Discrete Choice Model for Amtrak Acela Express Revenue Management

Cinzia Cirillo*, Pratt Hetrakul*, and Shahab Toobaie**

*Department of Civil & Environmental Engineering
1173 Glenn L. Martin Hall, Bldg #088
University of Maryland, College Park, MD 20742
USA
Tel. +1 301 405 6864
Fax +1 301 405 2585
e-mail: ccirillo@umd.edu
e-mail: hetrakul@umd.edu

**Performance Measurement, Marketing & Product Development
National Railroad Passenger Corporation (Amtrak)
60 Massachusetts Avenue NE,
Washington DC 20002,
USA
Tel + 202 906 2433
e-mail: toobais@amtrak.com

Dr. Cinzia Cirillo is an assistant professor at University of Maryland, Department of Civil & Environmental Engineering. She received her M.S. and Ph.D. in Civil Engineering from University of Naples and University of Torino respectively. Her research interest includes advanced discrete choice analysis, application of advanced statistical and econometrics methods to travel demand forecasting, and activity based models. Her recent research work focus on stochastic programming to solve complex discrete choice models with simulations.

Pratt Hetrakul is a Ph.D. candidate in Transportation at the University of Maryland. He received his B.E. in Civil Engineering from Chulalongkorn University and M.S. in Transportation from the University of Southern California. His research interest includes discrete choice modeling in transportation and the application of discrete choice to revenue management problem with emphasis on dynamic discrete choice modeling of passenger behavior and stochastic programming for revenue management problem. He is the corresponding author of this paper.

Shahab Toobaie is the Principal Officer of Performance Measurement at National Railroad Passenger Corporation (Amtrak). He is also currently a Ph.D. candidate in transportation at the University of Maryland. He received his B.S. in electrical engineering from Amirkabir University, and a first M.S. in systems engineering and a second M.S. in transportation from IRPD and the University of Maryland respectively. His research interests include traffic equilibrium, vehicle routing, and the application of statistical modeling and operations research to passenger railroad.

Discrete Choice Model for Amtrak Acela Express Revenue Management

ABSTRACT

In this paper, we propose a pricing strategy for Amtrak Acela Express focusing on business class passenger departing from Washington D.C. A two step process is proposed to model passenger demand. In the first step, passenger choice model of booking time is estimated using a multinomial logit model. In the second step, a linear regression determines passenger demand in response to fare price; the effects of departure day of week, and destination specific are incorporated. The proposed models, estimated on ticket reservation data, are incorporated into a non-linear programming problem to maximize expected revenue. The results indicate a potential for significant revenue improvements.

Keyword: Acela Express, Amtrak, Discrete Choice, Multinomial Logit, Railway, Revenue Management

Paper submitted: January 27, 2011

First Revision submitted: July 3, 2011

Second Revision submitted: August 31, 2011

Paper prepared for: Journal of Revenue and Pricing Management

1. INTRODUCTION

Acela Express is Amtrak's high speed rail service that operates along the Northeast Corridor between Washington D.C. and Boston. The service has become popular among business travelers due to its relatively high speed and comfort. Acela Express offers two service classes, first and business. The business class fare varies depending on departure date, departure day of week, departure time, the time the reservation is made, and customer demand for each departure. Different passengers groups are also subjected to different discount policy such as seniors, children, military, and group travel. A team of revenue and pricing analysts manage Acela Express pricing and fare bucket inventory based on current demand and various historical statistics and reports. Their effort is aimed at maximizing Acela Express revenue for Amtrak.

The Revenue Management (RM) problem in railway involves several dimensions, from capacity control to fare pricing depending on the objectives of the operator. Railways typically compete with other services such as airlines, automobiles, ferries, or buses; thus the pricing of the railways is influenced by the prices and the availability of these transportation alternatives. In addition to the price of competing travel alternatives, pricing for train also depends on its service attributes such as train speed, time of operation, and travel distance.

The attempts toward better understanding of passengers' behavior in the RM context have generated a number of studies based on the theory of Discrete Choice Analysis (DCA). DCA is an emerging research area in RM that offers a means for analysts to incorporate behavioral components into their model systems. DCA is based on the concept of random utility maximization and assumes that passengers make a choice among the set of products offered by taking product attributes into consideration. In RM problems, Multinomial Logit (MNL) models (Ben-Akiva and Lerman, 1985) or more advanced version of discrete choice models (Train, 2003) can be used to assess the impact of product attributes on rail passenger decisions.

This paper proposes a new fare strategy for Acela Express that allows the fare prices to be updated on a daily basis. The proposed fare strategy is based on advance booking, departure day of week, and destination specific effects. The passenger choice of booking day is estimated using Multinomial Logit (MNL) under the assumption that passengers choose the day to book the ticket that maximizes their expected utility. The aggregate market demand for each destination is modeled with Ordinary Least Squares (OLS) regression. The passenger choice model and the aggregate demand function are incorporated into an optimization module. This model system is formulated as an expected revenue maximization problem that gives the optimal fare strategy for each destination on a particular departure day over the sale horizon.

