The relationship between on-time performance and airline market share: a new approach

Yoshinori Suzuki *

Department of Logistics, Operations and MIS, College of Business, Iowa State University, 300 Carver Hall, Ames, IA 50011, USA

Received 4 March 1999; received in revised form 20 August 1999; accepted 24 September 1999

Abstract

We propose a new method of modeling the relationship between on-time performance and market share in the airline industry. The idea behind the method is that the passengers’ decision to remain (use same airline) or switch (use other airlines) at time \( t \) depends on whether they have experienced flight delays at time \( t - 1 \) or not. More specifically, we posit that the passengers who experienced flight delays are more likely to switch airlines for the subsequent flight than those passengers who did not experience delays. To capture such effect, we develop an aggregate-level Markovian type model that estimates the transition probability matrices separately for the passengers who experienced flight delays at time \( t - 1 \) and for those who did not experience delays. The model was calibrated with the US DOT data. The study results imply that, once experiencing flight delays, passengers are more likely to switch airlines. The results also imply that on-time performance affects a carrier’s market share primarily through the passengers’ experience, and not though the “advertisement” of performance. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Airline demand; On-time performance; Markov model; Non-linear regression

1. Introduction

Consider the following scenario. You and your colleague made a business trip from point A to point B. You used ABC airline, while your colleague chose airline XYZ. The ABC flight did not arrive on time because of some minor engine problems that were found just before the departure. Consequently, you arrived at point B 1 h behind the scheduled arrival time, and missed a very important business negotiation. Your colleague, on the other hand, arrived on time and was able to attend the important business negotiation that you had to miss. In this scenario, are you more
likely to switch to another airline for the next business trip than your colleague? Most likely, the answer is “Yes”.

The above example shows that the passengers’ choice of airlines may be affected by the on-time arrival experience of passengers. This example also suggests that the passengers may show stronger reaction (switch to other airlines) when they experience utility losses (service level below their expectation) than when they experience utility gains (service level above their expectation). This type of asymmetric pattern in human response to gains and losses is known as the loss aversion (Tversky and Kahneman, 1991), and was recently introduced to the transportation literature by Suzuki and Tyworth (1998). Simply put, the property of loss aversion suggests that human beings evaluate product/service quality relative to a certain reference point (e.g., expectation), and give heavier weights to the losses (negative deviations from the reference point) than to the equivalent-sized gains (positive deviations). Although developed recently, the property of loss aversion is supported by significant empirical evidence. For example, a recent empirical study by Hardie et al. (1993) utilized the concept of loss aversion in their multinomial logit model of consumer brand choice, and showed that the model performance can be improved significantly by incorporating the loss aversion effect.

The loss aversion property seems to have potential for providing important managerial implications in the air transportation area, especially for gaining deeper insights into the nature of the relationship between airline demand and customer service quality such as on-time performance. The loss aversion property, however, has not been utilized in the studies of airline demand. A typical approach to modeling the effect of on-time performance on airline passenger demand is to express a carrier’s sales (or market share) as a function of the carrier’s on-time performance and other exogenous variables. Both linear and non-linear specifications are widely used. Examples of this approach include the studies by Dresner and Windle (1992), Nako (1992), Ghobrial and Soliman (1992), Dresner and Xu (1995), Proussaloglou and Koppelman (1995), Yoo and Ashford (1996) and Suzuki (1998). This approach, however, implicitly assumes that all passengers are homogeneous, i.e., all of the passengers show identical airline choice behaviors at time t regardless of whether they have experienced flight delays at time t – 1 or not. This condition indicates that, if passengers are loss averse, the existing models will not capture the portion of data variance that is explained by the passengers’ loss aversion tendencies.

In this paper we model the aggregate-level relationship between airline market share and on-time performance by using the framework that captures the effect of passenger loss aversion. To incorporate the loss aversion tendencies of airline passengers, we formulate a model that permits separate estimation of airline choice probabilities for the passengers who experienced flight delays at time t – 1 and for those who did not. Although a model that possesses such a characteristic may best be constructed by using a discrete choice (disaggregate) modeling framework, we do not employ this approach for a practical reason. The disaggregate models can incorporate individual heterogeneity and may be more appropriate for modeling choice behaviors, but its use in the airline industry is difficult because of data unavailability. If using disaggregate choice models, one must use the data that are collected from a group of people over time (panel data), which probably are not available in the airline industry. The aggregate models, on the other hand, can be estimated by using the available data, such as the aggregate passenger travel data published by the US DOT. This condition implies that our model may be used quite conveniently by the airline management.
2. Model construction

2.1. Modeling airline market share

Our model formulation follows the aggregate Markovian type models (e.g., Lee et al., 1970; Horsky, 1976, 1977; Leeflang and Boonstra, 1982; Givon and Horsky, 1990). Let us assume that each passenger uses air transportation service once a month, or in other words, all passengers adhere to the average inter-travel time of one month.1 While this is a strong assumption,2 this type of assumption is implicitly made in virtually all existing market share models (Givon and Horsky, 1990). Under this assumption, the market share of each air carrier in a particular route can be written as follows:

\[ S_{it} = S_{it}^{-i} + S_{it}^{i}, \]

where \( S_{it} \) is carrier \( i \)'s share of total passengers transported in the route in month \( t \), \( S_{it}^{-i} \) is carrier \( i \)'s share that is obtained from the passengers who used other carriers' service in month \( t - 1 \), but used carrier \( i \)'s service in month \( t \) (switchers), and \( S_{it}^{i} \) is the share that is obtained from the passengers who used carrier \( i \)'s service in months \( t - 1 \) and \( t \) (non-switchers). If there are \( n \) carriers in the route, we have \( n \) demand functions of the form given in Eq. (1).

