PREDICTING AIR TRAVELERS' NO-SHOW AND STANDBY BEHAVIOR USING PASSENGER AND DIRECTIONAL ITINERARY INFORMATION

Laurie A. Garrow, Ph.D.*
Georgia Institute of Technology
School of Civil and Environmental Engineering
790 Atlantic Drive
Atlanta, GA 30322-0355
Ph: (404) 385-6565
Fax: (404) 894-2278
Email: Laurie.Garrow@ce.gatech.edu

Frank S. Koppelman, Ph.D.
Department of Civil and Environmental Engineering
Robert R. McCormick School of Engineering and Applied Science
Northwestern University
2145 Sheridan Road
Evanston, IL 60208-3109
Ph: (847) 491-8794
Fax: (847) 491-4011
Email: f-koppelman@northwestern.edu

Paper accepted for publication in the Journal of Air Transport Management.

June 23, 2004

* Corresponding author.

Abstract

This is the first study of airline travelers' no-show and standby behavior based on passenger and directional outbound/inbound itinerary data. The paper describes passengers' behavior based on estimation of a multinomial logit model for domestic U.S. itineraries departing in March 2001 or March 2002. This enables us to explore behavioral differences based on passenger and itinerary characteristics as well as identify differences in rescheduling behavior occurring after September 11, 2001. Benefits of using passenger data to improve forecasting accuracy and support a broad range of managerial decisions are described.

Keywords: No-show; Air-travel demand forecasting; Revealed preference rescheduling behavior

1. Background and motivation

Predicting passengers' no-show and standby behavior is an important component of airlines' profitability and revenue management. Since all passengers who book do not actually travel, airlines overbook to reduce the expected number of empty seats on a flight when there is demand for those seats. Cancellation and no-show models are used to forecast the expected number of passengers booked on a specific flight who will not travel on that flight. Cancellation models predict how many passengers inform the airline they do not intend to travel prior to the departure of their flights. No-show models estimate the number of remaining booked passengers, *i.e.*, passengers who have not cancelled, but fail to show for their flights.

Distinct from other transportation modes, the use of disaggregate passenger information in forecasting models has only recently been adopted for use in the study of air passenger behavior. There are only two published studies we are aware of that use passenger data to forecast no-show rates (Kalka & Weber, 2000; Pastor, 2000). Currently, many airlines forecast no-show rates using time-series models based on historical booking class or cabin no-show rates. These models consider differences in no-show rates due to flight-specific attributes such as departure time, day of week, month, capacity, origin, destination, etc. (Ruppenthal & Toh, 1983). To the extent that different types of passengers and itineraries exhibit distinct no-show and standby rates, current airline models based on historic cabin or booking class no-show rates cannot make accurate predictions when the underlying passenger and/or itinerary mix changes. Even modest improvements in no-show forecasts can translate to millions of dollars in annual revenue for a major U.S. carrier.

In addition to improving forecasting accuracy and generating incremental revenue, no-show models that incorporate passenger and itinerary information can help support a broad range of managerial decisions. For example, by identifying which passengers are more likely to divert to a competitor during irregular operations, revenue losses associated with departure delays can be quantified. By identifying which passengers are more likely to standby for different flights, airport security staffing levels required to screen standby passengers exhibiting particular characteristics can be determined. While the spectrum of business insights that can be explored is broad, the focus of this paper is to understand how day of departure rescheduling behavior as reflected in no-show and standby rates differs by passenger and itinerary types.

2. Conceptual Model

This section presents the conceptual framework used to study air travelers' rescheduling behavior. The conceptual model considers factors the airline industry considers important in addition to factors typically considered in the travel demand literature. Because the study is based on actual data from a major U.S. carrier, automated data cleaning routines the carrier uses that influence the interpretation of results are also included. Only day of departure rescheduling behavior is considered; rescheduling reflected prior to the itinerary's day of departure and captured in cancellation data is excluded from this analysis.

Information about factors the airline industry considers to be important in explaining no-show behavior can be obtained by looking at ways in which airlines tried to modify no-show behavior over the years. One form of intentional no-shows occurs when passengers make multiple bookings on one or more airlines (Ruppenthal & Toh, 1983; Scholin, 1983). Currently, some airlines try to prevent multiple bookings by using ticketing restrictions on low-yield bookings that require the passenger to pay for each reservation made (Pina, 1985). Some airlines also impose an exchange fee to redeem unused low-yield tickets (Anonymous, 1981) and issue refunds on unused tickets only if the passenger cancels in advance (American Airlines, 2003). In addition, when a passenger no-shows for the outbound itinerary, some airlines use automated programs to cancel the inbound itinerary (Scholin, 1983). Historically, airlines have tried many other policies like securing a credit card number at the time of reservation and charging a no-show penalty for failure to show (Anonymous, 1981). However, these business strategies do not prevent intentional no-shows that result from irrationalities in the carrier's prices. For example, many low-yield fares require a Saturday night stay. In some markets, it may be cheaper to buy two low-yield round trips than one high-yield round trip. In this case, the return portion of both low-yield round trips becomes a no-show[1].

The conceptual model of airline passengers' day of departure rescheduling behavior incorporates factors the airline industry considers to be important as well as factors from the travel demand literature (*e.g.*, the ability of individuals to reschedule depends on the available transportation supply). As shown in Figure 1, the initial scheduling activity occurs when the passenger makes a booking reservation. Day of departure rescheduling decisions for outbound and inbound itineraries are reflected in the no-show and standby alternatives. A standby is defined as a passenger who voluntarily takes a different itinerary with the carrier of interest. Standbys are further divided into those who wait at an airport hoping to take an earlier flight and the late standby who accepts a later flight as a result of missing his/her scheduled flight. Passengers who divert to competitor airlines select the no-show alternative. However, not all

[1] As an example, consider a passenger that wants to travel from Chicago to Boston on Monday and return on Tuesday. The passenger could buy two tickets: one departing from Chicago on Monday and "returning" after the following Saturday and one departing from Boston on Tuesday and "returning" after the following Saturday. Consequently, the return portion of both tickets becomes a no-show.

passengers who no-show divert to competitor airlines; as noted in the literature, no-shows can be caused for a variety of reasons and be a reflection of both the initial scheduling activity (*e.g.*, making multiple booking reservations on one or more carriers) and the day of departure rescheduling activity (*e.g.*, diverting to a competitor).