The remaining of this paper is organized as follows, in Section 2 we review the literature related to the application of discrete choice model in RM; we focus on methodologies that are applicable to our problem. Section 3 describes the data set, choice model specification, and results. Section 4 is devoted to the passenger demand function estimation. The optimization formulation and the deriving pricing scheme are in Section 5. The revenues generated by applying the results of the optimization problem are then compared to the actual ones. Conclusions and suggestions for future research are given in Section 6.

2. LITERATURE REVIEW

Discrete choice analysis has recently been introduced in RM where its applications mainly focus on hotel and airline industry. The approach enables analysts to account for preferences in product attributes in the RM policy. The approach overcomes the restricted assumption of independent demand model which assumes that passenger demand is completely independent from the controls being applied by the seller. To date, there has been a limited number of studies on the practicality and the effectiveness of the discrete choice in the revenue management problem, especially for the railway industry.

Talluri and van Ryzin (2004) analyzed a single leg revenue management problem where the consumer choice behavior is modeled using a general discrete choice model. The probability of purchasing a product in the choice set is calculated as a function of the available fares. In their case study, purchase transaction data are used for estimation; attributes considered are fare price and indicator variables for product restrictions. An Expectation Maximization (EM) method is proposed to overcome the problem due to incomplete data where only purchase transactions are available. In this case, it is not possible to distinguish a period without a customer arrival from a period in which there was an arrival but the arriving customer did not make a purchase. The model is used to study the buy-up and buy-down behavior of airline products where the RM problem involves the decision of fare product to be offered at each point in time.

The feasibility and benefit of the discrete choice analysis in RM was examined in Vulcano et *al.* (2008). This study relies on a choice-based approach where Multinomial Logit (MNL) model is used to estimate the choice of buy up, buy down, and diversion for the airline market. The unobservable shopping data are incorporated in the maximum likelihood estimation using a variation of the Expectation Maximization (EM) method. The choice set in this model consists of all the flights offered by multiple airlines on a given day between specific pairs of airports including a non-purchase decision which is specified as zero utility. The proposed choice model accounts for flight arrival time of day using the convex weight of time period. The results provide revenue improvement in the range of 1.4% to 5.3% for the specific market under analysis. The authors also suggest testing a model with unobservable segment (latent class) to improve accuracy of the model prediction. The method proposed by Vulcano et *al.* (2008) has proven to be both practically feasible and economically beneficial over the current airline RM practice. However, it is reasonable to assume that passengers decide on other choice dimensions than just on fares for a given departure time. For instance, Van Ryzin (2005) indicated that time to demand for a product is potentially a strategic aspect that customers consider when making ticket reservation.

Iliescu et *al*. (2008) estimated passenger cancellation model based on ticketing data. Two modeling approaches are compared, time-to-event model and MNL model. The dependent variable in the model is specified as the passenger’s choice of exchanging, cancelling, and keeping the original ticket. The survival method in time-to-event model exhibits features appropriate for the airline industry in two aspects, the presence of censored data, and the ability to incorporate time varying covariates. The MNL model provides good data fit and is consistent with the hypothesized scenario.

Comprehensive demand models are essential for the evaluation of policy measures such as dynamic pricing strategy and capacity utilization. In order to fulfill this need, methods based on dynamic interaction with the supply control should be developed. Whelan and Johnson (2003), and Whelan *et al*. (2005) estimated a Nested Logit model to evaluate the impact of fare structure on train overcrowding. The study is based on cross sectional Revealed Preference (RP) and Stated Preference (SP) data collected for the British Strategic Railway Authority. The model structure presents a lower nest corresponding to passenger's choice of ticket types and an upper nest corresponding to the decision of whether to travel by railway or not. Some of the interesting attributes describing ticket types are departure time restrictions and advance purchase requirements. Results show that SP data support the purchase, non-purchase behavior in the upper nest plausibly.

Amtrak is one of the few large passenger railways that are known to actively use RM techniques. Sibdari et *al.* (2007) studied dynamic pricing policy for Amtrak Auto Train. A revenue management model was developed for this service that allows passengers to bring their vehicles on the train. The method proposed relies on discrete-time multi product dynamic pricing model which is suitable for price policy and is updated on a daily basis. The choice model involves a multi stage decision process similar to the model proposed by Whelan and Johnson (2003) and by Whelan *et al*. (2005); passengers make the decision to buy or not to buy and whether to upgrade the accommodation or not. The data indicate that the relationship between time before departure and the average daily demand can be approximated by an exponential function. The analysis reveals that there was almost no reservation activity until 30 days before departure given a sale horizon of 330 days. In this analysis, passenger demand is specified to follow a Poisson random variable with specified mean where a passenger demand on a given day is a function of remaining time before departure, car accommodation price, and coach seat price.