The portion of the market share that is derived from the switchers can be specified as:

\[ S_{it}^{-i} = \sum_{k \neq i} \{ S_{kt-1} \lambda_k A_{ikt} \}, \]

where \( S_{kt-1} \) is the market share of carrier \( k \) \((k \neq i)\) in month \( t - 1 \) in the route, \( \lambda_k \) is an unknown parameter representing the passenger switching rate of carrier \( k \) (average proportion of passengers switching from carrier \( k \) to other airlines: \( 0 \leq \lambda_k \leq 1\ \forall k \)), and \( A_{ikt} \) is airline \( i \)'s share of switchers from airline \( k \) in month \( t \), or the fraction of switchers from airline \( k \) who are attracted by airline \( i \) (observe that \( S_{kt-1} \lambda_k \) represents the switchers from airline \( k \) to be allocated to other airlines during month \( t \)). In this paper, we call \( A_{ikt} \) as the index of carrier \( i \)'s attractiveness to the switchers from carrier \( k \).

The reason why \( \lambda \) parameter is estimated for each carrier is that the switching rates are likely to be influenced by carrier characteristics (e.g. airfare) and are likely to be different across carriers. We assume that \( \lambda_k \) is stationary (does not change over time). \( A_{ikt} \), on the other hand, is assumed to be impacted by on-time performance and other factors that fluctuate over time, and is considered non-stationary. Since \( A_{ikt}^{k} \) represents a percentage, \( A_{ikt}^{k} \) must satisfy the following:

---

1 The assumption of one month travel interval is directly or implicitly supported by several sources. It is well known in the airline industry that the top 20% of all air travelers (frequent travelers) account for 80% of the airline passenger revenue. These frequent flyers generally travel more than 10 times per year. This condition indicates that the passengers who account for 80% of airline revenue travel approximately every 4–5 weeks, on average. Although frequent travelers may pay higher fares per trip than nonfrequent travelers, it is probably safe to say that a substantial proportion of passengers fly approximately once every month. See, for example, Toh and Hu (1990) and Fortune (1988, pp. 169–172). The average inter-travel time of one month is also supported by an empirical study by Suzuki (1999).

2 The aggregate Markovian type models cannot be estimated without using this assumption. Though the use of this fixed interpurchase time assumption is a major limitation of the aggregate Markovian type models, they are widely used in the marketing literature for empirical analysis.
By imposing constraints (3a) and (3b), our model satisfies the mandatory condition of the Markovian type model that each row of the transition probability matrix sum to 1.

The portion of the market share that is derived from the non-switchers can be modeled as the product of airline $i$’s market share in month $t-1$ and its passenger retention rate $(1-\lambda_i)$:

$$S_{it} = S_{it-1}(1-\lambda_i).$$

(4)

After replacing $S_{it}^{-}$ and $S_{it}^{i}$ of Eq. (1) with the functional forms given in Eqs. (2) and (4), respectively, we obtain the following formula:

$$S_{it} = \sum_{k \neq i} \left\{ S_{it-1} \lambda_k A_{it}^k \right\} + S_{it-1}(1-\lambda_i).$$

(5)

Fig. 1 shows the implied Markovian transition probability matrix derived from Eq. (5).

2.2. Incorporating loss aversion effect

The market share model of the form given in Eq. (5) does not distinguish between utility gains and losses of passengers. Since all passengers expect on-time arrival, we assume that passengers obtain utility gains when their flights arrive on time and incur utility losses when their flights are delayed. To include such loss aversion effect we allow the switching rate parameter $(\lambda_i)$ to vary across two groups of passengers: the passengers who did not experience flight delays in month

\[3\] According to Winer (1988), one of the definitions of reference service/price is the expected levels of service/price.
Note that we approximate the number of passengers who used airline Sk in month $t - 1$ and did not experience flight delays in month $t - 1$. Under this condition, the $S_{it}^-$ function (Eq. 2) can be modified as follows:

$$S_{it}^- = \sum_{k \neq i} \{S_{kt-1}Q_{kt-1}A_{Git}^k\} + \sum_{k \neq i} \{S_{kt-1}(1 - Q_{kt-1})\lambda_{Ak}A_{Li}^k\},$$

where $Q_{kr}$ is the on-time performance (percentage of flights arriving on-time) of carrier $k$ in month $t$ in the route, $\lambda_{Gk}$ is the switching rate of carrier-$k$ passengers who experienced utility gains, $\lambda_{Ak}$ is the switching rate of carrier-$k$ passengers who experienced utility losses, $A_{Git}^k$ is carrier $i$'s attractiveness to the switchers from carrier $k$ who experienced utility gains, and $A_{Li}^k$ is carrier $i$'s attractiveness to the switchers from carrier $k$ who experienced utility losses. As before, $A_{Git}^k$ and $A_{Li}^k$ must satisfy the following conditions:

$$0 \leq A_{Git}^k \leq 1, \quad 0 \leq A_{Li}^k \leq 1 \quad \forall k, i, t, \quad i \neq k,$$

$$\sum_{i \neq k} A_{Git}^k = \sum_{i \neq k} A_{Li}^k = 1 \quad \forall k, t.$$  

Note that we approximate the number of passengers who used airline $k$ in month $t - 1$ and did not experience flight delays by $S_{kt-1}Q_{kt-1}$, and the number of passengers who used airline $k$ and experienced flight delays by $S_{kt-1}(1 - Q_{kt-1})$. The underlying assumption employed here is that airlines use approximately the same fleet sizes and/or loads in a given route. Also note that we differentiate the attractiveness of carrier $i$ to the switchers who did not experience flight delays ($A_{Li}^k$) from that to the switchers who experienced flight delays ($A_{Git}^k$). The reason is because the carrier preference of those passengers who experienced flight delays and that of those who did not experience delays may be different (e.g., passengers with flight delay experience may put more weight on on-time performance when choosing airlines for the subsequent flight).