As depicted in Figure 1, four categories are used to classify factors influencing the initial booking reservation and outbound and inbound travel decisions. They describe traveler characteristics, familiarity with the air transportation system, availability of viable alternatives, and trip characteristics. The categories overlap in the sense that traveler characteristics such as price insensitivity and frequent flyer status provide improved access to standby opportunities on the carrier of interest and competitors. Specifically, passengers with high-yield, unrestricted tickets can use those tickets for travel on most competitors. Similarly, elite members of a carrier's frequent flyer program with boarding priority over general and non-members have more access to standby opportunities.

[Insert Figure 1 about here]

In addition to providing access to standby opportunities, traveler characteristics like price sensitivity can influence the original scheduling activity. Price-sensitive passengers buying low-yield fares may search for and make multiple bookings on different carriers. These low-yield fares usually require ticketing, *i.e.*, must be purchased, within 24 or 48 hours after the booking is made. Many carriers use automated data cleaning routines that seek to remove bookings that have exceeded the ticketing time limit, however, due to the complexity of fare rules, some of these speculative low-yield bookings remain and are reflected in no-show rates. Alternatively, price-insensitive passengers may make multiple bookings on the same or different carriers using high-yield fares that do not need to be purchased until departure. This source of speculative bookings is not automatically cleaned from the data for outbound travel. The same is not true for inbound itineraries, *i.e.*, carriers may use an additional data cleaning routine to cancel the return portion of an itinerary when the passenger no-shows on the outbound portion.

The original booking and actual travel decisions of price-sensitive travelers may also be influenced by the availability of lower fares for non-preferred itineraries. Specifically, price-sensitive travelers may book an inferior, lower-priced itinerary with the intention of standing by for a more preferred itinerary with the same carrier. However, since some low cost and major carriers currently charge a standby fee, price-sensitive and/or infrequent travelers may have the perception that the same policy exists on all airlines. Thus, it is not clear if the availability of lower fares offered on less desirable itineraries at the time of booking influences a passenger's booking and travel decisions, specifically the passenger's intention to standby for a more preferred itinerary.

The second category used to describe air passengers' rescheduling behavior is familiarity of the air transportation system. Passengers who travel often and are members of one or more frequent flyer programs are likely to be more familiar with the air transportation system. These passengers may be more aware of unpublicized business practices, like the ability to use high-yield unrestricted tickets on competitors, and ever-evolving security procedures implemented after September 11, 2001, that influence passengers' ability to standby on the carrier or divert to competitors.

The third category used to describe air passengers' rescheduling behavior is the availability of viable alternatives. Specifically, it is assumed that passengers only standby for itineraries that have the same or higher level of service as the original itinerary. In addition, only itineraries arriving at the destination on the same day are considered, *i.e.*, rescheduling decisions occurring over multiple days are excluded from analysis.

Finally, trip characteristics including the rigidity of time constraints and uncertainty in the duration of the trip can influence rescheduling behavior. For example, a business traveler meeting a client early in the morning is probably less likely to late standby on the outbound itinerary. However, the same business traveler may be more likely to early standby on the inbound itinerary if the meeting with the client concludes earlier than anticipated.

In summary, it is important to highlight that the conceptual model used to study airline passengers' no-show and standby behavior is able to explore a broad range of analysis questions *because of the availability of disaggregate passenger data*. Indeed, only by using passenger-level data can we define standby choices, identify passengers' outbound and inbound itineraries, and explore behavioral differences due to traveler characteristics such as frequent flyer status. Empirical results offer a rich behavioral interpretation of passengers' rescheduling behavior and explore differences based on passenger and directional itinerary information derived from disaggregate data.

3. Data description and analysis methodology

Airline passengers' day of departure rescheduling behavior is based on variables derived from proprietary data sources from a major U.S. airline. The data is based on the carrier's actual booking, ticketing, flight/itinerary schedule, frequent flier membership class, and check-in information. The analysis uses data for continental U.S. itineraries departing in March 2001 and March 2002. March was selected because historically it experiences an average or lower than average monthly load factor. This results in fewer censored data points, *i.e.*, passengers who attempt to standby but are unable to do so due to no seat availability on competing itineraries. In addition, March was considered to be far enough after the events of September 11, 2001, so as not to be contaminated with short-term effects due to the terrorist attacks.

Four datasets, or segments, are used for this study. The datasets are distinguished by whether they are outbound or inbound itineraries and by whether they departed in March 2001 or March 2002. One-way itineraries are excluded from the analysis. Likelihood ratio tests were conducted to validate the need for separate outbound and inbound models. Specifically, a "restricted" model that combined March 2001 outbound and inbound data was compared to two "unrestricted" models that used data from only one segment. The likelihood ratio test, formally expressed as

$$-2[LL(\beta_{restricted} - \sum_{segments} LL(\beta_{unrestricted})] \sim \chi^2_{Number\ of\ restrictions}, \tag{1}$$

rejects the null hypothesis that the restricted model is the true model at the 0.005 level $\left(\hat{\chi}^2 = 149 >> \chi^2_{34,0.005} = 59\right)$. Similarly, a test of outbound/inbound segments on March 2002 also rejects the null hypothesis $\left(\hat{\chi}^2 = 164 >> \chi^2_{34,0.005} = 59\right)$.

Because the carriers' actual data contains millions of monthly booking transactions, a choice-based sample was selected with approximately equal choice frequencies for each of the four alternatives; show, early standby, late standby and no-show. The total number of observations in each dataset ranges from 1,791 to 3,674[2]. Population choice probabilities for show, no-show, early standby and late standby for the March 2001 outbound segment are 92.7%,

[2] The number of observations differs for each segment due to the underlying sampling process and data cleaning procedures.