3. PASSENGER CHOICE MODEL

3.1 Data Description

The ticket reservation data of Amtrak Acela Express are used for the analysis. Data were obtained from the Marketing and Product Development department at the National Railroad Passenger Corporation (Amtrak). The data contain booking information related to trips departing in March and April, 2009. In this study, we focus on business class passengers traveling from Washington D.C. to other stations in the Northeast Corridor; only reservations that were not cancelled and eventually contributed to the actual revenue are focused in our analysis. The data contain information in terms of trip origin, trip destination, fare class, reservation date, departure date, departure time, arrival time, fare price, and accommodation charge. The data recorded in March, 2009 which consists of 44,847 reservation records are used to estimate the passenger choice model and the demand function. The actual demand data recorded in April, 2009 are used to test the revenue under the new fare strategy proposed.

Our analysis does not account for seasonal effect such as Easter which fell in April, 2009 data (April 12 and April 19, 2009) mainly because the choice model and the demand function are estimated from March, 2009 data. Although we used April, 2009 data to test the revenue under new fare strategy, we did not account for seasonal effect from Easter in the analysis since we believe that the majority of the Acela Express passengers are business travelers whose traveling choices are not significantly impacted by Easter.

3.2 Multinomial Logit (MNL) Model Choice Design

We model passenger choice behavior by assuming that passengers maximize their utilities when choosing the day to book Acela Express tickets. Based on the data set, about 98 percent of the reservations occur no earlier than 30 days before departure, thus we assume the length of the sale horizon to be 31 days. This resulted in a choice set of 31 alternatives from booking day 1 (30 days before departure) to booking day 31 (departure day). We also assume that passengers know about their trip schedule 30 days in advance and have perfect information about the fare price evolution over the booking period. Consistently with the MNL formulation, individual customers' utilities for each alternative are assumed to be random variables. In our context, a set of booking day alternatives is denoted by *N*. For each passenger *i*, the utility of booking a ticket on day *n* assumes the form:

$$U_{in} = v_{in} + \varepsilon_{in}, \qquad (1)$$

where v_{in} is a deterministic term also called expected utility or nominal term. A random component ε_{in} is a mutually independent noise term following a Gumbel distribution. The expected utility is generally modeled as a linear in parameters combination of observable attributes,

$$v_{in} = \beta^T x_{in}, \qquad (2)$$

where β is an unknown vector of parameters to be estimated and x_{in} is a vector of attributes (explanatory deterministic values) of passenger *i* such as fare price. In our study, the booking choice model aims at representing the individual behavior in response to fare policy; passengers might decide to shift their booking day but are not allowed to leave the market. This model is essential for the revenue management problem based on fare optimization that will be discussed in the following Sections. Given the distinct service of Acela Express, we focus on a monopolistic model that does not account for service competition.

3.3 Passenger Choice Model Specification

We use ticket reservation data related to business class passengers traveling in March, 2009 to estimate the choice model specified in Section 3.2. There are three main difficulties in estimating choice model from the ticket reservation data. First, the ticket reservation data do not contain passengers' socioeconomic characteristics; therefore, individual specific factors influencing booking decisions such as trip purpose and personal income cannot be observed, these factors are believed to significantly influence booking behavior. Second, the only two available alternative specific attributes in the ticket reservation data are advance booking (number of day before departure), and fare price ($). Third, ticket reservation data only contain fare price information on the day that passengers make the reservation; the fare price that will be offered to the passengers on other days in the sale horizon is not known to the analyst. In order to overcome this limitation, the fare price on each booking day and for a particular destination, is computed by averaging the observed fares on each booking day in March, 2009.

The initial model specification consists of two generic parameters: advance booking (number of day before departure), and fare price ($). This specification resulted in a model with a positive coefficient for price because the majority of the Acela Express passengers book the ticket close to the departure date when fare prices are relatively high. However, it is reasonable for passengers to be insensitive to fare price on booking day close to departure given that they are more concerned about obtaining the ticket than about fare price. To address this problem, we allow passengers to have different price sensitivity on each booking day by specifying price estimates differently for each of the 31 booking days. This approach is similar to the Day From Issue (DFI) estimation proposed by Iliescu et *al.* (2008) in their Discrete Time Proportional Odds (DTPO) model. However, this specification is too complex and is not supported by the data; many parameters are not significant and the interpretation of the estimation results is difficult. To resolve this problem, we group the 31 booking days into 6 booking periods and define different price coefficients for each booking period in the final model. The booking periods are grouped such that booking days within the same booking period have approximately the same number of reservations. These 6 booking periods denoted by k are: (1) Booking day 1 to booking day 11, (2) Booking day 12 to booking day 20, (3) Booking day 21 to booking day 25, (4) Booking day 26 to booking day 29, (5) Booking day 30, and (6) Booking day 31.