When the loss aversion effect is incorporated into the model, the $S_{it}^-$ function becomes:

$$S_{it}^- = S_{it-1}Q_{it-1}(1 - \lambda_{Gi}) + S_{it-1}(1 - Q_{it-1})(1 - \lambda_{Li}).$$

Note that $(1 - \lambda_{Gi})$ is the retention rate of carrier-$i$ passengers who did not experience flight delays in month $t - 1$, whereas $(1 - \lambda_{Li})$ is the retention rate of carrier-$i$ passengers who experienced flight delays. After combining Eqs. (6) and (8), our model specification can be written as follows:

$$S_{it} = \sum_{k \neq i} \{S_{kt-1}Q_{kt-1}\lambda_{Gk}A_{Git}^k\} + \sum_{k \neq i} \{S_{kt-1}(1 - Q_{kt-1})\lambda_{Ak}A_{Li}^k\} + S_{it-1}(1 - \lambda_{Gi})$$

$$+ S_{it-1}(1 - Q_{it-1})(1 - \lambda_{Li}).$$

The behavioral assumption implied by Eq. (9) is that passengers follow a “two-step” process when making airline choice decisions. More specifically, passengers first decide whether or not to switch airlines based on their previous flight delay experiences, and then decide what other airlines to use based on the attractiveness of other airlines.

Eq. (9) indicates that if airline passengers are loss averse, $\lambda_L$ will be greater than $\lambda_G$ for all carriers. On the other hand, if passengers are loss neutral, $\lambda_L$ will be equal to $\lambda_G$. If $\lambda_L$ turns out to be smaller than $\lambda_G$, passengers are said to be loss prone. Fig. 2 expresses the model in matrix notation and shows that the model computes transition probability matrices separately for passengers who experienced flight delays in month $t - 1$ and for those who did not.
2.3. Airline attractiveness

In our model, the allocation of switchers from airline $k$ to other airlines ($n - 1$ airlines excluding airline $k$) is determined by the attractiveness of non-$k$ airlines ($A_{G_{kit}}$ and $A_{L_{kit}}$). This attractiveness can be considered as a function of multiple factors representing airline attributes:

$$A_{G_{kit}} = A_{L_{kit}} = f(X_{hit}, X_{h-ki}), \quad h = 1, 2, \ldots, H,$$

where $X_{hit}$ is the $h$th variable representing the attractiveness of carrier $i$ in month $t$, $X_{h-ki}$ is the $h$th variable representing the attractiveness of all the non-$k$ carriers in month $t$, and $H$ is the total number of carrier attractiveness variables used in the model. Observe that an attractiveness of carrier $i$ is not a function of carrier $k$’s attributes. Since we know that the switchers from airline $k$ will not use airline $k$ in month $t$, the switchers’ choice of airlines in month $t$ would simply be determined by the attributes of the airlines included in the switchers’ “choice set”, which does not include airline $k$. We specify $A_{G_{kit}}$ and $A_{L_{kit}}$ by using the following functions and constraints:

$$A_{G_{kit}} = \alpha_0 \frac{\exp(S_{hit})}{\sum_{j \neq k} \exp(S_{j-it})} + \sum_{h=1}^{H} \alpha_h \frac{\exp(X_{hit})}{\sum_{j \neq k} \exp(X_{j-it})},$$

$$A_{L_{kit}} = \beta_0 \frac{\exp(S_{hit})}{\sum_{j \neq k} \exp(S_{j-it})} + \sum_{h=1}^{H} \beta_h \frac{\exp(X_{hit})}{\sum_{j \neq k} \exp(X_{j-it})}.$$
Subject to
\[ \alpha_h \geq 0, \quad \beta_h \geq 0 \quad \forall h, \] (11c)
\[ \sum_{h=0}^{H} \alpha_h = 1, \quad \sum_{h=0}^{H} \beta_h = 1, \] (11d)
\[ \alpha_h, \beta_h = \text{constant across equations (carriers)} \quad \forall h, \] (11e)

where \( \alpha \) and \( \beta \) coefficients are the unknown parameters and \( i \in j \).

Eqs. (11a) and (11b) indicate that an airline’s attractiveness is represented by a linear combination of the airline’s attributes, each of which is measured in relation to those of the airlines included in the switchers’ “choice set”. Observe that we distinguish between \( A_{Gh}^k \) and \( A_{Lh}^k \) by allowing the unknown parameters to vary across Eqs. (11a) and (11b). In Appendix A we prove that Eqs. (11a) and (11b), when constraints (11c)–(11e) are imposed, satisfy the required conditions of the model (constraints (7a) and (7b)). Each attractiveness variable is exponentiated to allow for incorporating the variables with non-positive values.

