Passenger Demand Model for Railway Revenue Management

The Pennsylvania State University ❖ University of Maryland
University of Virginia ❖ Virginia Polytechnic Institute and State University
West Virginia University

The Pennsylvania State University ❖ The Thomas D. Larson Pennsylvania Transportation Institute
Transportation Research Building ❖ University Park, Pennsylvania 16802-4710
Phone: 814-863-1909 ❖ Fax: 814-863-37073
www.pti.psu.edu/mautc

PASSENGER DEMAND MODEL FOR RAILWAY REVENUE MANAGEMENT

FINAL REPORT

DTRT07-G-0003

Prepared for

University of Maryland

By

Dr. Cinzia Cirillo, Principal Investigator
Pratt Hetrakul, Graduate Research Assistant

Department of Civil & Environmental Engineering
1173 Glenn L. Martin Hall, Bldg #088
University of Maryland, College Park, MD 20742

January 2011

MAUTC-2009-01

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
MAUTC-2009-01		

4. Title and Subtitle Passenger Demand Model for Railway Revenue Management	**5. Report Date** January 2011 **6. Performing Organization Code**
7. Author(s) Dr. Cinzia Cirillo Pratt Hetrakul	**8. Performing Organization Report No.** MD-10-SP808B4A
9. Performing Organization Name and Address University of Maryland Department of Civil and Environmental Engineering 1173 Glenn Martin Hall College Park, MD 20742	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTRT07-G-0003
12. Sponsoring Agency Name and Address US Department of Transportation Research & Innovative Technology Admin UTC Program, RDT-30 1200 New Jersey Ave., SE Washington, DC 20590	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code** **(7120) STMD-MDOT/SHA**

15. Supplementary Notes

16. Abstract

In this paper, we have illustrated a fare pricing strategy for the Acela Express service operated by Amtrak. The RM method proposed is based on passenger's preference and products' attributes. Using sales data, a MNL model has been calibrated; the 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 in 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 booking data. The accuracy of the estimates depends on the market size; for instance, the model produces good results for station5 market which is the predominant market for Acela Express. Overall, we show that the proposed model in this paper is promising and can potentially lead to increase in revenue. It was demonstrated that the pricing strategy which accounts for choice behavior could potentially increase the revenue from 2.06 to 14.64 percent and 0.70 to 11.60 percent per day within the respective weeks of March and April. However, it should be noted that, as with any academic work, the model is based on some simplifying assumptions which might not fully comply with the real world problem. For example, Amtrak pricing strategy is more complicated than what presented in this paper. We did not account for cancellation behavior, various discounts, guest reward program, special fare plans or competition with non Acela trains or other modes of transportation. Also, the choice model is not tested independently to show if it accurately reflects customers' choice behavior in the market. So there is significant room to improve or extend this research.

Several research extensions 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 sales 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 model calibrated handles deterministic heterogeneity only. Mixed logit models could be adopted to address random heterogeneity in customer behavior. Both latent classes and random parameters logit models have the potential to improve the accuracy of the customer choice model.

To conclude, our booking data can be used to study cancellation behavior for high quality rail 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).

17. Key Words	18. Distribution Statement
Acela Express, Amtrak, booking data, cancellation behavior	No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	23	

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

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 predominantly attracts business travelers. Two service classes are offered: 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 attempts toward better understanding of passengers' behavior in the RM environment have been practiced in several aspects. Choice based RM is an active research area that offers means for analysts to incorporate a choice behavior framework into RM systems. This framework is based on the assumption that passengers make choice among the set of products offered by taking product attributes into consideration. 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 fares to be changed on a daily basis. The proposed fare strategy is based on fare price, departure day of week and destination specific effects. The passenger choice of booking day is estimated using multinomial logit (MNL) and assuming that passengers choose the day to book the ticket that maximizes their expected utility. The aggregate market demand for each destination market is modeled with ordinary least squares (OLS) regression. The passenger choice model and the aggregate demand function are incorporated in 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

This paper is organized as follows, in Section 2 we review 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 sets, 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 real ones. Conclusions and suggestions for future research are given in Section 6.