The independent variables included in the final model are advance booking (number of days before departure), fare price ($), destination specific dummies, and long distance dummy. The advance booking coefficient enters the model as a generic parameter while the fare price, destination specific dummies, and long distance dummy enter the model as booking period specific parameters. Destination specific dummies and long distance dummy are included to account for passenger heterogeneity across different markets. Accounting for destination specific effects is motivated by Iliescu et *al*. (2008) which shows promising results in terms of significance of the estimates. In our study, these destination markets are chosen from high demand markets which are believed to exhibit specific effects toward passengers' booking behavior. A high demand market is assumed to have limited seat capacity; passengers are more likely to book the ticket in advance to make sure that a ticket is available on the day they travel. Several destinations have been tested in model calibration for their significance in explaining passengers' behavior. Three stations are found to significantly influence choice behavior: Boston South Station (BOS), New York Penn Station (NYP), and Philadelphia 30th street station (PHL). The effect of long distance is motivated by Whelan et *al*. (2008). In their study, it is found that leisure trips are in general long distance trips. In our context, by taking into account long distance trip variable we expect to capture specific effects deriving from trip purpose, and associated unobservable factors such as trip flexibility. For this problem, long distance trips are assumed to have travel time greater than or equal to two hours.

The resulting utility of passenger i booking the ticket on day n which falls within the booking period k can be expressed as:

$$\begin{aligned} U_i(n,k) &= (\beta_{adv} \times advbking(n)) + (\beta_{fare}^k \times fare(n)_i) + (\beta_{BOS}^k \times BOS_i) \\ &+ (\beta_{NYP}^k \times NYP_i) + (\beta_{PHL}^k \times PHL_i) + (\beta_{long}^k \times LONG_i) + \varepsilon_i, \end{aligned} \tag{3}$$

where the independent variables and their associated index are:

n = Booking day, $n \in \{1,\ldots,31\}$

k =Booking period, $k \in \{1,\ldots,6\}$

$advbking$ = Advance booking (number of day before departure)

$fare$ = Fare price ($)

BOS = Boston destination dummy (1 if trip destination is BOS, 0 otherwise)

NYP = New York destination dummy (1 if trip destination is NYP, 0 otherwise)

PHL = Philadelphia destination dummy (1 if trip destination is PHL, 0 otherwise)

$Long$ = Long distance dummy (1 if trip travel time ≥ 2 hours, 0 otherwise)

ε_i = A mutually independent noise term following a Gumbel distribution

The probability of passenger i booking on day n can be calculated by using the logit probability formulation as:

$$\Pr(bookingday = n) = \frac{\exp[U_i(n,k)]}{\sum_{m=1}^{31} \exp[U_i(m,k)]} \tag{4}$$

3.4 Choice Model Results and Interpretations

The results obtained from the choice model calibration are reported in Table 1; most of the estimates have the expected sign and are statistically significant. However, destination specific parameters are not significant in some booking periods.

Fare price estimates show high statistical significance and the expected sign. The monotonically increasing value of price estimates from booking period 2 to booking period 6 indicates that passengers become less price sensitive as time approaches departure which is in line with the expectation. Specifically, in booking period 6, the small positive price estimate indicates that passengers are insensitive to price on the day of departure. This result is reasonable for the Acela Express service, given that the majority of the passengers are business oriented travelers who book the ticket close to the departure date and are not very sensitive to fare price. The smaller magnitude of the price estimate in booking period 1 compared to other booking periods (2 to 5) could be explained by the relatively low number of passenger booking in this period.

The Boston dummy is coherent with the expected pattern. The monotonically decreasing value of this variable with respect to booking period implies that it is preferable to book the ticket for Boston as early as possible to ensure the availability of the seat. Boston is a popular long distance market, with travel time by railway comparable to travel time by plane.

The New York destination shows an opposite trend to the one observed for Boston, the increasing value of the estimates with respect to booking time implies that passengers prefer to book the ticket closer to the departure date preferably in booking period 5 and 6 respectively. Philadelphia destination follows the same pattern as that of Boston, the most preferable booking periods being respectively periods 2 and 1.

Long distance variable shows statistically significant estimates except for booking period 3. The long distance estimate pattern is in line with the expectation; the earlier booking periods being more preferable to travelers, with booking period 2 being the most preferred. This could be explained by the fact that driving to these long distance destinations is onerous and that traveling by bus is relatively time consuming and uncomfortable. It is then sensible to assume that passengers book the ticket for these destinations early enough to ensure the availability of seats.

The advance booking parameter indicates that it is generally more preferable to book and to pay the ticket as late as possible. The advance booking coefficient is statistically significant and has a negative sign indicating a strong preference toward late booking.