The first term of Eqs. (11a) and (11b) represent the “carrier specific effects”, which are equivalent to the “carrier constants” that are used in many cross-section time-series econometric airline demand studies. Because of the way our model is formulated, the “carrier effect” cannot be captured in a standard manner by using carrier dummy variables (because of constraint (7b): see Appendix A). For this reason, we use a “heuristic” approach to capture the carrier effects. That is, instead of using carrier dummy variables, we use a variable that represents the ratio of one-period lagged market share of carrier \( i \) to the one-period lagged summed market shares of non-\( k \) carriers, to capture carrier specific effects. This approach is similar to the standard dummy variable approach (one factor fixed effect model). The dummy variable approach captures the portion of data variance not explained by the explanatory variables (carrier effect) by the differences of average dependent-variable values across carriers. Our approach, on the other hand, captures the unexplained portion of data variance (not explained by \( X_{hit} \)) by the differences of one-period lagged dependent-variable values across carriers. Thus, while the standard dummy variable approach measures the “carrier effect” of carrier \( i \) by the average (across all time periods) market share of carrier \( i \), our approach measures this effect by the one-period lagged market share of carrier \( i \). From the model interpretation viewpoint, the first terms of Eqs. (11a) and (11b) imply that, holding other effects (\( X_{hit} \) effects) fixed, switchers from carrier \( k \) will be distributed to all non-\( k \) carriers according to their respective (one-period lagged) market shares.

Eqs. (11a)–(11e) indicate that the carrier attractiveness (\( A_{Gh}^k \) and \( A_{Lh}^k \)) must be an increasing function of each of the arguments (\( X_{hit} \)). This pattern suggests that if the model includes the variables whose higher values tend to lower an airline’s attractiveness (e.g., airfares), we must attach a minus sign to these variables so that \( A_{Gh}^k \) and \( A_{Lh}^k \) become increasing functions of all the variables.

---

4 Although we could have used the average market share, instead of the one-period lagged market share, to capture carrier specific effects, the use of this approach brings about estimation bias. Notice that the average market share is endogenous (average market share of carrier \( i \) is a function of \( S_{it} \) \( \forall t \)).
2.4. The model

The ultimate form of the model can be obtained by replacing $A_{Gi}^k$ and $A_{Li}^k$ of Eq. (9) with Eqs. (11a) and (11b), respectively, and by adding a stochastic term. By doing so, we obtain the following formula:

$$S_{it} = (1 - \lambda_{Gi})S_{it-1}Q_{it-1} + (1 - \lambda_{Li})S_{it-1}(1 - Q_{it-1}) + \sum_{k \neq i} \left[ \alpha_0 \lambda_{Gk} \frac{\exp(S_{it-1})}{\sum_{j \neq k} \exp(S_{it-1})} S_{kt-1}Q_{kt-1} \right]$$

$$+ \sum_{h=1}^{H} \sum_{k \neq i} \left[ \alpha_h \lambda_{Gk} \frac{\exp(X_{hit})}{\sum_{j \neq k} \exp(X_{hjt})} S_{kt-1}Q_{kt-1} \right]$$

$$+ \sum_{k \neq i} \left[ \beta_0 \lambda_{Lk} \frac{\exp(S_{it-1})}{\sum_{j \neq k} \exp(S_{jt-1})} S_{kt-1}(1 - Q_{kt-1}) \right]$$

$$+ \sum_{h=1}^{H} \sum_{k \neq i} \left[ \beta_h \lambda_{Lk} \frac{\exp(X_{hit})}{\sum_{j \neq k} \exp(X_{hjt})} S_{kt-1}(1 - Q_{kt-1}) \right] + \epsilon_{it}.$$  \hspace{1cm} (12)

Eq. (12) is subject to constraints (11c)–(11e), as well as to the followings:

$$0 \leq \lambda_{Gi} \leq 1, \quad 0 \leq \lambda_{Li} \leq 1 \quad \forall i. \hspace{1cm} (13)$$

Eq. (12) can be estimated by using the non-linear least-squares method. Note, however, that the residual terms of Eq. (12) will generally be serially and contemporaneously correlated because of the nature of the data and the restrictions imposed on the parameter values. Thus, the model should be estimated by using the generalized least-squares (GLS), rather than the standard ordinary least-squares (OLS). Further, if the data are not well-conditioned, the estimated parameter values will not be bounded between 0 and 1, in which case the model must be estimated via the use of a constrained optimization technique.

3. Model properties

The model has several interesting characteristics. First, the model is capable of estimating the switching rate parameters separately for passengers who experienced flight delays ($\lambda_{Li}$) and for those who did not experience delays ($\lambda_{Gi}$). Thus, if airline passengers are loss averse with respect to on-time performance, the model can capture such loss aversion tendency by providing significantly different estimates for $\lambda_{Li}$ and $\lambda_{Gi}$, where $\lambda_{Li} > \lambda_{Gi} \forall i$.

Second, the model estimates the airline attractiveness separately for the switchers who experienced flight delays ($A_{Li}^k$) and for those who did not ($A_{Gi}^k$) by allowing the unknown parameters to vary across $A_{Gi}^k$ and $A_{Li}^k$. Such a model formulation may provide us some useful insights into the passengers’ airline switching behavior. For example, if the estimation results indicate that $\alpha_1 < \beta_1$ (where $X_{hit}$ represents the on-time performance of carrier $i$ in month $t$), the results imply that the switchers who experienced flight delays in month $t - 1$ would put more weight on the on-time performance attribute when choosing airlines in month $t$, than those who did not experience flight delays in month $t - 1$. 
Third, because the model employs the Markovian type modeling framework, it possesses the characteristic that the summed market shares of \( n \) carriers will always be 1. Whether a model possesses this characteristic or not is an important issue, because the market-share models that do not meet this condition make little intuitive sense.