6.3%, 0.89% and 0.13%, respectively. Similarly, population choice probabilities for the March 2001 inbound segment are 86.9%, 10.0%, 2.99%, and 0.13%, respectively.

Multinomial logit (MNL) models are used to estimate the probability that each passenger will show, no-show, or standby for an earlier or later itinerary with the carrier of interest. We assume the show and no-show alternatives appear in every passenger's choice set while the early (late) standby alternative is included in the choice set only if there is at least one itinerary of the same or higher level of service that arrives earlier (later) than the passengers' original booked itinerary. Exogenous sampling maximum likelihood (ESML) estimators are used, which for choice-based MNL models with a full set of alternative specific constants produce consistent estimates of all parameters except for the alternative specific constants; however, since population rates are known, consistent estimators can be recovered by subtracting log (sample rate/population rate) from the estimated constants. McFadden's proof of this property is reported in Manski and Lerman (1977). Because the MNL is known to have limitations, *e.g.*, it imposes restrictive substitution patterns across alternatives; more advanced nested logit (NL) models were also estimated. We show, in a separate analysis, that the ESML estimator can also be used with NL models for choice-based samples and that consistent estimators can be recovered by subtracting $\mu_m \times$ log (sample rate / population rate) for alternative i from the estimated constants where μ_m is the logsum coefficient associated with nest m, $0 < \mu_m \leq 1$, $i \in m$. See Garrow (2004) for more information. Because the behavioral interpretations are similar for the MNL and NL models, only MNL results are discussed in this paper.

An iterative modeling approach guided by judgment and statistics was used to find a preferred model specification; more than 250 different utility specifications were estimated during this process. Variables that were consistently insignificant across the four segments and/or that had counter-intuitive signs are excluded from the final specification. The final model specification includes flight, itinerary, and passenger variables that are described in detail in the next section.

Before presenting empirical results, it should be emphasized that the data used for the study is derived from a single major U.S. carrier. Thus, the results presented here are more likely to be applicable to large network carriers than to low cost or international carriers. Nonetheless, the methodological contribution of using passenger data to predict no-show behavior provides insight into the broad range of managerial and analysis questions that can be answered. To the extent that these factors influence no-show rates and vary across a carrier's flights, the forecasting benefit of using passenger and directional itinerary information can be generalized to other carriers.

4. Empirical results

The presentation of model results is divided into three sections. First, an interpretation is presented using the March 2001 outbound segment. This provides a fundamental understanding of how flight, itinerary, and passenger variables influence the probability a passenger will show, no-show, or standby for an earlier or later itinerary. The next two sections expand on this foundation by, first, evaluating differences in outbound versus inbound models and, second, by estimation and comparison of March 2001 versus March 2002 models. Table 1 summarizes model results for all four segments and is referred to throughout this section.

4.1 Base model interpretation: March 2001 outbound segment

4.1.1 Flight/itinerary variables

Flight/itinerary variables include dummy variables for the itinerary's departure day of week, departure time period, and duration. Also included are carrier capacity and carrier presence measured at the itinerary's departure city. The parameter estimates and reference levels associated with each of these variables is shown in Table 1.

The positive day of week parameter of 0.26 associated with the early standby alternative relative to other choices and days of the week indicates that passengers are more likely to early standby for outbound itineraries late in the work week. Day of week was not found to be statistically significant in predicting differences among show, no-show, and late standby choices. This finding is of particular significance because many analysts managing inventory levels on flights believe day of week to be one of the most important drivers of no-show behavior (Garrow, 2001). However, when disaggregate, passenger-level data is used, the importance of day of week diminishes. This finding confirms an earlier study of no-show behavior by Kalka and Weber (2000).

Departure time variables are defined for four time periods: itineraries departing from 6 – 9 a.m., 9:01 a.m. – 4 p.m., 4:01 – 7 p.m., and after 7 p.m. (7:01 p.m. – midnight).[3] Departure time variables associated with the no-show and early standby alternatives are statistically significant. Passengers are more likely to no-show for flights departing very early in the morning (before 9 a.m.) and late in the evening (after 7 p.m.). Departure time coefficients

[3] No flights depart between midnight and 6 a.m.

associated with the early standby alternative are all negative relative to the reference time period of 7:01 p.m.-midnight. In addition, these coefficients are larger in absolute magnitude for itineraries departing earlier in the day. Because the early standby alternative is included in the choice set only when at least one viable alternative is available, this reflects a strong resistance of passengers to early standby earlier in the day.[4] Standby results are particularly important given that the check-in data provided for this study and used in the airline's current no-show model often marks these standby passengers as no-shows; as part of this study, we matched apparent no-shows with boardings on earlier (later) flights to identify early (late) standbys. Elimination of this classification error eliminates inflated no-show rates that are sensitive to load factors on the carrier's earlier itineraries. This sensitivity, for early standby, is greatest for flights departing later in the day.

The relationship between early standbys and denied boardings in the carrier's current no-show model is best clarified via an example. Figure 2 depicts a hypothetical flight with capacity for 90 passengers. We assume that there are 100 bookings for this flight of which 90 show, five no-show, and five successfully standby for an earlier flight resulting in a full flight and no denied boardings. Because early standbys are classified as no-shows in the carrier's current model, 10% of future bookings made on a similar flight are expected to no-show leading to a full cabin with no denied boardings. However, if the earlier flights are full and there are no seats available for standby passengers from the hypothetical flight, 95 passengers will "show" for the future flight, five of which will become denied boardings.

[4] Note this interpretation changes due to censoring effects if a passenger attempted to standby for an earlier itinerary but was unsuccessful because there were no available seats. This effect is considered negligible because of the small incidence of flights departing full – approximately 7.8% in March 2001 and 6.4% in March 2002.

[Insert Figure 2 about here]

Thus, when a carrier's load factors suddenly increase[5] in a market, not only is the carrier more likely to see denied boardings, but these denied boardings are most likely to occur on flights departing later in the day. These denied boardings are particular costly to airlines because a later itinerary may not be available and the airline must pay for overnight accommodations and/or pay a competitor airline to accommodate the denied boarding passengers (usually at the high-yield coach price). By distinguishing early standbys from no-show passengers, the sensitivity of current no-show forecasts to load factors on the carriers' competing itineraries can be accounted for and over-bookings can be reduced to avoid a high level of denied boardings.

