly lower on-time departure rates for flights which originate from an airline's hub. Worse on-time performance by hub airlines may reflect the banking of flights by hub airlines in an effort to minimize passenger connection time. The result is a difficult to maintain flight schedule. Hence carriers may choose not to depart on time in order to accommodate connecting passengers who are late arriving at the hub. We find negative coefficient estimates across all three models for *airline hub origination.* This suggests that these flights have fewer cancellations, however, none of the airline hub origination estimates achieve statistical significance. Using a longer sample period, Rupp and Holmes (2004) report significantly lower cancellation rates for hub airline origination flights. Upon first glance, the flight operations decision appears independent of whether the plane is destined for an airline's hub. If we expand our statistical threshold to include marginal significance (10% level), we find airline hub destination flights are more likely to be canceled. This result is somewhat surprising given that we expected carriers to place a high value on accommodating connecting passengers. Let us now examine some logistical variables.

The first two models suggest that flight decisions are made independent of flight *distance*. The third model, however, indicates that longer flights are more likely to push back from the gate on time. *Hours until next flight*, which measures the lag time between flights, is inversely correlated with on-time departure rates in all three models. In addition, in the yield specification, significantly fewer flight cancellations occur when there is a longer wait before the next flight. These results are consistent with carriers opting to delay a departure in order to accommodate late arriving passengers.

The logistical variable with the most explanatory power is the *renormalized schedule departure time*. Consistent with Mazzeo (2003), we also find that flight schedules deteriorate during the day with significantly fewer on-time departures later in the day. Specifically, the marginal effects from model (1) of -.3026 suggests that flights which depart at 8 p.m. instead of 8 a.m. (which is a 0.5 increase in the *renormalized schedule departure time*) have 15.1 percentage points lower on-time departure rates (or about 18% less likely to be on time). Given that the plane does not depart on time, we also find significantly lower cancellation rates (and hence more delays) occur for flights scheduled later in the day.

Last flight of day is included in the estimation on the premise that a carrier avoids canceling the last flight since would cause a considerable loss in passenger goodwill. We find no evidence that supports (or denies) this claim. Instead, we find that the *last flight of day* is more likely to depart on time. This suggests that carriers improve on time performance before the end of the day. In all three models, routes with more frequent daily service are more likely to experience flight cancellations. This indicates that carriers maybe opportunistically consolidating flights on frequently served routes. Finally, origination airports with more daily operations have significantly lower on-time departure rates, however, we find no link between airport congestion and flight cancellations. Likewise, busy destination airports also have lower on-time departure rates, yet no effect on flight cancellations. Next we present the weather variables.

As expected weather plays an important role in the flight operations decision. Precipitation especially in the form of frozen precipitation at both origination and destination airports contribute to more cancellations and fewer on-time departures. Cold weather at the destination can also lead to fewer on-time departures.

4.1 Airline hub size

Table VI presents nested logit estimations for various airline hub sizes. Overall, most of the results in model (4) are similar to the previously discussed models. First we will discuss the hub airline findings, then we will briefly mention the few parameters which experience a change in statistical significance. Recall that in the previous specification we found airline hub origination flights were less likely to depart on time and neither more, nor less likely to be canceled. We find similar results as small, medium, and large airline hub origination flights all have significantly worse on time departure rates compared to non-hub airlines. The marginal effects for on time flights at large and medium airline hubs (-0.0385 and -0.0410, respectively) are more than twice as large as the small airline hub (-0.0169), with the larger hubs experiencing worse on time performance. Mayer and Sinai (2003b) document similar size effects of airline hubs in their study of flight delays. We also interact seating capacity with the size of airline's hub destination. The only notable finding is that large planes destined for a large airline hub have significantly fewer flight cancellations.

The inclusion of the three hub sizes results in a couple of notable changes. First, routes with higher load capacity are now significantly less likely to register an on-time departure. Second, the total number of airport operations at both origination and destination airports no longer attributed to fewer on-time departures.

5 Conclusion

Not yet completed.

References

‘Aviation and Transportation Security Act’, 2002, *US Code Congressional & Administrative News*, S. 1447, US Government Printing Office, Washington, D.C..

Brueckner, Jan K. 2002. ‘Airport Congestion when Carriers have Market Power’, *American Economic Review*, 92:5, pp. 1357-1375.

Domencich, Thomas A. and Daniel McFadden. 1975. Urban Travel Demand : A Behavioral Analysis : A Charles River Associates Research Study. Amsterdam : North-Holland Pub. Co. ; New York : American Elsevier.