4. Empirical analysis

4.1. Variable selection

Like many contemporary airline passenger demand models, we measure an airline’s attractiveness by its on-time performance (see, e.g., Suzuki, 1998; Dresner and Xu, 1995; Nako, 1992), airport dominance (Dresner and Windle, 1992; Borenstein, 1991; Borenstein, 1989), and safety record (Suzuki, 1998; Bruning et al., 1985; Hu and Bruning, 1984; Comm, 1993):

\[
A_{i,t}^k = A_{i,t}^k = f(Q_{i,t-3}, Q_{-kt-3}, DOM_{i,t}, DOM_{-kt}, CSH_{i,t}, CSH_{-kt}),
\]

where \( Q_{-kt-3} \) is the on-time performance of carriers other than \( k \) in month \( t - 3 \), \( DOM_{i,t} \) is the airport dominance (number of flight departures/arrivals performed at origin and destination airports) of carrier \( i \) in month \( t \), \( DOM_{-kt} \) is the airport dominance of carriers other than \( k \) in month \( t \), \( CSH_{i,t} \) is the number of passenger fatalities experienced by carrier \( i \) in months \( t, t - 1 \) and \( t - 2 \) (3-months moving average), and \( CSH_{-kt} \) is the passenger fatalities experienced by carriers other than \( k \) in months \( t, t - 1 \) and \( t - 2 \).

Only the lagged on-time performance variables (\( Q_{i,t-3} \) and \( Q_{-kt-3} \)) are included in Eq. (14), because passengers do not have access to the up-to-date on-time performance information of carriers. The reason is because the US DOT makes the on-time performance information of carriers available to the public 40 days after the end of the month. For example, the on-time performance of carriers for the month of January will be available to the public approximately on 10 March. Thus the latest on-time performance data which is accessible by the passengers purchasing tickets after 10 March are the January data. But since most of the air tickets are purchased several days before the actual flights, most of the tickets purchased after 10 March will be reflected in the passenger enplanement figures of the month of April. Thus, \( Q_{i,t-3} \) and \( Q_{-kt-3} \) should be used in Eq. (14). On the other hand, only the “current” airport dominance variable is included in Eq. (14) because an airline’s attractiveness is determined by how dominant the airline is for the month of actual travel. Note that when an airline is on strike, this airline’s airport dominance variable becomes 0 or very close to 0. This condition indicates that, when an airline is on strike, the decrease in market share for this airline is explained by the decrease in its airport dominance variable. Thus, our model does not miss the market share fluctuation due to strikes. The safety variable contains both the current and the lagged passenger fatality information (3-months moving average).

The price attribute (airfare) is not included in Eq. (14) because of data unavailability (to our knowledge, time-series data of carriers’ average airfares for a specific route are not publicly available). For this reason, we capture the price effect by the “carrier-specific constants” (\( z_0 \) and \( \beta_0 \) coefficients). The \( z_0 \) and \( \beta_0 \) coefficients also capture the effects of other “unique” characteristics of carriers, such as the quality of frequent flyer programs. We do not include service variables
other than on-time performance (e.g., baggage mishandling and overbooking) because these data are not available by route or by airport.

4.2. Data

We obtained data from the following US DOT data sources: Data Bank 28DM-T-100, Air Travel Consumer Report, and the US DOT web-site. We use the Atlanta Hartsfield Airport (ATL) – Chicago O’Hare Airport (ORD) route data for model calibration, because this route is one of the most competitive routes in US. The data consist of monthly observations of three carriers from 1990 to 1997. The total sample size is 288 (279 after adjusting for lag operations). The carriers included in the data set are American (AA), Delta (DL), and United (UA) airlines. They are the only carriers who served this route continuously from 1990 to 1997. Table 1 shows the descriptive statistics of the model variables.

It should be noted that between 1990 and 1997, carriers other than AA, DL, and UA were also present in the route, for some carriers entered into the route and some carriers abandoned (exited) the route during this time period. Thus, by including only the three carriers in our data set, we are ignoring the market share of other carriers. This approach, however, can be validated if we make the following assumptions: (1) when the new entrant comes into the route it takes shares from the three “existing” carriers in a manner proportional to their shares. (2) when a carrier exits the route this carrier’s share is distributed to the three carriers in a manner proportional to their shares. This is the constant ratio assumption that is often used in the aggregate market share models (e.g., Kalyanaram and Urban, 1992). This assumption is also quite similar to that used in the logit choice models which assumes that successive entrants draw shares from all earlier entrants proportional to their respective shares. The logit choice models are widely used in the air transportation literature for modeling passengers’ carrier choice behaviors (e.g., Nako, 1992; Ghobrial and Soliman, 1992; Proussaloglou and Koppelman, 1995; Yoo and Ashford, 1996).

4.3. Definition of measures

*Market share.* The market share is computed by dividing the total number of passengers transported by carrier *i* in month *t* in the route by the total number of passengers transported by the three carriers included in our data set in month *t* in the route. The total number of passengers transported by carrier *i* in month *t* is defined as the sum of (1) passengers enplaned at ATL and deplaned at ORD by carrier *i* in month *t*, and (2) passengers enplaned at ORD and deplaned at

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics*</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Market share (average)</td>
</tr>
<tr>
<td>On-time % (average)</td>
</tr>
<tr>
<td>Flight arrivals/month</td>
</tr>
<tr>
<td>Passenger fatalities/month</td>
</tr>
</tbody>
</table>

*Chicago (ORD) and Atlanta (ATL) figures combined (1/1990–12/1997). |

<sup>b</sup> AA = American, DL = Delta, UA = United.
ATL by carrier $i$ in month $t$. Since the US DOT data are not aggregated by the booking class level, the total number of passengers transported by a carrier in a given month equals the sum of this carrier’s passengers in all booking classes for that month.