2. LITERATURE REVIEW

Revenue Management problem in railway involves several dimensions from capacity control to fare pricing, depending on the objectives of the operator. In our study, we assume that the primary objective of Amtrak in managing Acela Express service is to maximize the ticket sale revenue. Choice based RM exhibits more realistic assumptions than independent demand by accounting for customer consideration of product attributes when modeling the choice process. This method is appealing because it enables analysts to exploit preferences for attributes such as departure day of week, fare price, and destination specific effects in ticket booking behavior. 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, particularly for the railway industry.

Talluri and van Ryzin (2004) propose MNL model of buy up and buy down in a RM context based on behavioral choice model. Their problem is based on single leg RM problem where customer choice is modeled to specify the probability of purchasing each fare product as a function of the set of available fare products. In their case study, the purchase transaction data is used in their choice model estimation in which the choice set is constituted by sets of fare offered including the no-purchase choice. The attributes incorporated in their choice model are price, indicator variables for product restrictions and other variables not identified in their paper. The feasibility and benefit of the choice based RM has been examined in Vulcano et *al.* (2008). A MNL model is used to incorporate the customer choice of buy up, buy down and diversion for the airline market between New York City and Florida. Three data sources used in estimating their customer choice behavior are flight schedule data, revenue accounting data, and availability data. The choice set is constituted by all flights available on a given day between a specific pair of airports. The schedule and availability data file is used to constitute the choice set where the true fare paid from the revenue accounting file is replaced with the average revenue reported in the availability file in order to ensure the consistency between alternatives being purchased and not being purchased. The attributes specified in their choice model are arrival time of flight, fare price, and the indicator of flying on a particular day. The arrival time of flight was defined by the convex weight of four overlapping time slots: morning (between 5 AM and 11 AM), noon (between 9 AM and 3 PM), afternoon (between 1 PM and 7 PM), and evening (between 5 PM and midnight) to represent arrival times that fall in multiple time slots. The fare price attribute used in their paper is the base fare divided by 1,000 due to numerical scaling issues. The indicator of flying on a particular day is included as a dummy to account for availability of the choice on a particular departure day. The results obtained show revenue improvement ranging from 1.4 percent to 5.3 percent. The method proposed by Vulcano et *al.* (2008) has proved to be both practically feasible and economically beneficial over the current airline RM practice. On the other hand, it is also reasonable to assume that passenger can make their decision on other choice dimensions opposed to the traditional fare product on a given departure time. For instance, Van Ryzin (2005) indicates that time to demand for the products is potentially a strategic aspect that customers could consider when making ticket reservation.

Comprehensive demand models are essential for the evaluation of policy measures such as dynamic pricing strategy and capacity utilization. In order to fulfill such need, methods based on dynamic interaction with the supply control should be developed. Whelan and Johnson (2003), and Whelan *et al.* (2005) estimate 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 (RP) data collected on behalf of 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 in a credible manner.

Sibdari et *al*. (2007) study dynamic pricing policy for Amtrak Auto Train. A revenue management model is 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 being updated on a daily basis. The choice model involves a multi stage decision process similar to the ones 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 also reveals that given a sale horizon of 330 days, there is almost no reservation activity until 30 days before departure. Passenger demand is specified to follow a Poisson random variable with specified mean corresponding to the function of the remaining time, and the associated price.

3. PASSENGER CHOICE MODEL

3.1 Data Description

Booking data for Acela Express were obtained from Marketing & Product Development department at National Railroad Passenger Corporation (Amtrak). The data consist of Acela Express booking data related to trips departing in March and April 2009. In our study, we focus on business class passenger traveling from Washington D.C. to other stations in the north-eastern corridor, and only to those reservations who did not cancel their trips and that eventually contributed to the actual revenue. The database contains information in terms of trip origin, trip destination, fare class, reservation date, departure date, departure time, arrival time, fare price and accommodation charge. In order to estimate the passenger choice model and the demand function we use data recorded in March 2009, that consist of 51,002 reservation records. The actual demand recorded for April 2009 is used for the purpose of revenue testing under new fare strategies. The data set relative to March 2009 indicates that business class passenger predominantly book the ticket no earlier than 30 days before departure.