Itinerary duration variables associated with the early standby alternative are defined for three intervals: itineraries less than 180 minutes, itineraries between 181 and 300 minutes, and itineraries greater than 300 minutes. These variables may be interpreted as differential early standby alternative specific constants for distinct time periods. Duration coefficients become more negative as duration increases, which indicates that passengers are less likely to early standby for longer itineraries. Intuitively, this makes sense because passengers on longer itineraries have a smaller amount of time available in the day for non-travel activities and may have less flexibility in rescheduling travel and non-travel activities.

Carrier capacity variables measure the number of seats offered by the carrier for itineraries that are scheduled to depart in the defined periods before/after the booked itinerary and have a level of service equivalent to or higher than the level of service of the booked

[5] It is the sudden shift in load factors that leads to denied boarding problems. This is because gradual shifts in load factors can be captured in current airline no-show models that are based on time-series models.

itinerary. Carrier capacity coefficients are positive for itineraries arriving at their destination within five hours of the booked itinerary. Coefficients associated with the early standby alternative tend to be larger in magnitude than those associated with the late standby alternative for each time period. This is consistent with a conservative, risk-averse decision strategy on the part of the traveler[6]. Further, the value of carrier capacity associated with the early standby alternative decreases as the amount of time between the original booked itinerary and an earlier itinerary increases while the effect of later carrier capacity is relatively constant for the late standby alternative. This reflects a higher degree of control for passengers making early standby versus late standby choices; passengers who arrive at the airport early can retain their seat on their original flight while standing by for an earlier flight, whereas passengers who arrive late and fail to rebook no longer have a confirmed seat and are subject to seat availability on later flights, the show behavior of booked passengers on those flights, and the carrier's standby priority rules.

Departure city presence variables measure the carrier's market presence as a function of the number of flights to all destinations at the itinerary's departure city. Market presence increases the probability a passenger will standby when the carrier of interest has dominant market share in the departure city[7]. Formally, departure city presence is modeled, in terms of the number of flights, as:

$$\beta_1\left(\frac{Market\ share_{carrier\ of\ interest}}{Market\ share_{top\ competitor}}\right)^{\alpha} \text{ if Market Share}_{\text{carrier}} > \text{Market Share}_{\text{top competitor}} \qquad (2)$$

[6] At the time the data was collected, there were no standby fees. Thus, there is no risk in standing by for an earlier flight but a potentially large risk in standing by for a later one. The latter occurs because the airline is not obligated to accommodate passengers that fail to check-in before their original itinerary's departure.

[7] Market presence was found to be statistically insignificant when the carrier did not have dominant market share.

where β_1 is the estimated parameter reported in Table 1 and α is 0.5. Parametric values for α were estimated over the range of 0.4 to 2.0 and 0.5 was found to give the most significant results. The departure city presence beta coefficients associated with early and late standby alternatives are significant (0.14 and 0.33, respectively) indicating passengers are more likely to standby for earlier or later itineraries if booked on the dominant carrier for the departure city. The parameter for the late standby alternative is twice as large as that for the early standby alternative, which means that market presence has a much stronger effect in predicting late standby behavior (by reducing the risk of no available flights) than early standby behavior.

[Insert Table 1 about here]

4.1.2 Passenger variables

Passenger variables include dummy variables for e-ticket purchases, booking class, frequent flyer status, and group size. As shown in Table 1 for the March 2001 outbound segment, e-ticket is a very powerful predictor of no-show rates because it helps discriminate among speculative and confirmed bookings; bookings that are not e-tickets have either not been paid for or have been paid for and confirmed via another purchase medium like paper tickets. In the outbound March 2001 segment, the parameter associated with the no-show e-ticket variable is –1.81, indicating passengers with e-tickets are much less likely to no-show than passengers without e-tickets.

Booking class variables indicate whether the passenger booked a first or business class fare, a high yield coach fare, or a low-yield coach fare. Passengers traveling in first and business classes are less likely to no-show and less likely to standby for earlier or later itineraries. These passengers may be less willing to standby because they might have to travel in a lower-premium cabin. This is particularly relevant for passengers booked in business class in trans-continental U.S. markets. Since these markets are not served exclusively with three-cabin planes, the passenger's standby itineraries may be on flights that do not have a business cabin.

Frequent flyer coefficients reveal whether the passenger is an elite or general member of the carrier's frequent flyer program. Frequent flyer members, who are more loyal to the carrier than non-members, are less likely to no-show and are more likely to early standby as compared to non-members. Frequent flyer members may be able to take advantage of more early standby opportunities than non-members because they receive boarding priority.

Finally, group variables indicate whether the passenger is traveling with at least one other person. Passengers traveling in groups are more likely to show than passengers traveling alone. Empirically, the group variables associated with the no-show, early standby, and late standby alternatives are all negative (-0.60, -0.71, and -0.50, respectively) relative to the show alternative.

4.2 Outbound and inbound differences: March 2001 segments

Given the preceding analysis of how flight/itinerary and passenger variables influence choice rates, this section focuses on the similarities and differences between outbound and inbound itineraries. A comparison of population rates for March 2001 outbound and inbound segments reveals that early standbys occur more than three times as often on inbound itineraries (3.0%) than outbound ones (0.89%). Late standby rates are identical (0.13%) for both outbound and inbound segments. Counter to intuition, no-show rates are higher for inbound itineraries

(10.0%) than outbound itineraries (6.2%). This finding is somewhat surprising given that if a passenger no-shows on the outbound itinerary, the carrier may automatically cancel the inbound itinerary (thereby deleting the inbound record from the dataset). However, similar to increases in early standby rates, no-show rate increases can reflect a greater amount of rescheduling activity for inbound itineraries. As described in the conceptual model, no-show rates will increase if more passengers with tickets purchased for travel on the carrier of interest use their tickets to fly on competitors[8]. They will also increase if passengers fail to rebook their return itinerary for a later departure date prior to the time their original booked itinerary departs. Finally, no-show rates on inbound itineraries can reflect irrationalities in the carriers' pricing system, causing passengers to purchase two sets of low fare tickets and using only the outbound portion of each set, as described earlier.