"Table 1 about here"

The choice model has been validated on the same set of stations to which the fare optimization procedure is applied (as explained in Section 5). Model predictions from the model estimated on the entire set of data and observed choice for different time periods have been compared in the within sample validation (Table 2). A random sample of observations extracted from March 16 to March 22, 2009 has been selected for a hold out sample validation (Table 3). Results indicate that the choice model performs better for stations with large sample size such as Station 5, 6, 7, and 8 especially on booking day 31. Given that smaller stations only account for less than 5% of the entire market, it can be said that model predictions reproduce reasonably well the distribution of the demand over the sale horizon.

"Table 2 about here"

"Table 3 about here"

4. PASSENGER DEMAND FUNCTION

In this paper, the demand function is estimated to predict the aggregate passenger volume for each destination market. The demand function is estimated with Ordinary Least Squares (OLS) regression by taking into account product attributes. Although the demand function provides similar level of aggregation to the choice model, the estimation procedure adopted is different. In choice model estimated with Multinomial Logit model, the observation used in the analysis is an individual reservation record. For each reservation record, it is reasonable to assume dependency between choices given that only one booking day choice can be chosen among alternatives. In demand function estimated with OLS, the observation used in the analysis is the passenger volume on each booking day. The objective of the demand function is not to predict the day in which passengers book the ticket, but rather to predict the aggregate demand volume based on the fare price (along with other trip attributes) on each booking day. Thus, the demand of one booking day estimated from OLS does not necessarily influence the demand of other booking days. The aggregate demand for a particular destination market is the sum of estimated demand on each booking day over the sale horizon.

Other than the constant, the independent variables included in the final model are advance booking square, departure day of week dummies, fare price, and booking day specific dummies. The advance booking square is used to represent the non-linear relationship between demand and advanced booking observed from the data set. Initially, we estimated one model for each destination using the same specification. However, with this approach, the relatively high booking demand close to departure associated with relatively high fare price results in a model with a positive price estimate. This is because, unlike the classical demand model, fare price is not a completely independent variable. Amtrak Revenue Management periodically changes fare prices in response to the demand to maximize Acela Express revenue.

To address this problem, we group the booking day into 5 booking periods and estimate 5 independent demand functions for each destination market. These 5 booking periods denoted by k' are: (1) Booking day 1 to booking day 11, (2) Booking day 12 to booking day 20, (3) Booking day 21 to booking day 25, (4) Booking day 26 to booking day 29, and (5) Booking day 30 to booking day 31. The reason we have 5 booking periods in the demand function while we have 6 booking periods in the choice model is because in the demand function, each booking day only represents a single passenger demand volume. On the other hand, in the choice model, each booking day has multiple reservations; each reservation represents one observation of passenger booking choice. This results in a relatively low number of observations for booking day 30 and 31 in the demand function. Therefore, we combine these two booking days into one period (booking period 5) to allow for sufficient sample size in the model estimation. With this approach, we assume that, in each booking period, the price is almost independent of demand and can be used as an independent variable in the demand function. This assumption is not far from the reality, as Amtrak price changes happen in piecewise manner. In other words, fare prices do not change continuously and instantaneously in response to any small fluctuation in demand. On the other hand, the demand responds to changes in the fare price continuously and instantaneously. This approach allows us to compare the demand of the booking period subjected to different fare price throughout the month and to obtain a service demand that is sensitive to price within the booking period. The passenger demand on booking day n which falls into booking period k' takes the form:

$$\begin{aligned} Demand(n,k') = {} & \alpha_0^{k'} + (\alpha_{advsq}^{k'} \times advbking^2) + (\alpha_{mon}^{k'} \times MON) + (\alpha_{tues}^{k'} \times TUES) \\ & + (\alpha_{wed}^{k'} \times WED) + (\alpha_{thurs}^{k'} \times THURS) + (\alpha_{fri}^{k'} \times FRI) + (\alpha_{sat}^{k'} \times SAT) \\ & + (\alpha_{fare}^{k'} \times fare(n)) + \sum_{i \in k'} (\alpha_{bkday\_i}^{k'} \times BookDay\_i) + \varepsilon, \end{aligned} \tag{5}$$

where: $advbking$ = Advanced booking (number of day before departure)

$MON,...,SAT$ = Departure day of week dummies

$BookDay\_i$ = Booking day specific dummies

ε = Error term

Given that unrealistic results were obtained for some stations, the methodology is applied to 8 out of the 15 destinations; these stations are renamed from Station 1 to Station 8 for data confidentiality reasons. Due to space limitations, we only show the demand functions estimated for 3 out of these 8 destinations (see Table 4 to Table 6) to illustrate our results. Note that the estimated demand obtained from the demand function is unconstrained meaning that it is not bounded by a particular value and therefore allows for the latent demand in the revenue estimation. Therefore, in periods where historically it was at capacity, the estimated demand obtained from demand function does not have such restriction. However, the capacity is accounted in the objective function.