3.2 MNL Choice Design

Choice behavior is modeled by assuming that passengers maximize their utilities when choosing the day to book an Acela Express ticket. The sale horizon is assumed to be 31 days long, since this is the time interval when the majority of the reservations are registered to occur; the choice set is then constituted by 31 booking days. We 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, the passenger is assumed to maximize its utility when booking the ticket and 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,

$$v_{in} = \beta^T x_{in}, \tag{2}$$

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 is aimed to model the response of passenger behavior in response to fare policy by shifting their booking day without leaving the market. This model is essential for the revenue management process in fare optimization which will be discussed in the coming Sections. The choice model is a monopolistic model that does not account for service competition due to its distinct nature.

3.3 Passenger Choice Model Specification

We use sales data relative to business class passengers travelling in March 2009 to estimate the choice model specified in Section 3.2. There are three main difficulties in estimating the MNL model of booking day choice from sales data. First, the sales data do not contain passengers' socioeconomic characteristics; therefore, individual specific factors influencing booking decisions cannot be observed. For instance, trip purpose and personal income have potentially a high impact on booking behavior. Second, the only two choice specific attributes that can be used to specify the choice model are the advanced purchase with respect to the departure day and the ticket price. These first two difficulties are alleviated in our problem by incorporating destination specific effects, and by defining estimated parameters differently for each booking time period. Third, sales data only contain fare price information on the day the passengers make a reservation; the fare price that will be offered to the passenger on other days within the sale horizon is not known to the analyst. Thus, the associated fare price for each booking day and for each destination is computed by averaging fares offered throughout the entire month of March 2009.

The independent variables that enter the final models are advance booking (number of day), fare ($), destination specific and long distance dummy. We believe that passengers consider the tradeoff between the time to book the ticket and the associated fare price. These two variables enter the model as exogenous factors. The specific attributes for high demand markets are taken into account as dummy variables. Accounting for destination specific effects has been motivated by Iliescu et *al*. (2008) who show 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 passenger's 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 affect choice behavior: Boston South Station (BOS), New York Penn Station (NYP), and Philadelphia 30th street station (PHL), which are included in the model as dummy variables. The long distance factor is also taken into account and specified as a dummy variable. The effect of long distance is motivated by Whelan et *al*. (2008). In their study, it was 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.

To account for differences in marginal effects with respect to booking time, the parameters for each explanatory variable are specified differently for each booking period. This approach is similar to the day from issue (DFI) estimation proposed by Iliescu et *al.* (2008)’s in their DTPO model. In our early estimation trials, we have tried to specify 31 different parameters, one for each booking day. However, the model complexity is not supported by the data, many parameters are not significant and the interpretation of the estimation results is difficult. We have then decided to aggregate into periods the booking days that have approximately the same number of reservations, which results in six booking periods. The booking period group is shown later in Table 1.

The resulting utility of passenger *i* booking the ticket on day *n* can be expressed as:

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

where the independent variables and their associated index are:

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

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

$advbking$ = Number of day booking in advance of departure

$fare$ = Fare price in US dollars ($)

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

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

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

3.4 Choice Model Results and Interpretation

The results obtained from the choice model calibration are reported in Table 1. The model shows a good level of fit; the rho-squared with respect to zero is 0.304. The majority of the estimates has the expected sign and is statistically significant. However, destination specific parameters appear to be not significant in some booking periods.

Fare price estimates show high statistical significance and have the expected sign. The monotonically increasing value of price estimates from booking period 2 to booking period 6 is intuitive and indicates that passengers become less price sensitive as time approach departure. Specifically, in booking period 6, the relatively small magnitude of the positive price indicates that passenger become insensitive to price on the day of departure. This result is reasonable for the Acela Express service, given that the majority of the passengers travels for business purpose and are not very sensitive to fare price. The smaller magnitude of 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.

Boston destination dummy variable shows an 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. This is intuitive given that this market can be considered as a long distance market; the transportation mode with comparable travel time is air. Thus booking times are comparable to those observed for airlines.

New York destination shows an opposite trend to that of Boston, the increasing value of the estimates with respect to booking time imply 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 period is booking period 2 and 1 respectively.

Long distance show statistically significant estimates except for booking period 3. The long distance estimate show intuitive pattern; the earlier booking periods being more preferable to travelers and booking period 2 being the most preferred. This could be explained by the fact that driving to reach these long distance destinations is particularly onerous and that travelling by bus is relatively long 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 if everything is equal, it is generally more preferable to book and to pay the ticket as late as possible. The advance booking coefficient is statistically significant and have a negative sign indicating a strong preference toward late booking.