Greene, William H. 2000. *Econometric Analysis 4th Edition.* Upper Saddle River, NJ: Prentice Hall.

Mayer, Christopher and Todd Sinai. 2003a. ‘Network Effects, Congestion Externalities, and Air Traffic Delays: or Why All Delays Are Not Evil’, *American Economic Review*, 93:4, pp. 1194-1215.

Mayer, Christopher and Todd Sinai. 2003b. ‘Why Do Airline Schedules Systematically Underestimate Travel Time?’, working paper, Wharton School of Business.

Mazzeo, Michael J. 2003. ‘Competition and Service Quality in the U.S. Airline Industry’, *Review of Industrial Organization*, 22:4, pp. 275-296.

Morrison, Steven and Clifford Winston. 1989. ‘Enhancing the Performance of the Deregulated Air Transportation System’, *Brookings Papers on Economic Activity: Microeconomics,* 1, pp. 61-112.

_____. 1995. *The Evolution of the Airline Industry.* Washington, D.C.: The Brookings Institution.

Rupp, Nicholas G. and George M. Holmes. 2004. 'An Investigation Into the Determinants of Flight Cancellations' East Carolina University Department of Economics Working Paper.

Rupp, Nicholas G., George M. Holmes, and Jeff DeSimone. 2004. 'Airline Schedule Recovery after Airport Closures: Empirical Evidence since September 11th' NBER working paper #9744.

Suzuki, Yoshinori. 2000. 'The Relationship between On time Performance and Airline Market Share', *Transportation Research: Part E: Logistics and Transportation Review*, 36:2, pp. 139-54.

Vuong, Quang H. 1989. 'Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses', *Econometrica*, 57:2, pp. 307-333.

Figure 1: The flight operations decision process

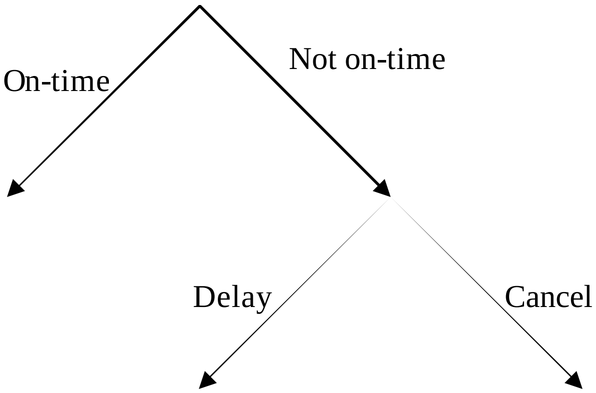

Table I: Summary of Major Carriers On-time Performance[1], 1995-2002

Year	Scheduled Flights	On-Time Departures	Late Departures	Flights Canceled	Pecent On-time Departures	Pecent Late Departures	Percent Canceled
1995	5,327,435	4,315,691	919,839	91,905	81.01	17.27	1.73
1996	5,351,983	4,120,963	1,102,484	128,536	77.00	20.60	2.40
1997	5,411,843	4,369,447	944,633	97,763	80.74	17.45	1.81
1998	5,384,721	4,225,308	1,014,904	144,509	78.47	18.85	2.68
1999	5,527,884	4,281,989	1,091,584	154,311	77.46	19.75	2.79
2000	5,683,047	4,176,404	1,319,153	187,490	73.49	23.21	3.30
2001	5,967,780	4,551,576	1,185,006	231,198	76.27	19.86	2.79*
2002	5,267,770	4,421,223	781,566	64,981	83.93	14.84	1.23
Average	5,490,308	4,307,825	1,044,896	137,587	78.55	18.98	2.34

[1] Source: Bureau of Transportation Statistics (www.bts.gov/oai/on_time_2002/), accessed 1/30/2004. Major carriers include: Alaska, America West, American, Continental, Delta, Northwest, Southwest, TWA, United, US Airways.

*Because of the shutdown of the air transportation system as a result of the terrorist attacks on September 11, 2001, the BTS granted air carriers waivers that we would not count the forced cancellations against the air carriers' on-time performance ratings. Hence, the "Percent Canceled" does not reflect these forced cancellations. For historical purposes, however, the number of "Flights canceled" includes all cancellations regardless of reason.