The fare price estimates have the expected sign for the majority of the models. The square of the advanced purchase could only be included in some of the models estimated due to the difficulties encountered in applying the proposed regression procedure. Note that for each booking period, the number of booking day specific dummies is equal to number of booking day in the period minus one. This is because the booking day specific dummy of the last booking day in the period is absorbed in the constant term. However, in the case that no reservation is observed on a particular booking day from the sample, it is not possible to estimate the booking day specific dummy. The station which shows the best model fit is Station 5; this can be explained by the large sample available for this market. The results for booking period 5 across all destinations show the best model fit when compared to other booking periods for the same destination. For this booking period, the fare strategy does not significantly vary across observations, which results in the demand to be relatively stable and therefore in a good model fit.

"Table 4 about here"

"Table 5 about here"

"Table 6 about here"

The validation of the demand function on a hold-out sample extracted from the period March 16 to March 22, 2009 is reported in Table 7. Results indicate that the volumes of ticket sales predicted are comparable to the actual demand.

"Table 7 about here"

5. FARE OPTIMIZATION

5.1 Optimization Procedure

The optimization procedure for revenue maximization incorporates both the passenger choice models based on discrete choice models and the demand functions obtained from the linear regressions. By assuming that Acela Express aims to maximize revenue from ticket sales, a fare strategy that can be updated on a daily basis is proposed. With this approach, day specific fare prices within the sale horizon are strategically calculated from this optimization problem.

In this specific case study, we optimize the fare prices over a representative week in March (March 16 to March 22, 2009). The booking choice models and the passenger demand functions estimated from ticket reservation data in March, 2009 are incorporated in the fare optimization problem to represent passenger response to RM policy. The fare strategy resulting from the optimization process is compared to the current fare policy by comparing "model" revenue to the real revenue registered from March 16 to March 22, 2009. However, in reality, the earliest departure day in which this new fare strategy can be applied is 31 days after the prices are computed. Therefore, we also test the revenue in April, 2009 by imposing these fare strategies to the week of April 20 to April 26, 2009 and we assess the performance of the fare price estimated in March, 2009. This allows us to test how the fare prices estimated from sale data in one month performs the following month.

5.2 Problem Formulation

Revenue optimization is formulated by maximizing the revenue of each station for each departure day. The problem is formulated as an expected revenue maximization problem:

$$\max_{fare_n} \text{Revenue}=\left[\min\left[\sum_{n=1}^{31} Demand_n(fare_n), Capacity\right]\right]\times\left[\sum_{n=1}^{31}\left[fare_n \times \Pr(day\_n)\right]\right] \qquad (6)$$

The first term represents the accommodatable demand volume on a particular departure day. The accommodatable demand is calculated as the minimum between the predicted demand and the train capacity to ensure that only demand within the capacity contributes to revenue. The predicted demand is calculated by summing the estimated demand obtained from the demand function over the entire sale horizon. The train capacity for each destination market is approximated with historical sale data by assuming that actual demand in March, 2009 was at 80 percent load factor. We do not account for the capacity redistribution in this analysis since we treat the fare optimization for each origin destination pair as an independent problem. The second term in equation (6) represents the expected fare price expressed as the day specific fare $(fare_n)$ weighted by the probability that passengers book the ticket on that booking day ($\Pr$(day *n*)). Thus, the overall formulation represents the expected revenue per day for each destination with the day specific fare $(fare_n)$ as decision variables. We assume that Acela Express fare strategy is subjected to the predetermined fare bound restriction and that the fare prices only increase monotonically as time approaches departure. The incremental amount to fare price from one day to the next is assumed to be within the bound limit (assumed to be \$ 5.00). The assumptions used in this research do not necessarily represent the actual Acela Express RM policy. The corresponded constraints for problem in equation (6) according to our RM control assumptions are:

$$fare_{lb} \le fare_n \le fare_{ub} \qquad (7)$$

$$fare_n \leq fare_m \text{ ; for all } m > n \tag{8}$$

$$fare_m - fare_n \leq incremental\_allowance(\$) \text{ ; for all } m \geq n \tag{9}$$

The first constraint imposes bounds on fare prices for each destination. These bounds are assumed to be the maximum and the minimum of the average day specific fare prices recorded for the entire sale horizon in March, 2009. The second constraint ensures that the new fare price increases monotonically with respect to booking day, while the last constraint ensures that the increment of price on each day does not exceed the incremental allowance. The classifications of all the variables are:

$fare_n$ = Real Decision variable $\in R_+^{31}$

$fare_{lb}$, $fare_{ub}$ =Lower and upper bound on fare price of each destination respectively

In order for the objective function to represent expected revenue realistically, when the estimated demand exceeds capacity, the unmet demand must not concentrate only in the later part of the sale horizon, otherwise the high fares subjected to this later period (due to monotonicity assumption) would not be able to contribute to expected fare calculation in the objective function because the capacity is already depleted. Therefore, we assume that the unmet demand is not only concentrated in the later part of the sale horizon, but instead spread evenly over the sale horizon. To accommodate this assumption, the booking limit on each booking day is imposed on a first come first serve basis which results in the total unmet demand to be spread evenly across the sale horizon. The booking limit is set using the information on the estimated demand and on the corresponding unmet demand.