Table 1 Passenger advance booking choice model result

I. Alternative specific estimates	**Fare price**	T-stat	**BOS**	T-stat	**NYP**	T-stat	**PHL**	T-stat	**LONG**	T-stat
Booking period1 (day 1-11)	-0.0029*	-2.0	0.3082	0.9	-0.3010*	-2.6	0.0303	0.1	-0.5191*	-4.7
Booking period2 (day 12-20)	-0.0223*	-21.0	0.3501	1.1	-0.4095*	-3.1	0.8036*	2.5	0.9610*	3.5
Booking period3 (day 21-25)	-0.0157*	-19.6	-0.4483	-1.4	-0.4366*	-3.3	-0.11	-0.4	-0.1939	-0.8
Booking period4 (day26-29)	-0.0132*	-23.1	-0.6941*	-2.2	-0.2819*	-2.2	-0.422	-1.4	-0.6633*	-2.8
Booking period5 (day30)	-0.0056*	-9.9	-0.9606*	-2.8	0.07	0.5	-0.7294*	-2.4	-1.6420*	-6.9
Booking period6 (day31)	0.0018*	3.9	-0.9760*	-3.1	-0.0761	-0.6	-0.9678*	-3.2	-2.4790*	-10.7
II. Generic estimate	**Advbk**	T-stat					Rho-square wrt. Zero			0.3040
	-0.2269*	-65.5					Loglikelihood			-107,192
							Number of observations			44,847

*Statistically significant at 0.05 level

4. PASSENGER DEMAND FUNCTION

In our paper, we estimate the demand function to project the aggregate number of reservations throughout the sale horizon with respect to product attributes such as fare price and departure day of week. The model estimates the demand for each destination with respect to fare price and to other non price attributes on each booking day by ordinary least square (OLS) regression model. Initially, we estimated one model for each destination by specifying advance booking as a dummy variable. However, with this approach, the relatively high booking demand close to departure coping with the associated higher fare 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 we estimate 5 independent demand functions for each destination market. In 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 reality as Amtrak price changes happen in piecewise manner. In other words, fare prices don't change continuously and instantaneously in response to any small fluctuation in demand. However, the demand, on the other hand, responses almost continuously and instantaneously to changes in the fare price. This approach allows us to compare the demand of the same booking period with different associated fare price throughout one month departure horizon and to obtain a service demand that is sensitive to price within the booking period. The demand function for each booking period (k*) takes the form:

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

where: $advbking$ = Advanced purchase (in number days before departure) for booking day n

$MON, \ldots, SAT$ = Departure day of week dummies

$BookDay\_i$ = Booking day specific intercepts

We observe from our data that demand has a non-linear relationship with advanced purchase; a quadratic formulation is adopted here to account for this non-linear effect. Due to unrealistic results of some stations and data confidentiality concern, we selected 8 out of 15 destinations for the analysis where destination stations are renamed into station1 to station8. Due to space limit, we only show demand function results of 3 stations from 8 stations as examples in our analysis. The demand function estimates are shown in Table 2 to Table 4. The fare price estimates exhibit 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 difficulties in applying the proposed regression procedure. The station which shows the best model fit is station 5, which could 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. This could be caused by the fact that for this booking period the fare strategy does not have significant difference from observation to observation, thus the model fit is comparatively good. Some of the demand function estimates provide unrealistic signs, for instance positive fare estimate throughout all the booking period. Stations which show unrealistic results are not considered in our optimization process (discussed in Section 5) due to the inability to incorporate positive fare coefficients into the fare optimization process.