**On-time performance.** This variable represents the overall percentage of flights arriving within 15 min of scheduled arrival time. Since on-time performance data are available by airport (not by route), we calculated, for each carrier, the weighted average of on-time performances at ATL and ORD for month $t$ by using the following formula:

$$Q_{it} = \frac{P_{it}(ATL \rightarrow ORD)Q_{it}(ORD) + P_{it}(ORD \rightarrow ATL)Q_{it}(ATL)}{P_{it}(ATL \rightarrow ORD) + P_{it}(ORD \rightarrow ATL)},$$

where $P_{it}(ATL \rightarrow ORD)$ is the number of passengers transported by carrier $i$ in month $t$ from ATL to ORD, $P_{it}(ORD \rightarrow ATL)$ is the number of passengers transported by carrier $i$ in month $t$ from ORD to ATL, $Q_{it}(ORD)$ is the on-time performance of carrier $i$ in month $t$ at ORD, and $Q_{it}(ATL)$ is the on-time performance of carrier $i$ in month $t$ at ATL.

**Airport dominance.** We measure the airport dominance of carrier $i$ in month $t$ by adding the number of flight arrivals realized by carrier $i$ in month $t$ at ORD and that at ATL.

**Safety Record.** As mentioned previously, the safety record of carrier $i$ is measured by the number of passenger fatalities experienced by the carrier (3-months moving average). This variable is not route specific – i.e., the variable represents the total passenger fatalities, regardless of the route, experienced by carrier $i$ in months $t$, $t-1$ and $t-2$. We attach a minus sign to this variable because increases in passenger fatalities decrease an airline’s attractiveness.

### 4.4. Method

We use the generalized least-squares method that addresses both the autocorrelation and the contemporaneous correlation of the residual terms. We assume that the residual terms follow an AR(1) process, i.e., that the model can be specified as follows:

$$y_{it} = f(\theta, x_{it}) + u_{it}, \quad u_{it} = \rho_i u_{it-1} + \varepsilon_{it}, \quad (15)$$

where $f(\theta, x_{it})$ represents the right-hand side of Eq. (12), and $\rho_i$ is an autocorrelation coefficient. Observe that Eq. (15) indicates that one autocorrelation coefficient is estimated for each airline. The model parameters and the autocorrelation coefficients can be estimated simultaneously by performing a non-linear least-squares regression on the following equation:

$$y_{it} = f(\theta, x_{it}) + \rho_i (y_{it-1} - f(\theta, x_{it-1})) + \varepsilon_{it}. \quad (16)$$

We minimize the sum of squared residuals (SSR) of Eq. (16) by using the Levenberg–Marquardt numerical optimization algorithm. To eliminate residual contemporaneous correlation, we combine the above method with the seemingly unrelated regression method. We impose the inequality parameter constraints (constraints (11c) and (13)) by adding a penalty function to the objective function (SSR). For a detailed discussion of the non-linear parameter estimation with a penalty function, see, for example, Bard (1974).

---

5 Strictly speaking, the on-time data by route are available from the US DOT. However, since the data do not go back to 1990, these data cannot be used in our study.
4.5. Results

Table 2 shows the estimation results. The results indicate that the switching rate of passengers who experienced flight delays ($\lambda_{Li}$) is consistently higher than that of passengers who did not experience delays ($\hat{\lambda}_{Gi}$) for all carriers. This condition implies that, once experiencing flight delays, passengers are more likely to switch airlines. The test of null hypothesis that $\lambda_{Li} = \hat{\lambda}_{Gi} \forall i$ (i.e., passengers are loss neutral) was rejected with a significant $p$-value ($p < 0.001$). This test result provides two important implications. First, passengers are likely to be loss averse with respect to on-time performance. Second, the proposed model provides a significantly better fit (significantly better mean squared error and adjusted $R^2$) than the model which does not incorporate the loss aversion effect (Eq. 5).

Table 2 also indicates that the attractiveness of an airline to the switchers with delay experience and that to the switchers without delay experience may be quite different. The results imply that the switchers without delay experience tend to choose airlines according to the degree of airport dominance (observe that $a_2$ is the largest-value $a$ coefficient and is statistically significant). This finding is consistent with those of the past air transportation studies. Many air transportation demand studies report that passengers tend to select carriers with higher airport dominance, for such carriers generally have better frequent flyer programs, computer reservation systems, and