Table II: Summary Statistics for the 1% sample of 2002 U.S. domestic flights by major carriers[1] (n=48,863)

	Mean	Standard Deviation	min	max
Flight Outcomes				
Proportion ontime[2]	0.861	0.346	0	1
Proportion delayed	0.126	0.331	0	1
Proportion canceled	0.013	0.114	0	1
Economic Variables				
Average one-way air fare (quarterly)	157.48	65.56	0	556.50
Average revenue (in $10,000s)	1.6883	1.1603	0	14.9425
Potential revenue (in $10,000s)	2.5172	1.5848	0	20.0404
Yield (average revenue passenger mile)	0.329	0.301	0	5.059
Route load capacity (monthly average)	0.661	0.127	0.057	0.985
Seating capacity of aircraft (in 100's)	1.557	0.527	0.150	4.950
Seating capacity*hub destination	0.647	0.857	0	4.950
Seating capacity*small airline hub dest.	0.170	0.507	0	4.950
Seating capacity*medium airline hub dest.	0.287	0.662	0	4.950
Seating capacity*large airline hub dest.	0.191	0.558	0	4.950
Route Competition Variables				
Carriers on route	1.571	0.736	1	5
Monopoly	0.553	0.497	0	1
Large duopoly carrier	0.236	0.425	0	1
Small duopoly carrier	0.104	0.305	0	1
Airport Competition Variables				
Airport concentration at origination	0.408	0.220	0.123	1
Airport concentration at destination	0.408	0.219	0.123	1
Airline hub origination	0.397	0.489	0	1
Small airline hub origination	0.120	0.325	0	1
Medium airline hub origination	0.161	0.367	0	1
Large airline hub origination	0.116	0.320	0	1
Airline hub destination	0.401	0.490	0	1
Small airline hub destination	0.121	0.327	0	1
Medium airline hub destination	0.167	0.373	0	1
Large airline hub destination	0.113	0.317	0	1
Logistical Variables				
Distance (in 100's miles)	7.604	5.850	0.36	49.62
Hours until next flight	5.965	6.175	0	24
Renormed scheduled departure time	0.564	0.193	0.01	1
Last flight of day	0.246	0.431	0	1
Daily total carrier flights on route	6.290	4.202	1	32
Daily total origination airport flights (100's)	6.158	4.955	0.02	19.95
Daily total destination airport flights (100's)	6.128	4.887	0.01	19.84
Weather Variables				
Rain origination (hundredths of an inch)	9.51	33.04	0	804
Rain destination (hundredths of an inch)	9.23	31.66	0	804
Minimum temperature origination (Farenheith)	50.69	17.26	-26	92
Minimum temperature destination (Farenheith)	50.77	17.30	-29	92
Frozen precipitation origination (hundredths)	0.58	4.91	0	99
Frozen precipitation destination (hundredths)	0.64	5.21	0	99

Note: Variables are constructed from the original data set of every domestic flight by major carriers in 2002.

[1] The sample includes the following major carriers: Alaska, American, American Eagle, American West, Continental, Delta, Northwest, Southwest, United, and US Airways.

[2] Ontime flights push back from the gate no more than 15 minutes after their scheduled departure time.

Table III: Nested Logit estimations of flight operation decisions for 1% sample of all 2002 U.S. domestic flights