Table 2 Station 1 Demand function result

Booking period 1(Day1-11)			Booking period 2(Day12-20)			Booking period 3(Day21-25)			Booking period 4 (Day26-29)			Booking period 5 (Day30-31)		
	Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat
(Constant)	1.820	1.600	(Constant)	16.531	2.104	(Constant)	1.650	1.664	(Constant)	5.492	2.579	(Constant)	5.865	0.681
AdvbkSq	0.002	1.739							AdvbkSq	-0.040	-1.213			
Fare	-0.007	-1.375	Fare	-0.057	-1.497	Fare	0.002	0.371	Fare	-0.010	-1.061	Fare	0.004	0.118
Mon Dep	-0.065	-0.136	Mon Dep	-6.354	-1.380	Mon Dep	-0.315	-0.779	Mon Dep	-0.093	-0.106	Mon Dep	5.990	1.280
Tues Dep	0.545	1.082	Tues Dep	-7.519	-1.488	Tues Dep	-0.546	-1.237	Tues Dep	-0.200	-0.200	Tues Dep	0.228	0.052
Wed Dep	1.340	2.328	Wed Dep	-5.486	-1.138	Wed Dep	-0.531	-1.171	Wed Dep	0.125	0.123	Wed Dep	-1.697	-0.366
Thurs Dep	0.087	0.162	Thurs Dep	-5.606	-1.185	Thurs Dep	-0.163	-0.330	Thurs Dep	-0.874	-0.901	Thurs Dep	-2.982	-0.606
Fri Dep	0.727	1.761	Fri Dep	-4.764	-1.095	Fri Dep	0.389	0.861	Fri Dep	-0.050	-0.058	Fri Dep	0.458	0.104
Sat Dep	-0.148	-0.219	Sat Dep	-6.399	-1.297	Sat Dep	-0.332	-0.514	Sat Dep	-1.184	-0.880	Sat Dep	-1.455	-0.220
Book Day1	-0.054	-0.069	Book Day12	-2.188	-0.393	Book Day21	-0.068	-0.168	Book Day27	-0.271	-0.421	Book Day30	-2.932	-1.233
Book Day2	-1.713	-2.254	Book Day13	-2.330	-0.383	Book Day22	-0.320	-0.765	Book Day28	0.263	0.373			
Book Day3	-0.224	-0.353	Book Day14	-1.949	-0.319	Book Day23	0.614	1.632						
Book Day4	-1.452	-2.255	Book Day15	-0.981	-0.147	Book Day24	0.298	0.785						
Book Day5	-0.966	-1.763	Book Day16	-2.683	-0.506									
Book Day6	-0.041	-0.063	Book Day17	-1.335	-0.270									
Book Day7	-0.592	-0.963	Book Day18	-1.344	-0.236									
Book Day8	-0.401	-0.750	Book Day19	7.085	1.371									
Book Day9	-0.488	-0.824												
Book Day10	-0.284	-0.542												
R Square		0.265	R Square		0.102	R Square		0.146	R Square		0.078	R Square		0.149

Table 3 Station 5 Demand function result

Booking period 1(Day1-11)			Booking period 2(Day12-20)			Booking period 3(Day21-25)			Booking period 4 (Day26-29)			Booking period 5 (Day30-31)		
	Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat
(Constant)	28.999	8.464	(Constant)	36.398	4.829	(Constant)	106.553	3.598	(Constant)	304.186	2.097	(Constant)	622.579	2.230
AdvbkSq	-0.016	-4.459				AdvbkSq	-0.757	-8.406	AdvbkSq	-3.492	-3.700			
Fare	-0.059	-3.169	Fare	-0.064	-1.183	Fare	-0.153	-0.712	Fare	-1.153	-1.238	Fare	-2.659	-1.513
Mon Dep	-3.791	-2.680	Mon Dep	-7.461	-2.899	Mon Dep	6.741	1.017	Mon Dep	54.432	2.480	Mon Dep	154.914	3.893
Tues Dep	-2.660	-1.786	Tues Dep	-3.150	-1.117	Tues Dep	20.845	2.837	Tues Dep	29.003	1.131	Tues Dep	292.668	6.378
Wed Dep	0.453	0.299	Wed Dep	-6.486	-2.171	Wed Dep	34.130	3.751	Wed Dep	53.104	2.008	Wed Dep	306.067	5.554
Thurs Dep	-0.792	-0.532	Thurs Dep	-7.384	-2.419	Thurs Dep	40.334	4.477	Thurs Dep	77.027	2.664	Thurs Dep	285.056	4.989
Fri Dep	1.117	0.743	Fri Dep	1.253	0.406	Fri Dep	27.562	2.912	Fri Dep	71.676	2.404	Fri Dep	233.648	4.123
Sat Dep	-6.725	-4.299	Sat Dep	-17.913	-6.509	Sat Dep	-18.054	-2.653	Sat Dep	-61.375	-2.569	Sat Dep	-121.852	-3.060
Book Day1	-0.104	-0.052	Book Day12	-7.331	-2.205	Book Day22	-8.010	-1.550	Book Day27	-10.791	-0.727	Book Day30	-88.365	-4.159
Book Day2	-1.710	-0.900	Book Day13	-6.897	-2.127	Book Day23	-10.422	-2.084	Book Day28	-16.384	-1.029			
Book Day3	-1.626	-0.917	Book Day14	-7.171	-2.230	Book Day24	-1.125	-0.211						
Book Day4	1.051	0.607	Book Day15	-5.125	-1.580									
Book Day5	-2.010	-1.175	Book Day16	-2.226	-0.696									
Book Day6	-2.206	-1.315	Book Day17	6.965	2.112									
Book Day7	-2.470	-1.441	Book Day18	2.161	0.691									
Book Day8	-1.169	-0.660	Book Day19	-1.994	-0.642									
Book Day9	-0.253	-0.142												
Book Day10	1.002	0.546												
R Square		0.247	R Square		0.265	R Square		0.567	R Square		0.293	R Square		0.766