Table 2
Estimation results

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Switching rate parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\lambda}_G$ (American)</td>
<td>0.3598</td>
<td>0.1184</td>
<td>3.04***</td>
</tr>
<tr>
<td>$\hat{\lambda}_G$ (Delta)</td>
<td>0.2901</td>
<td>0.0668</td>
<td>4.35***</td>
</tr>
<tr>
<td>$\hat{\lambda}_G$ (United)</td>
<td>0.3298</td>
<td>0.0862</td>
<td>3.83***</td>
</tr>
<tr>
<td>$\hat{\lambda}_L$ (American)</td>
<td>0.6114</td>
<td>0.0526</td>
<td>11.62***</td>
</tr>
<tr>
<td>$\hat{\lambda}_L$ (Delta)</td>
<td>0.3589</td>
<td>0.0401</td>
<td>8.96***</td>
</tr>
<tr>
<td>$\hat{\lambda}_L$ (United)</td>
<td>0.4346</td>
<td>0.0571</td>
<td>7.61***</td>
</tr>
<tr>
<td><strong>Attractiveness parameters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0$ (Carrier effect)</td>
<td>0.0284*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1$ (On-time)</td>
<td>0.0087</td>
<td>0.2606</td>
<td>0.03</td>
</tr>
<tr>
<td>$a_2$ (Airport dominance)</td>
<td>0.9542</td>
<td>0.4680</td>
<td>2.04**</td>
</tr>
<tr>
<td>$a_3$ (Passenger fatality)</td>
<td>0.0087</td>
<td>0.0450</td>
<td>0.19</td>
</tr>
<tr>
<td>$\beta_0$ (Carrier effect)</td>
<td>0.9028*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_1$ (On-time)</td>
<td>0.0389</td>
<td>1.6883</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_2$ (Airport dominance)</td>
<td>0.0315</td>
<td>1.8887</td>
<td>0.02</td>
</tr>
<tr>
<td>$\beta_3$ (Passenger fatality)</td>
<td>0.0268</td>
<td>0.1352</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Autocorrelation coefficients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$ (American)</td>
<td>0.9875</td>
<td>0.0166</td>
<td>59.52***</td>
</tr>
<tr>
<td>$\rho$ (Delta)</td>
<td>0.9449</td>
<td>0.0181</td>
<td>52.12***</td>
</tr>
<tr>
<td>$\rho$ (United)</td>
<td>0.9794</td>
<td>0.0285</td>
<td>34.36***</td>
</tr>
<tr>
<td>Sample size</td>
<td>279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fit statistic ($R^2$)</td>
<td>0.985</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* These parameters are specified as follows: $a_0 = 1 - a_1 - a_2 - a_3; \beta_0 = 1 - \beta_1 - \beta_2 - \beta_3$.

** Significant at the 95% significance level.

*** Significant at the 99% significance level.
closer relationships with local travel agents (Borenstein, 1989; Borenstein, 1991; Dresner and Windle, 1992; Nako, 1992). The airline choice of switchers with delay experience, on the other hand, appears to be impacted by the “carrier-specific” effect, but not by the on-time performance, airport dominance, or safety record (observe that $\beta_0$ is the largest-value $\beta$ coefficient and $\beta_1$, $\beta_2$, $\beta_3$ are not significant).

Our empirical results provide interesting information about the nature of the relationship between on-time performance and airline market share. Observe in Table 2 that $\beta_1$, $\beta_3$, $\beta_1$, and $\beta_3$ are not statistically significant. Therefore, passengers probably do not look at the on-time performance and the safety record of carriers when switching airlines. This pattern, however, does not indicate the lack of causal relationship between on-time performance and passengers’ choice of airlines (or market share). Recall that the switching rate of passengers with delay experience is significantly higher than that of passengers without delay experience. Thus, an airline’s on-time performance significantly affects the airline’s market share by changing the retention rate (or switching rate) of passengers for that airline. The on-time performance of an airline, however, appears to have minimal impact on attracting switchers from other airlines, as mentioned above. In short, our study results imply that on-time performance may affect market share only through the passengers’ experience (delay experience), and not through the “advertisement” of performance.

5. Conclusions and limitations

We have proposed an aggregate-level model of on-time performance and market share for air carriers. A unique aspect of the model is that it estimates the transition probability matrices separately for the passengers who experienced flight delays in month $t - 1$ and for those who did not experience delays. Because of this characteristic, the model can capture the loss aversion effect of on-time performance on market share. The model, when calibrated with the US DOT data, showed several interesting findings. First, passengers are likely to be loss averse with respect to on-time performance. The results indicate that, for all carriers included in our data set, the switching rate of passengers with delay experience is higher than that of passengers without delay experience. Second, when passengers switch airlines, their choices may be affected by the airport dominance, but not by other airline attributes. This pattern implies that airlines that possess greater dominance at origin and/or destination airports may attract more switchers. Third, the estimation results imply that on-time performance affects an airline’s market share through the passengers’ experience, but not the “advertisement” of performance. Note, however, that we do not propose generalization of these results, for the model was estimated by using only a single-route data. Data analysis for more routes and airlines is required in the future.

The model has its limitations. First, the model assumes that the switching rate parameters ($\lambda_{Li}$ and $\lambda_{Gi}$) are stationary (constants). It is possible, however, that a passengers’ decision to stay or switch from a particular carrier is not only a function of delay/non-delay experience, but also a function of other variables such as the airport dominance of the carrier in question. An interesting extension of the model, therefore, is to formulate a model where $\lambda_{Li}$ and $\lambda_{Gi}$ would be treated as non-stationary by expressing them as functions of some exogenous factors.

Second, our model’s definition of “on-time arrival” is rather arbitrary. In our model, the flights are considered as “on-time” if arriving at the destination airport within 15 min of scheduled
arrival time, but are considered as “delayed” otherwise. Although this definition is also used by the US DOT (and by many other airline studies: e.g., Dresner and Xu, 1995; Suzuki, 1998), it may not represent the true definition of “on-time arrival” as perceived by the passengers. The future research may want to explore this issue by estimating our model with alternative definitions of “on-time arrival” (e.g., 10 min behind schedule, 20 min behind schedule, etc.), and examining which definition provides the best fit to empirical data (provided that the on-time data for alternative definitions are available).