5.3 Optimization Results

The optimization problem in equation (6) is solved as a non-linear programming problem with LINGO 12.0, the optimization software by Lindo System Inc. (Lindo System Inc., 2010). The non-linearity nature of this problem is influenced by the probability function of the MNL model in equation (4). The corresponding fare prices by day of week and for each destination are shown in Figure 1 to Figure 8 for Station 1 to Station 8 respectively.

"Figure 1 about here"

"Figure 2 about here"

"Figure 3 about here"

"Figure 4 about here"

"Figure 5 about here"

"Figure 6 about here"

"Figure 7 about here"

“Figure 8 about here”

The results obtained will be discussed by destination station and by day of week (Figure 1 to Figure 8). The fare strategy by day of week for Station 1 depicted in Figure 1 conforms to the associated demand function in Table 4. For instance, Table 4 shows that demand reaches the highest peak on Wednesday and the lowest on Saturday particularly in the booking period 1 (Day 1-11), booking period 2 (Day 12-20), and booking period 4 (Day 26-29). The fare strategy in Figure 1 suggests that the fare price to be charged should be higher on the day with high demand (Wednesday) and lower on the day with low demand (Saturday) throughout the sale horizon. The changes in the demand function across different booking periods due to day of week effects also influence the price strategy. For instance, when we compare between Monday and Wednesday for Station 1, in booking period 2 (Day 12-20) Wednesday is shown to have greater demand compared to Monday, thus the corresponding fare price in Figure 1 in the beginning of the booking period 2 (Day 12-20) is lower than on Wednesday. However, in booking period 3 (Day 21-25), Monday shows a higher passenger number than Wednesday, thus the optimization takes this effect into account and the fare price for Monday starts increasing before the end of booking period 2 (Day 17) and finally matches Wednesday fare at the beginning of the booking period 3 (Day21). Other stations’ fare strategies show a similar pattern to the one occurred in Station1.

For Station 2, the departure days with high demand obtained from the associated demand function are Monday, Wednesday, and Friday. According to our results, Monday shows high demand in booking period 2 to 5 (Day 12-31), Wednesday shows high demand in the booking period 1 (Day 1-11), and Friday shows high demand in booking period 1, and 2 (Day 1-20). The lowest demand is on Saturday in booking period 1 (Day 1-11) and 4 (Day 26-29). Thus, the fare strategy by day of week in Figure 2 suggests that the highest fare is applied to Monday, Wednesday, and Friday and the lowest to Saturday.

For Station 3, the departure day with the highest demand is Friday and booking periods 3 to 5 (Day 21-31). Our results indicate that the lowest demand is on Saturday in booking period 1 (Day 1-11), 3 (Day 21-25), and 5 (Day 30-31). Thus, the fare strategy by day of week in Figure 3 suggests applying the highest fare on Friday and the lowest fare on Saturday.

For Station 4, the departure day with high demand is Wednesday and booking periods 2 (Day 12-20) and 4 (Day 26-29). The lowest demand is calculated for Saturday in booking periods 1 (Day 1-11), 3 (Day 21-25), 4 (Day 26-29), and 5 (Day 30-31). The fare strategy by day of week in Figure 4 imposes the highest fare to Friday and the lowest fare to Saturday.

For Station 5, the departure day with high demand (Table 5) is Tuesday and booking periods 2 (Day 12-20) and 5 (Day 30-31). In Figure 5, it can be seen that the highest fare is applied to Tuesday and the lowest fare on Saturday.

For Station 6, the departure day with high demand is Thursday and booking periods 1 (Day 1-11), 3 (Day 21-25), 4 (Day 26-29), and 5 (Day 30-31). According to our result, the lowest demand is calculated on Saturday from booking period 2 to 5 (Day 12-31). The optimal fare strategy in Figure 6 suggests applying the highest fare on Thursday and the lowest fare on Saturday.

For Station 7, departure days with high demand obtained from the associated demand function is Tuesday with high demand in booking period 2 (Day 12-20) and 5 (Day 30-31). According to our result, the lowest demand is shown to be Saturday in all booking periods. Thus, the fare strategy by day of week in Figure 7 suggests the highest fare on Tuesday with the lowest fare on Saturday.

For Station 8, Thursday is shown to have the highest peak in booking period 3 to 5 (Day 21-31) (see Table 6). The lowest demand is predicted on Saturday for all booking periods. However, due to the relatively low fare price for this station, the fare gap imposed by the optimization problem is smaller when compared to nearby stations. The day of week effect is not sufficient to influence the fare strategy by day of week (Figure 8).