Table 4 Station 8 Demand function result

Booking period 1(Day1-11)			Booking period 2(Day12-20)			Booking period 3(Day21-25)			Booking period 4 (Day26-29)			Booking period 5 (Day30-31)		
	Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat		Beta	T-stat
(Constant)	8.032	3.969	(Constant)	1.218	0.427	(Constant)	9.006	1.368	(Constant)	25.821	0.876	(Constant)	-27.493	-0.438
AdvbkSq	-0.003	-1.376							AdvbkSq	-0.832	-2.892			
Fare	-0.020	-1.570	Fare	0.030	1.176	Fare	0.002	0.030	Fare	-0.026	-0.103	Fare	0.811	1.420
Mon Dep	-0.143	-0.148	Mon Dep	1.365	1.062	Mon Dep	9.400	3.423	Mon Dep	26.292	3.428	Mon Dep	58.549	4.161
Tues Dep	1.561	1.629	Tues Dep	5.933	4.308	Tues Dep	15.760	5.249	Tues Dep	21.164	2.539	Tues Dep	127.369	8.371
Wed Dep	1.697	1.654	Wed Dep	2.235	1.456	Wed Dep	17.200	5.412	Wed Dep	27.782	3.175	Wed Dep	126.381	7.006
Thurs Dep	2.161	2.142	Thurs Dep	1.431	0.903	Thurs Dep	22.381	6.885	Thurs Dep	31.801	3.354	Thurs Dep	128.991	6.627
Fri Dep	1.929	1.911	Fri Dep	2.597	1.733	Fri Dep	10.876	3.265	Fri Dep	26.899	2.778	Fri Dep	79.858	4.012
Sat Dep	-0.762	-0.619	Sat Dep	-0.736	-0.459	Sat Dep	-2.315	-0.751	Sat Dep	-6.522	-0.818	Sat Dep	-5.564	-0.372
Book Day1	-1.241	-0.896	Book Day12	-0.769	-0.489	Book Day21	-10.180	-4.077	Book Day27	-2.405	-0.472	Book Day30	-69.534	-8.922
Book Day2	-1.471	-1.116	Book Day13	-1.345	-0.879	Book Day22	-9.089	-3.804	Book Day28	-1.836	-0.341			
Book Day3	-1.067	-0.830	Book Day14	0.367	0.229	Book Day23	-2.895	-1.181						
Book Day4	-0.352	-0.316	Book Day15	-1.546	-0.921	Book Day24	0.352	0.145						
Book Day5	-1.769	-1.577	Book Day16	1.343	0.869									
Book Day6	0.270	0.237	Book Day17	1.252	0.821									
Book Day7	-2.035	-1.810	Book Day18	-0.567	-0.381									
Book Day8	-2.331	-2.113	Book Day19	0.160	0.105									
Book Day9	-2.157	-1.946												
Book Day10	0.136	0.120												
R Square		0.117	R Square		0.177	R Square		0.524	R Square		0.317	R Square		0.873

5. FARE OPTIMIZATION

5.1 Optimization Procedure

The passenger choice model and the demand function estimated are used to solve a fare optimization problem. The proposed fare setting scheme is updated on a daily basis and is based on the assumption that Amtrak aims to maximize the revenues from Acela Express ticket sales. The day specific fare structure proposed within the sale horizon is strategically based on the booking choice model, while the fare pattern is obtained by maximizing the expected revenue.