Model (1) - Average revenue	Ontime			Cancel		
	Coeff	Std Error	Marg.Eff.	Coeff	Std Error	Marg.Eff.
Economic Variables						
Average revenue (in $10,000s)	0.0139	0.0291	0.0016	0.2090 *	0.1020	0.0021
Route load capacity (monthly average)	-0.4240	0.2375	-0.0496	-5.9200 **	0.4635	-0.0582
Seating capacity of aircraft (in 100's)	-0.0112	0.0620	-0.0014	-0.6412 **	0.2378	-0.0067
Seating capacity*hub destination (in 100's)	-0.1069	0.0564	-0.0122	-0.2962	0.1980	-0.0020
Route Competition Variables						
Monopoly	0.1558 **	0.0559	0.0175	-0.2180	0.1721	-0.0040
Large duopoly carrier	0.0896	0.0514	0.0101	-0.0937	0.1645	-0.0019
Small duopoly carrier	0.0644	0.0616	0.0072	-0.4050 *	0.1908	-0.0050
Airport Competition Variables						
Airport concentration at origination	-0.2682 **	0.0726	-0.0303	0.1406	0.2237	0.0043
Airport concentration at destination	-0.0276	0.0708	-0.0031	0.0306	0.2263	0.0006
Airline hub origination	-0.1918 **	0.0423	-0.0217	-0.1815	0.1599	0.0001
Airline hub destination	0.1493	0.1001	0.0170	0.6104	0.3138	0.0049
Logistical Variables						
Distance (in 100's miles)	0.0012	0.0040	0.0001	-0.0114	0.0148	-0.0001
Hours until next flight	-0.0202 **	0.0055	-0.0023	-0.0382	0.0211	-0.0002
Renormed scheduled departure time	-2.6750 **	0.1160	-0.3026	-1.5877 **	0.3097	0.0115
Last flight of day	0.4657 **	0.0757	0.0527	0.1870	0.2710	-0.0029
Daily total carrier flights on route	-0.0031	0.0047	-0.0003	0.0490 **	0.0129	0.0006
Daily total origination airport flights (100's)	-0.0106 *	0.0043	-0.0012	0.0020	0.0137	0.0001
Daily total destination airport flights (100's)	-0.0152 **	0.0045	-0.0017	-0.0012	0.0143	0.0001
Weather Variables						
Rain origination (hundredths of an inch)	-0.0055 **	0.0004	-0.0006	0.0035 **	0.0009	0.0001
Rain destination (hundredths of an inch)	-0.0040 **	0.0004	-0.0004	0.0007	0.0010	0.0000
Minimum temperature origination	-0.0012	0.0013	-0.0001	0.0001	0.0045	0.0000
Minimum temperature destination	-0.0032 *	0.0013	-0.0004	-0.0007	0.0045	0.0000
Frozen precipitation origination (hundredths)	-0.0144 **	0.0024	-0.0016	0.0125 **	0.0045	0.0003
Frozen precipitation destination (hundredths)	-0.0110 **	0.0026	-0.0012	0.0184 **	0.0041	0.0003
Constant	4.7546 **	0.2780		3.0366 **	0.6231	
Log-likelihood	-20,299					
Observations	146,589					

Note: Regressions include indicator variables for carrier, day of week and month. Marginal effects are the probability that the average flight is ontime or canceled. * and ** indicate 5% and 1% significance levels.

Table IV: Nested Logit estimations of flight operation decisions for 1% sample of all 2002 U.S. domestic flights

Model (2) - Potential revenue	Ontime			Cancel		
	Coeff	Std Error	Marg.Eff.	Coeff	Std Error	Marg.Eff.
Economic Variables						
Potential revenue (in $10,000s)	-0.0017	0.0195	-0.0001	0.1340 *	0.0647	0.0014
Route load capacity (monthly average)	-0.4004	0.2325	-0.0485	-5.4616 **	0.4645	-0.0534
Seating capacity of aircraft (in 100's)	0.0097	0.0603	0.0007	-0.6250 **	0.2343	-0.0067
Seating capacity*hub destination (in 100's)	-0.1046	0.0565	-0.0120	-0.2970	0.1986	-0.0020
Route Competition Variables						
Monopoly	0.1588 **	0.0557	0.0178	-0.2105	0.1722	-0.0039
Large duopoly carrier	0.0893	0.0513	0.0100	-0.0850	0.1657	-0.0018
Small duopoly carrier	0.0633	0.0617	0.0069	-0.3971 *	0.1914	-0.0049
Airport Competition Variables						
Airport concentration at origination	-0.2625 **	0.0725	-0.0296	0.1485	0.2235	0.0043
Airport concentration at destination	-0.0222	0.0707	-0.0025	-0.5736	0.6245	-0.0052
Airline hub origination	-0.1890 **	0.0425	-0.0215	-0.1950	0.1616	-0.0001
Airline hub destination	0.1475	0.1003	0.0170	0.5995	0.3139	0.0048
Logistical Variables						
Distance (in 100's miles)	0.0025	0.0039	0.0003	-0.0108	0.0146	-0.0001
Hours until next flight	-0.0204 **	0.0055	-0.0023	-0.0384	0.0211	-0.0002
Renormed scheduled departure time	-2.6695 **	0.1165	-0.3024	-1.5839 **	0.3105	0.0116
Last flight of day	0.4669 **	0.0757	0.0528	0.1877	0.2713	-0.0030
Daily total carrier flights on route	-0.0031	0.0047	-0.0003	0.0489 **	0.0129	0.0005
Daily total origination airport flights (100's)	-0.0108 *	0.0043	-0.0012	0.0021	0.0137	0.0001
Daily total destination airport flights (100's)	-0.0154 **	0.0045	-0.0017	-0.0010	0.0143	0.0002
Weather Variables						
Rain origination (hundredths of an inch)	-0.0055 **	0.0004	-0.0006	0.0035 **	0.0009	0.0001
Rain destination (hundredths of an inch)	-0.0040 **	0.0004	-0.0004	0.0007	0.0010	0.0000
Minimum temperature origination	-0.0012	0.0013	-0.0001	0.0003	0.0045	0.0000
Minimum temperature destination	-0.0032 *	0.0013	-0.0004	-0.0007	0.0045	0.0000
Frozen precipitation origination (hundredths)	-0.0145 **	0.0024	-0.0016	0.0126 **	0.0045	0.0003
Frozen precipitation destination (hundredths)	-0.0112 **	0.0026	-0.0012	0.0183 **	0.0042	0.0003
Constant	4.7205 **	0.2698	0.5348	2.7056 **	0.6083	-0.0214
Log-likelihood	-20,299					
Observations	146,589					