These findings support the intuition that high demand days have a larger impact on revenue maximization than low demand days. In particular, the departure day of week in the demand function is specified as a constant term while the price coefficient accounts for passenger sensitivity to fare. This constant term of departure day of week could be viewed as an intercept term in the regression model. When the day of week intercept is relatively high compared to other days within the week, the solution deriving from the optimization problem suggests relatively higher fare compared to other days. The decrease in passenger demand due to higher fare in this case is outweighed by the higher demand based on day of week effect; charging higher fare price on this day compared to other days within the week contributes to higher daily revenue.

The revenue improvements for each departure day in March (March 16 to March 22) and April (April 20 to April 26) representative week are shown in Table 8 and 9. The total revenue improvement includes the stations which we do not apply the optimization and provide null revenue improvement. The total revenue improvement ranges from 1.92 to 13.76 percent and from 0.65 to 10.60 percent for March and April representative weeks respectively. Table 8 and 9 indicate that the application of the proposed fare strategy results into significant revenue improvement on Monday, Tuesday, and Saturday.

"Table 8 about here"

"Table 9 about here"

6. CONCLUSIONS

In this paper, we have analyzed a fare pricing strategy for the Acela Express service operated by Amtrak. The RM method proposed is based on passenger preferences and product attributes. Using ticket reservation data, a MNL model has been calibrated; random utility theory has been applied to explain passengers' choice of booking time under a range of hypothetical sale horizons. In order to capture aggregate passengers' response to fare price, a demand function based on OLS regression has been incorporated in the procedure. This approach is appealing because it allows product attributes such as departure day of week, fare price and destination specific effects to be taken into account in the RM problem. The two models are incorporated into a mathematical formulation that maximizes the expected revenues for each departure day and for each destination market.

Our analysis provides a method for estimating choice behavior and passenger demand in response to RM strategies from readily available ticket reservation data. The proposed modeling framework appears to be promising; the deriving pricing strategy leads to a potential increase in revenues ranging from 1.92 to 13.76 percent and from 0.65 to 10.60 percent per day within the weeks of March and April respectively. However, it should be noted that, as with any academic work, the model is based on some simplifying assumptions which might not fully represent the real world problem. For example, Amtrak pricing strategy is more complicated than what is presented in this paper. We do not account for cancellation behavior, various discounts, guest reward program, special fare plans or competition with non-Acela trains or other modes of transportation.

Based on the experience acquired with this case study, several research directions are suggested. The new pricing strategy should be tested in terms of market acceptance and pricing response. Due to lack of socioeconomic information from our ticket reservation data, it would be desirable to calibrate a latent class model by identifying different passenger segments in terms of trip purpose or socioeconomic characteristics. The choice model proposed in this study only handles deterministic heterogeneity. Mixed logit model could be adopted to address random heterogeneity in customer behavior. Both latent classes and random coefficient logit models have the potential to improve the accuracy of the customer choice model. In term of fare optimization, the possibility to introduce capacity redistribution is suggested. In this case, the optimization problem has to account for all the origin destination pairs which share the same train capacity. This approach is expected to be very effective on revenue maximization.

To conclude, our ticket reservation data can be used to study ticket booking behavior for high quality railway services. The optimization routine based on choice behavior and different time horizons could be adopted by other operators that sell products on-line (i.e. shippers, couriers).

REFERENCES

Ben-Akiva, M.E. and Lerman, S.R. (1985) Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge: MIT Press.

Iliescu, D.C., Garrow, L.A., and Parker, R.A. (2008) Hazard models of US airline passengers' refund and exchange behavior. Transportation Research B 42, pp. 229–242.

Lingo User Manual (2010). Lindo System Inc., Chicago, IL. http://www.lindo.com/

Sibdari, S., Lin, K. Y., and Chellappan, S. (2007) Multiproduct revenue management: An empirical study of Auto train at Amtrak. Journal of Revenue and Pricing Management Vol0, pp.1-13.

Talluri, K. T. and van Ryzin, G. J. (2004) Revenue management under a general discrete choice model of consumer behavior. Management Science, 50, pp.15-33.

Talluri K. T. and van Ryzin G. J. (2004a). The Theory and Practice of Revenue Management. Kluwer, New York, NY, pp. 561-563.

Train, K.E. (2003) Discrete Choice Methods with Simulation: Cambridge University Press, pp. 38-47.

Van Ryzin. (2005) Models of demands. Journal of Revenue and Pricing Management, 4(2), pp.204-210.

Vulcano, G., van Ryzin, G., and Wassim, C. (2008) Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. Revised: January, May and June 2009. Forthcoming in *MSOM*.

Whelan, G. and Johnson, D. (2004) Modeling the impact of alternative fare structures on train overcrowding. International Journal of Transportation Management, 2(1), pp.51-58.

Whelan, G., Toner, J., Shires, J., Batley R., Wardmand, M., and Nash, C. (2008) Optimal Fare Regulations for Britain's Railways. Transportation Research Part E 44, pp. 807-819.