Our procedure aims to optimize the fare price over a representative week in March (March 16th to March 22nd) and April (April 20th to April 26th). The booking choice model and the passenger demand function estimated from sales data in March 2009 are used to formulate a fare optimization problem by representing passenger's response to fare strategy. 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 16th to March 22nd. In reality, the earliest departure date to which this fare strategy could be imposed is 31 days after the prices are computed. Therefore, we test the revenue in April by imposing these fare strategies to the week April 20th to April 26th and we assess the performance of the fare price estimated in March. This also allows us to check whether the fare prices estimated from sales data in a specific month performs the next month.

5.2 Problem Formulation

Revenue optimization is formulated for each departure day and for each station as an expected revenue maximization problem:

$$\max \text{ Revenue} = [\sum_{n=1}^{31} Demand_n(fare_n) + y \times \{Capacity - \sum_{n=1}^{31} Demand_n(fare_n)\}]$$

$$\times [\sum_{n=1}^{31} \{fare_n \times \Pr(\text{day } n)\}] \qquad (6)$$

The first term represents the total predicted passenger demand on a particular departure day and is calculated by summing the booking demand over the entire sale horizon. This predicted demand is controlled by the decision variable *y* which ensures that only demand within capacity limit contributes to the revenues. When demand exceeds capacity (y = 1), the spillover demand will be subtracted from the predicted demand. The train capacity for each destination market is approximated through historical sales data by assuming that actual demand in March 2009 was at 80 percent load factor. The second term in equation 6 represents the expected fare price and is expressed as the sum of product between day specific fare $(fare_n)$ and the probability that passenger book the ticket on that booking day ($\Pr(\text{day } n)$) throughout the sale horizon. Thus the overall formulation represents the expected revenues from a particular departure day; the decision variables are day specific fare $(fare_n)$.

We assume that Acela Express fare strategy is subjected to the predetermined fare bound restriction and its fare price only increase monotonically as time approaches departure. The incremental amount of fare price from one day to the next is assumed to be within the bound limit (assumed to be \$ 5.00). However, this assumption serves only for the purpose of this academic research but does not necessarily represent the actual revenue management policy currently used for Acela Express. The corresponded constraints for problem in Equation (6) according to our RM control assumptions are:

$$Capacity - Demand \leq M \times (1 - y) \tag{7}$$

$$Demand - Capacity \leq M \times y \tag{8}$$

$$fare_{lb} \leq fare_n \leq fare_{ub} \tag{9}$$

$$fare_n \leq fare_m \text{ ; for all } m \geq n \tag{10}$$

$$fare_m - fare_n \leq incremental\_allowance(\$) \text{ ; for all } m \geq n \tag{11}$$

The first two constraints force *y* to one when demand exceeds capacity and zero otherwise by incorporating the large number *M*. When demand is less than capacity, the left hand side of the first constraint is positive, thus *y* is forced to zero in order for the right hand side to equal *M* with the second constraint satisfying the condition. On the other hand, when demand is greater than capacity, the left hand side of the second constraint is positive, thus *y* is forced to one in order for the right hand side to equal *M* with the first constraint satisfying the condition. The third constraint imposes bound on fare price. These bounds are assumed to be the maximum and the minimum of the average day specific fare prices recorded for the entire sale horizon of March 2009. The fourth constraint ensures that the new fare price increase monotonically with respect to booking day and the last constraint ensures that the increment of price on each day does not exceed the incremental allowance (assumed to be \$5.00). The classifications of all the variables are:

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

y = Binary Decision variable (equal 1 when demand exceeds capacity, 0 otherwise)

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

M = Exorbitantly large number

5.3 Optimization Result

Optimization results are obtained from solving equation (6) for representative week in March (March 16th to March 22nd). The corresponding fare prices by day of week and for each destination are shown in Figure 1 to Figure 8 for station 1 to 8 respectively.