Note: Regressions include indicator variables for carrier, day of week and month. Marginal effects are the probability that the average flight is ontime or canceled. * and ** indicate 5% and 1% significance levels.

Table V: Nested Logit estimations of flight operation decisions for 1% sample of all 2002 U.S. domestic flights

Model (3) - Yield	Ontime			Cancel		
	Coeff	Std Error	Marg.Eff.	Coeff	Std Error	Marg.Eff.
Economic Variables						
Yield (average revenue passenger mile)	0.3131 **	0.0699	0.0353	0.0195	0.1871	-0.0031
Route load capacity (monthly average)	-0.4003	0.2331	-0.0306	-5.6628 **	0.4617	-0.0588
Seating capacity of aircraft (in 100's)	-0.0027	0.0497	0.0008	-0.4291 *	0.2035	-0.0047
Seating capacity*hub destination (in 100's)	-0.1100 *	0.0557	-0.0118	-0.2300	0.1999	-0.0014
Route Competition Variables						
Monopoly	0.1260 *	0.0567	0.0148	-0.2100	0.1700	-0.0037
Large duopoly carrier	0.0821	0.0522	0.0096	-0.1376	0.1620	-0.0024
Small duopoly carrier	0.0602	0.0627	0.0080	-0.4534 *	0.1885	-0.0057
Airport Competition Variables						
Airport concentration at origination	-0.2671 **	0.0729	-0.0307	0.1953	0.2204	0.0050
Airport concentration at destination	-0.0421	0.0709	-0.0050	0.0728	0.2236	0.0013
Airline hub origination	-0.1965 **	0.0417	-0.0219	-0.1079	0.1564	0.0009
Airline hub destination	0.1550	0.1000	0.0160	0.5722	0.3136	0.0047
Logistical Variables						
Distance (in 100's miles)	0.0092 *	0.0036	0.0010	0.0055	0.0134	0.0000
Hours until next flight	-0.0220 **	0.0055	-0.0024	-0.0425 *	0.0209	-0.0002
Renormed scheduled departure time	-2.7412 **	0.1190	-0.3053	-1.5969 **	0.3029	0.0112
Last flight of day	0.4856 **	0.0761	0.0542	0.2350	0.2690	-0.0025
Daily total carrier flights on route	-0.0029	0.0047	-0.0005	0.0477 **	0.0129	0.0006
Daily total origination airport flights (100's)	-0.0100 *	0.0043	-0.0011	0.0002	0.0136	0.0001
Daily total destination airport flights (100's)	-0.0145 **	0.0046	-0.0016	-0.0016	0.0142	0.0001
Weather Variables						
Rain origination (hundredths of an inch)	-0.5384 **	0.0367	-0.0616	0.3180 **	0.0808	0.0092
Rain destination (hundredths of an inch)	-0.3953 **	0.0355	-0.0448	0.0822	0.0953	0.0051
Minimum temperature origination	-0.0012	0.0013	-0.0001	-0.0002	0.0045	0.0000
Minimum temperature destination	-0.0032 *	0.0013	-0.0004	-0.0014	0.0045	0.0000
Frozen precipitation origination (hundredths)	-1.3800 **	0.2378	-0.1588	1.1877 **	0.4502	0.0278
Frozen precipitation destination (hundredths)	-1.0101 **	0.2705	-0.1188	1.8538 **	0.4082	0.0313
Constant	4.6782 **	0.2784	0.5209	2.8133 **	0.6114	-0.0182
Log-likelihood	-20,290					
Observations	146,589					

Note: Regressions include indicator variables for carrier, day of week and month. Marginal effects are the probability that the average flight is ontime or canceled. * and ** indicate 5% and 1% significance levels.