Pairwise comparisons of outbound and inbound models were conducted. Coefficients for those variables that are significantly different between the models at the 0.02 level are highlighted and include duration, departure time, e-ticket, and booking class variables. All other coefficients are not significantly different and, in general, the magnitude of the differences is relatively small. That is, the following patterns, already discussed for the outbound itineraries, are observed for inbound itineraries as well.

Formally, the probability individual *n* selects alternative *i* in a MNL model is:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \tag{3}$$

[8] Passengers with paper tickets for high-yield coach, business, and first class fares can use these tickets to fly a competitor airline. Passengers with low-yield coach fares must have their tickets reissued for a high-yield fare class before being able to fly a competitor airline. This is done only during irregular operations or for denied boardings when the carrier cannot accommodate the passengers on its own itineraries.

Comparisons between coefficients for the inbound and outbound models should be based on the entire vector of parameters for each model and the corresponding covariance matrices. However, it more intuitive and provides better behavioral insights to compare common parameters across different datasets[9].

As seen in Table 1, inbound passengers are more likely than outbound passengers to early standby for itineraries that are less than 180 minutes. In addition, early standby behavior is more likely to occur during the 4:01 – 7 p.m. departure time interval, particularly on inbound itineraries. This is consistent with a risk-averse decision strategy in which passengers book return flights later in the day and/or during the evening peak period and standby for earlier flights if their meetings or other trip activities conclude early. Also, passengers without e-tickets are less likely to no-show (more likely to show) on inbound itineraries than outbound itineraries. The knowledge that a passenger has an e-ticket may be less important for inbound itineraries since un-ticketed bookings that were not automatically cancelled when they exceeded the ticketing time limit can have the inbound itinerary automatically cancelled when the passenger no-shows on the outbound itinerary. Since the e-ticket variable helps discriminate between confirmed and speculative bookings, similar results are expected for airlines that use automated data cleaning processes to cancel the return itinerary when passengers no-show on the outbound itinerary. Finally, high yield passengers are more likely to early and late standby on inbound itineraries. To the extent that high yield passengers are business travelers, this reflects a higher degree of uncertainty in when meetings or other trip activities conclude for business travelers.

[9] A more complex approach based on comparison in choice probabilities was used to verify that differences in parameter estimates translated into differences in choice probabilities. See Garrow (2004) for details.

Additional insight is provided by computing average choice probabilities for a hypothetical passenger and market. Table 2 contains these choice probability forecasts using March 2001 data for a hypothetical passenger who has a low-yield booking, is a general member of the carrier's frequent flyer program, is traveling alone, is traveling on an itinerary that departs on Saturday through Tuesday with a flight time between 181 and 300 minutes, is booked on a carrier that does not have dominant market share in the departure city but has other itineraries, one per arrival and departure time period with seat capacity of 120 seats. These choice probabilities are compared along three dimensions: e-ticket, departure time intervals, and outbound/inbound itineraries. As can be seen in Table 2, passengers with e-tickets are more likely to show and this effect is stronger for outbound than inbound itineraries. In addition, the later in the day an itinerary departs, the more likely a passenger is to early standby. The effect of departure time is stronger for inbound itineraries, particularly for the 4:01-7 p.m. departures, *e.g.*, the e-ticket inbound itinerary early standby rates are 0.77%, 1.60%, 3.33%, and 3.63%, respectively, and are 1.9, 3.0, 3.5 and 2.2 times greater than those for the e-ticket outbound itinerary, respectively.

[Insert Table 2 about here]

4.3 Effects of 9/11: March 2001 and March 2002 segments

This section focuses on the similarities and differences between March 2001 and March 2002 itineraries. Likelihood ratio tests were conducted to validate the need for separate models. Specifically, a "restricted" model that combined March 2001 and March 2002 outbound data was compared to two "unrestricted" models that used data from only one segment. The likelihood ratio test rejects the null hypothesis that the restricted model is the true model at the 0.005 level

$\left(\hat{\chi}^2 = 74 >> \chi^2_{34,0.005} = 59\right)$. Similarly, a test of March 2001/March 2002 inbound data also rejects the null hypothesis $\left(\hat{\chi}^2 = 70 >> \chi^2_{34,0.005} = 59\right)$.

Many structural schedule changes occurred between March 2001 and March 2002 due to the terrorist attacks of September 11, 2001. Some of these effects are reflected in flight/itinerary variables. Note that airlines implemented different schedule reduction strategies following the terrorist attacks and as such, the schedule changes discussed here do not generally apply to the industry as a whole. The year-over-year (YOY) reduction in flight departures for continental U.S. markets was 21%. The number of markets served by one flight a day remained the same (138 in March 2001 compared to 139 in March 2002), however the number of markets served by multiple flights a day decreased from 280 to 235. Markets no longer served by the carrier of interest were either dropped from the schedule or converted to markets served exclusively by express partners (only mainline service is reflected in this analysis). The average flight interarrival time for markets served by multiple daily flights decreased slightly from 125 to 122 minutes, presumably due to eliminating service or reducing to a single itinerary in markets with relatively low service and high interarrival times. Further, as seen in Table 3, the distribution of interarrival times changed. The percentage of flights with interarrival times between 91 and 150 minutes increased from 19% to 25%. As reflected in Table 1, this distributional shift may account for YOY differences in early standby carrier coefficients for this interarrival time period (0.27 in 2001 versus 0.14 in 2002 for outbound and 0.19 versus 0.04 for inbound).