Third, the model assumes that the number of passengers who did not experience flight delays can be approximated by multiplying the market share and the on-time performance (e.g., \( S_{kt-1}Q_{kt-1} \) represents the passengers who used airline \( k \) in month \( t - 1 \) and did not experience flight delays). As mentioned previously, this assumption is valid only if airlines use approximately the same fleet sizes and/or loads in a given route. This condition may be true in many routes, for carriers generally use the fleets that match the route demand (e.g. use smaller airplanes for low-demand routes). This condition, however, may not hold for some routes and/or carriers. It may be necessary to test the validity of this assumption empirically in the future.

Fourth, the model assumes that all passengers have approximately the same inter-travel time (one month). Although this type of assumption is used in virtually all of the market share models (in particular the aggregate-level Markovian models), the model suffers from this assumption because the model is only as strong as the “one-month” inter-travel time assumption is valid (i.e., if the assumption does not hold, the study results may have to be interpreted differently). We made the “one-month” inter-travel time assumption by referring to both anecdotal and empirical evidence, but more empirical evidence may be needed to support the use of this assumption in empirical analyses.

In sum, our model should be considered as an aggregate model of airline demand and on-time performance that also provides, to some degree, insights into the passengers’ carrier switching and non-switching behaviors. The model does not substitute for disaggregate models of airline choice. Rather, our framework should be viewed as a “quick and dirty” method of approximating the passengers’ airline switching behaviors from the available (aggregate) data. Our approach, however, may become a useful tool for investigating the nature of the relationship between on-time performance and market share by using the available data.

Acknowledgements

The author would like to thank for the generous funding provided by The Pennsylvania State University’s Center for Logistics Research. Thanks also go to Brenda Lantz for providing many useful comments.

Appendix A. Satisfying constraints (7a) and (7b)

To show that all \( A^k_{Gil} \) and \( A^i_{Lil} \) are bounded between 0 and 1 (constraint (7a)), we only need to prove that \( A^k_{Gil} \geq 0 \) and \( A^i_{Lil} \geq 0 \) \( \forall k, i, t, \ (i \neq k) \) because if all \( A^k_{Gil} \) and \( A^i_{Lil} \) are non-negative, each of them cannot be larger than 1 because of the constraint (7b). To satisfy the conditions \( A^k_{Gil} \geq 0 \)
and $A^k_{Gli} \geq 0$, we impose the non-negativity constraints (constraint (11c)) on the $\alpha$ and $\beta$ coefficients of Eqs. (11a) and (11b). Since all of the variables in Eqs. (11a) and (11b) can take only the non-negative values (ratio of exponentiated variables), all $A^k_{Gli}$ and $A^k_{Lli}$ will be non-negative if the $\alpha$ and $\beta$ coefficients are non-negative. Thus, constraint (11c) satisfies the conditions of (7a), given that (7b) is true.

To discuss how to meet the conditions of (7b), first observe that all variables in Eqs. (11a) and (11b) have the property that sum of variable values across $i$ carriers ($i \neq k$) is 1:

$$\sum_{i \neq k} \left[ \frac{\exp(S_{it-1})}{\sum_{j \neq k} \exp(S_{jt-1})} \right] = 1, \quad \sum_{i \neq k} \left[ \frac{\exp(X_{git})}{\sum_{j \neq k} \exp(X_{hjt})} \right] = 1 \quad \forall h.$$  \hspace{1cm} (A.1)

Now, assuming that $\alpha$ and $\beta$ coefficients are constant across carriers, we can derive the following by utilizing (A.1) ($A^k_{Gli}$ is used in the following discussion, but the same logic applies to $A^k_{Lli}$):

$$\sum_{i \neq k} A^k_{Gli} = \sum_{i \neq k} \left[ \frac{\alpha_0 \exp(S_{it-1})}{\sum_{j \neq k} \exp(S_{jt-1})} \right] + \sum_{i \neq k} \sum_{h=1}^H \left[ \frac{\beta_h \exp(X_{hit})}{\sum_{j \neq k} \exp(X_{hjt})} \right]$$

$$= \alpha_0 \sum_{i \neq k} \left[ \frac{\exp(S_{it-1})}{\sum_{j \neq k} \exp(S_{jt-1})} \right] + \sum_{h=1}^H \left[ \alpha_h \sum_{i \neq k} \left[ \frac{\exp(X_{hit})}{\sum_{j \neq k} \exp(X_{hjt})} \right] \right]$$

$$= \sum_{h=0}^H \alpha_h.$$  \hspace{1cm} (A.2)

Thus, to satisfy the conditions of (7b) for $A^k_{Gli}$, all we need is to set $\alpha_0 + \alpha_1 + \cdots + \alpha_H = 1$ (constraint (11d)) and $\alpha_h = \text{constant across carriers}$ (constraint (11e)). Using the same procedure, it can be shown that, to satisfy the conditions of (7b) for $A^k_{Lli}$, we require that $\beta_0 + \beta_1 + \cdots + \beta_H = 1$ (constraint (11d)) and $\beta_h = \text{constant across carriers}$ (constraint (11e)). Hence, by imposing constraints (11c)–(11e), Eq. (11a) and (11b) satisfy the conditions (7a) and (7b).

References


US Department of Transportation Air Travel Consumer Report, various editions.

US Department of Transportation Data Bank 28DM-T-100, various editions.