Table VI: Nested Logit estimations of flight operation decisions for 1% sample of all 2002 U.S. domestic flights

Model (4) - Airline hub size	Ontime			Cancel		
	Coeff	Std Error	Marg.Eff.	Coeff	Std Error	Marg.Eff.
Economic Variables						
Average revenue (in $10,000s)	0.0332	0.0295	0.0036	0.1640	0.1050	0.0014
Route load capacity (monthly average)	-0.4696 *	0.2312	-0.0483	-5.8608 **	0.4666	-0.0581
Seating capacity of aircraft (in 100's)	-0.0493	0.0648	-0.0052	-0.5150 *	0.2525	-0.0050
Seating capacity*small airline hub dest.	-0.0780	0.0754	-0.0086	-0.2362	0.2679	-0.0017
Seating capacity*medium airline hub dest.	-0.0563	0.0988	-0.0063	-0.0698	0.3891	-0.0002
Seating capacity*large airline hub dest.	-0.1448	0.1051	-0.0153	-1.2583 *	0.4942	-0.0120
Route Competition Variables						
Monopoly	0.2158 **	0.0598	0.0246	-0.2870	0.1840	-0.0054
Large duopoly carrier	0.1240 *	0.0532	0.0141	-0.1592	0.1682	-0.0030
Small duopoly carrier	0.0770	0.0622	0.0090	-0.4281 *	0.1905	-0.0054
Airport Competition Variables						
Airport concentration at origination	-0.2319 **	0.0759	-0.0262	0.0915	0.2279	0.0034
Airport concentration at destination	-0.0017	0.0733	-0.0002	0.0225	0.2339	0.0003
Small airline hub origination	-0.1509 **	0.0483	-0.0169	-0.2179	0.1774	-0.0008
Medium airline hub origination	-0.3628 **	0.0643	-0.0410	0.0628	0.2196	0.0045
Large airline hub origination	-0.3403 **	0.0765	-0.0385	0.0725	0.2897	0.0044
Small airline hub destination	0.1286	0.1216	0.0141	0.4810	0.3832	0.0038
Medium airline hub destination	-0.1078	0.1880	-0.0126	0.4968	0.6930	0.0065
Large airline hub destination	0.1043	0.1988	0.0100	2.2720 **	0.7957	0.0234
Logistical Variables						
Distance (in 100's miles)	-0.0021	0.0041	-0.0002	-0.0060	0.0153	0.0000
Hours until next flight	-0.0214 **	0.0055	-0.0024	-0.0376	0.0212	-0.0002
Renormed scheduled departure time	-2.6995 **	0.1152	-0.3035	-1.5845 **	0.3100	0.0115
Last flight of day	0.4783 **	0.0759	0.0539	0.1818	0.2722	-0.0031
Daily total carrier flights on route	-0.0029	0.0046	-0.0004	0.0479 **	0.0127	0.0005
Daily total origination airport flights (100's)	-0.0049	0.0047	-0.0006	-0.0054	0.0146	0.0000
Daily total destination airport flights (100's)	-0.0094	0.0050	-0.0011	-0.0052	0.0159	0.0000
Weather Variables						
Rain origination (hundredths of an inch)	-0.5439 **	0.0372	-0.0617	0.3418 **	0.0860	0.0094
Rain destination (hundredths of an inch)	-0.3961 **	0.0353	-0.0448	0.0691	0.0975	0.0049
Minimum temperature origination	-0.0017	0.0013	-0.0002	-0.0001	0.0046	0.0000
Minimum temperature destination	-0.0035 **	0.0013	-0.0004	0.0004	0.0045	0.0000
Frozen precipitation origination (hundredths)	-1.4054 **	0.2377	-0.1596	1.2002 **	0.4489	0.0278
Frozen precipitation destination (hundredths)	-1.0570 **	0.2624	-0.1208	1.8457 **	0.4134	0.0311
Constant	2.7863 **	0.2751	0.5489	2.7863 **	0.6467	-0.0216
Log-likelihood	-20,289					
Observations	146,589					

Note: Regressions include indicator variables for carrier, day of week and month. Marginal effects are the probability that the average flight is ontime or canceled. * and ** indicate 5% and 1% significance levels.