[Insert Table 3 about here]

While the difference in departure time coefficients associated with the early standby alternative are not statistically different between March 2001 and March 2002, there is an interesting relationship between changes in flight schedule and early standby departure time coefficients. Specifically, coefficients associated with the early standby alternative reflect the unwillingness of passengers to early standby the earlier in the day the itinerary departs (changes due to elimination of early standby alternatives is captured in the generation of choice sets). This resistance increases in March 2002, particularly for flights departing after 9 a.m. relative to flights departing before 9 a.m. A closer look at flight frequency by time of day reveals why this effect is not as strong for the 6-9 a.m. category (note, reduction of early standby opportunities is captured directly in the generation of choice sets, so parameter estimates reflect passengers' desire to early standby). Specifically, the effect is not as strong in 2002 due in part to added security measures implemented after the terrorist attacks. These security measures increased the time required for travelers to check-in and clear security checkpoints. Few travelers (and employees!) wanted to arrive at the airport at 4 or 5 a.m. to make a 6 a.m. departure.

[Insert Figure 3 about here]

Security measures implemented after 9/11 may also influence changes in frequent flyer behavior and coefficients. In particular, after 9/11 positive baggage match was implemented for all domestic U.S. flights. This policy requires airlines to remove checked baggage for passengers that check-in, but do not board the flight. Prior to 9/11, it was possible for a passenger to check baggage, clear security, and standby for an earlier flight. After 9/11, once a passenger checks baggage, the passenger may not be able to standby for an earlier flight. To the extent that all travelers arrived earlier to the airport due to variability and uncertainty in security check-in times, this result suggests that frequent flyer members, who travel frequently for business and are less likely to check baggage, were less affected by new positive baggage security measures implemented after the terrorist attacks and able to take advantage of early standby opportunities. As seen in Table 1, early standby coefficients associated with elite members relative to non-members increase from 0.57 to 1.06 for outbound itineraries while coefficients for general members relative to non-members increase from 0.33 to 0.75. Both of these differences are statistically significant at the 0.05 level. Thus, although the population rate of early standbys remains relatively unchanged for the outbound segment (0.89% to 0.92%), the rate is not consistent across passenger categories.

5. Summary and conclusions

This study is one of the first in the airline industry to be based on disaggregate passenger data. To the best of our knowledge, this is the most comprehensive study of airline passengers' no-show behavior published to date and the only one that has examined airline passenger's standby behavior. Most importantly, the feasibility and importance of using detailed passenger data in airline forecasting models has been demonstrated. From a practical perspective, incorporating passenger and directional itinerary information allows carriers to develop more accurate forecasting models that capture shifts in no-show rates due to changes in the underlying passenger demands and itinerary mixes. Moreover, it is only by using disaggregate passenger data that a broad range of analysis and managerial questions could be explored. To the extent that airlines incorporate passenger data into their existing business process and forecasting models, a broader range of business insights can be gleaned.

Acknowledgements

The authors would like to thank the U.S. carrier who supported this study and Gregory Coldren, a doctoral student at Northwestern University, who provided assistance in obtaining and interpreting flight and itinerary variables. The authors are also grateful for the helpful comments of reviewers, including individuals who reviewed an earlier draft that was presented at the 83rd Annual Transportation Research Board meeting.

REFERENCES

American Airlines. (2003). Detailed fare rules. Retrieved June 30, 2003, from www.americanairlines.com.

Anonymous (1981, December 10 & 20). A crackdown on no-shows: NY Air, Air Florida – stick vs. carrot. *Transport Central*, 29(13-14), 3-4.

Garrow, L. (2001). *Factors airline analysts think to be important drivers of no-show behavior: Input from no-show user council.* Elk Grove Village, IL: United Airlines.

Garrow, L. (2004). *Comparison of choice models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior*. Published doctoral dissertation. Northwestern University, Evanston, Illinois.

Kalka, K., & Weber, K. (2000). PNR-based no-show forecast. Presented at the AGIFORS Reservation and Yield Management Study Group, New York City.

Manski, C., & Lerman, S. (1977). The estimation of choice probabilities from choice based samples. *Econometrica*, 45, 1977-1988.

Pastor, J. (2000). What exactly is data mining. Presented at the AGIFORS Reservation and Yield Management Study Group, New York City.

Pina, M. (1995, October 5). On-line bookings spur rise in airlines' no-shows. *Travel Weekly,* 54(79), pp. 1, 4.

Ruppenthal, K., & Toh, R. (1983). Airline deregulation and the no-show/overbooking problem. *The Logistics and Transportation Review,* 19(2), 111-121.

Scholin, A. (1983, August). The no-show problem. *Airline Executive*, 7(8), 39-40.

LIST OF TABLES AND FIGURES

Table 1 MNL model results

	3/01 Outbound	3/01 Inbound	3/02 Outbound	3/02 Inbound
Constants (ref. = show)				
Alternative specific constant NS	1.33 (11.8)	1.52 (13.3)	1.59 (10.0)	1.92 (14.0)
ASC ESB: Duration ≤ 180 mins	-0.42 (2.0)	0.25 (1.8)	-0.28 (1.0)	1.04 (5.7)
ASC ESB: 180 < duration ≤ 300 mins	-0.55 (2.5)	0.09 (0.5)	-0.58 (1.9)	0.67 (3.3)
ASC ESB: Duration > 300 mins	-0.87 (3.5)	-0.53 (2.9)	-0.60 (2.1)	0.23 (1.0)
Alternative specific constant LSB	-0.37 (3.4)	-0.23 (2.0)	-0.95 (5.6)	-0.59 (4.1)
Day of Week (ref.=Sunday-Tuesday)				
Wed – Fri ESB	0.26 (2.9)	0.35 (4.7)	0.42 (3.5)	0.33 (3.7)
Dept. Time (ref. = after 7 pm and for NS 6-9 am)				
Depart 9:00 am – 7:00 pm NS	-0.29 (3.3)	-0.21 (2.4)	-0.14 (1.1)	-0.22 (2.2)
Depart 6:00 am – 9:00 am ESB	-1.46 (7.2)	-1.58 (7.1)	-1.49 (5.0)	-1.72 (6.6)
Depart 9:01 am – 4:00 pm ESB	-1.16 (7.3)	-0.85 (8.4)	-1.43 (7.1)	-1.32 (9.7)
Depart 4:01 pm – 7:00 pm ESB	-0.58 (3.6)	-0.10 (1.0)	-0.86 (4.3)	-0.49 (3.6)
Carrier Capacity (100's seats) before scheduled departure for ESB				
Arrive 1-90 mins earlier	0.39 (7.0)	0.28 (6.4)	0.36 (4.9)	0.32 (6.1)
Arrive 91-150 mins earlier	0.27 (4.6)	0.19 (4.5)**	0.14 (2.1)	0.04 (0.9)**
Arrive 151-300 mins earlier	0.07 (2.0)	0.07 (2.4)	0.06 (1.5)	0.06 (2.1)
Carrier Capacity (100's seats) after scheduled departure for LSB				
Arrive 1-90 mins later	0.08 (1.6)	0.11 (2.4)	0*	0*
Arrive 91-150 mins later	0.15 (3.0)	0.21 (4.1)	0.13 (2.0)	0.19 (3.3)
Arrive 151-300 mins later	0.11 (3.6)	0*	0.09 (2.5)	0.15 (4.3)
Schedule Presence (Ratio of total flights for dominant carrier vs. nearest competitor)^0.5				
Departure city presence ESB	0.14 (3.1)	0.14 (4.6)	0.17 (2.4)	0.21 (4.5)
Departure city presence LSB	0.33 (8.8)	0.27 (7.6)	0.58 (9.2)	0.37 (6.6)
E-ticket NS	-1.81 (20.2)	-1.49 (19.8)	-2.16 (18.4)	-1.74 (19.9)
Booking Class (ref. = low-yield)				
First and business NS	-0.74 (3.9)	-0.19 (1.3)	0.05 (0.2)	-0.43 (2.5)
First and business ESB	-1.01 (4.2)	-1.09 (6.6)	-1.58 (4.8)	-1.10 (5.8)
First and business LSB	-1.03 (6.4)	-1.56 (6.0)	-1.52 (5.1)	-1.53 (5.4)
High yield NS	0.08 (0.7)	0.26 (2.7)	-0.10 (0.8)	-0.07 (0.6)
High yield ESB	-0.25 (2.1)	0.11 (1.2)	0.08 (0.5)	-0.10 (0.9)
High yield LSB	-0.46 (4.5)	-0.07 (0.6)	-0.77 (5.3)	-0.46 (3.4)
Frequent Flyer (ref. = not a member)				
General member NS	-0.54 (4.6)	-0.59 (5.3)	-0.81 (5.1)	-0.33 (2.6)
General member ESB	0.33 (2.4)**	0.37 (3.9)	0.75 (4.3)**	0.39 (3.2)
General member LSB	-0.63 (5.6)	-0.44 (3.8)	-0.17 (1.1)	-0.01 (0.1)
Elite member NS	-0.13 (1.0)	-0.02 (0.2)	-0.32 (1.9)	-0.26 (1.8)
Elite member ESB	0.57 (4.2)**	0.31 (2.5)	1.06 (5.4)**	0.26 (1.9)
Elite member LSB	-0.47 (3.7)	-0.11 (0.9)	0.29 (1.6)	-0.06 (0.3)
Group Size (ref. = travel alone)				
Group 2+ NS	-0.60 (5.2)	-0.46 (5.0)	-0.62 (4.3)	-0.35 (2.9)
Group 2+ ESB	-0.71 (6.0)	-0.63 (7.2)	-0.67 (4.0)	-0.71 (6.2)
Group 2+ LSB	-0.50 (4.6)	-0.33 (2.9)	-0.44 (2.8)	-0.13 (0.9)
Model Fit Statistics				
LL Zero / LL Constants	-3575 / -3539	-4798 / -4681	-2203 / -2182	-3428 / -3292
LL Model	-3112	-4160	-1825	-2868
ρ^2_{zero} / $\rho^2_{constant}$	0.130 / 0.121	0.133 / 0.111	0.172 / 0.164	0.163 / 0.129
Number of cases / Number of variables	2,761 / 34	3,674 / 33	1,791 / 33	2,729 / 33

KEY: Coefficient (t-stat). Highlighting shows pairwise t-test of 3/01 out- and in-bound models significant at 0.02.
*Constrained at zero to avoid negative value. **Subset of pairwise t-tests of '01 and '02 models significant at 0.05.

Table 2
Choice probability forecasts for a hypothetical passenger and itinerary

	E-ticket				No e-ticket			
	6-9 am	9am-4 pm	4-7 pm	7pm-mid	6-9 am	9am-4 pm	4-7 pm	7pm-mid
Outbound								
SH	96.99%	96.85%	96.44%	94.92%	85.85%	85.74%	85.42%	81.15%
NS	2.54%	2.54%	2.53%	3.32%	13.74%	13.72%	13.67%	17.35%
ESB	0.40%	0.54%	0.96%	1.68%	0.35%	0.48%	0.85%	1.44%
LSB	0.07%	0.07%	0.07%	0.07%	0.07%	0.07%	0.07%	0.06%
Inbound								
SH	94.76%	93.97%	92.32%	91.10%	82.43%	81.83%	80.58%	77.37%
NS	4.35%	4.32%	4.24%	5.16%	16.80%	16.68%	16.42%	19.45%
ESB	0.78%	1.60%	3.33%	3.63%	0.68%	1.39%	2.90%	3.08%
LSB	0.11%	0.11%	0.11%	0.11%	0.10%	0.10%	0.09%	0.09%

Note: Two decimal places shown for presentation purposes.

Table 3

Comparison of 2001 and 2002 flight service

	March 2001	March 2002	YOY Change
Distribution of interarrival times			
0 < interarrival time ≤ 90 mins	15,265 (26%)	9,109 (20%)	-6,156 (-40%)
90 < interarrival time ≤ 150 mins	11,266 (19%)	11,412 (25%)	+146 (1.3%)
150 < interarrival time ≤ 300 mins	13,426 (23%)	10,953 (24%)	-2,473 (-18%)
300 < interarrival time	17,829 (31%)	14,106 (31%)	-3,723 (-21%)
Average interarrival time (min)	125	122	-3 (-2.4%)
Number of markets served			
1 flight / day	138	139	+1 (0.7%)
> 1 flight / day	280	235	-45 (-16%)

KEY: Number (%).

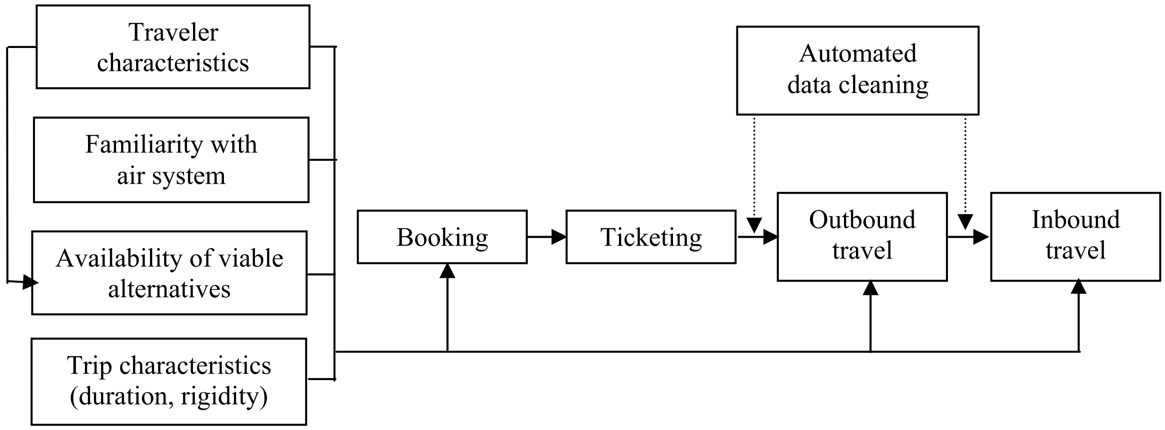

Fig. 1. Conceptual model of airline passengers' rescheduling behavior.

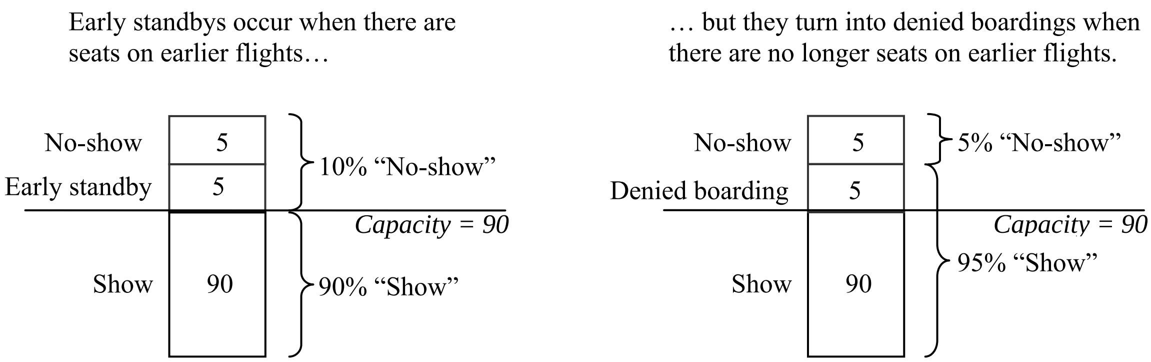

Fig 2. Relationship between early standbys and denied boardings.

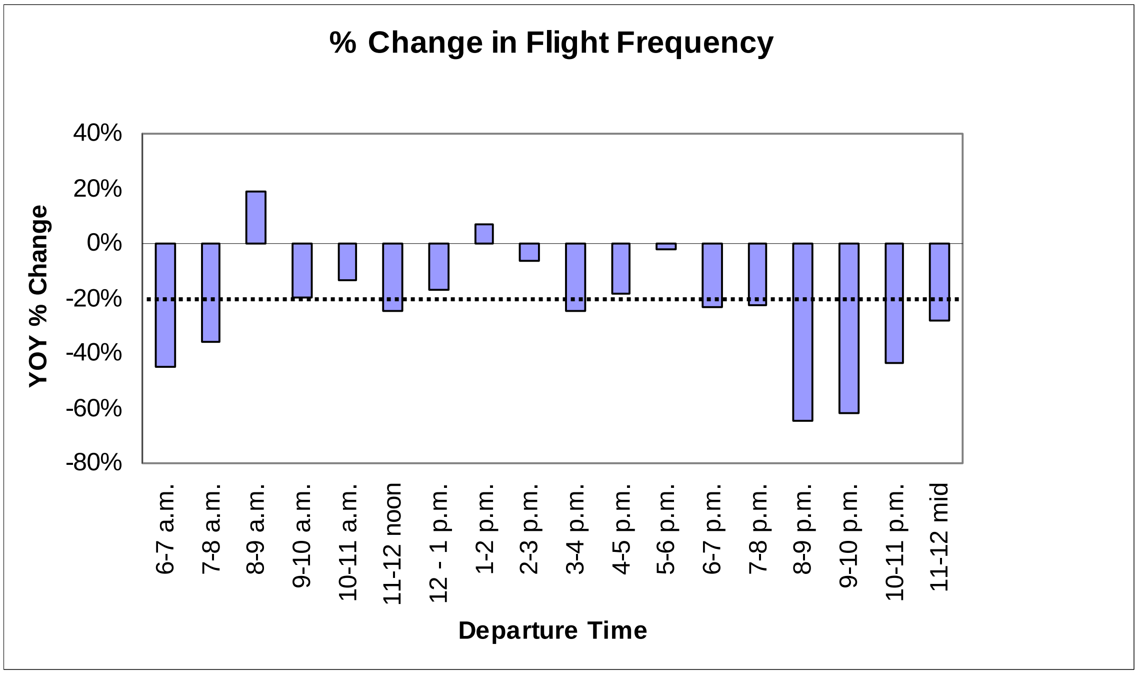

Fig. 3. Percent change in flight frequency from March 2001 to March 2002 by time of day