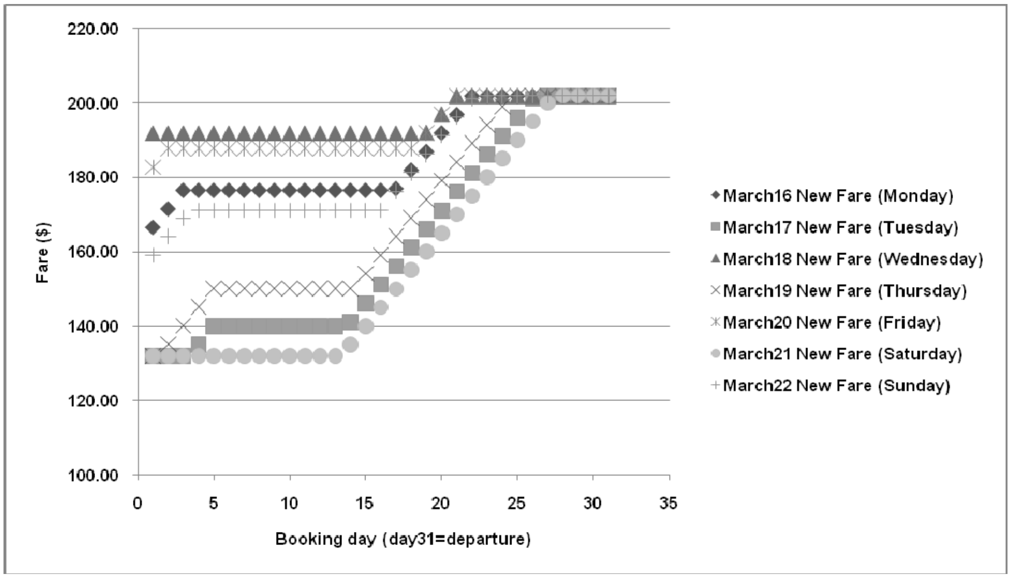

Figure 1 Station 1 Fare Strategy in March 16-22

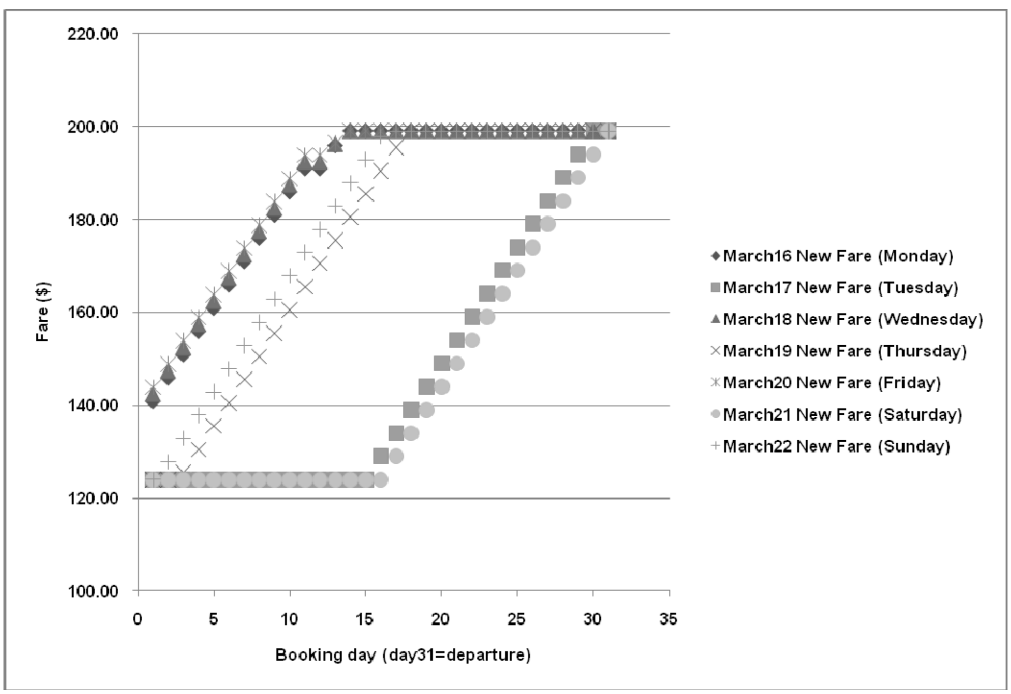

Figure 2 Station 2 Fare Strategy in March 16-22

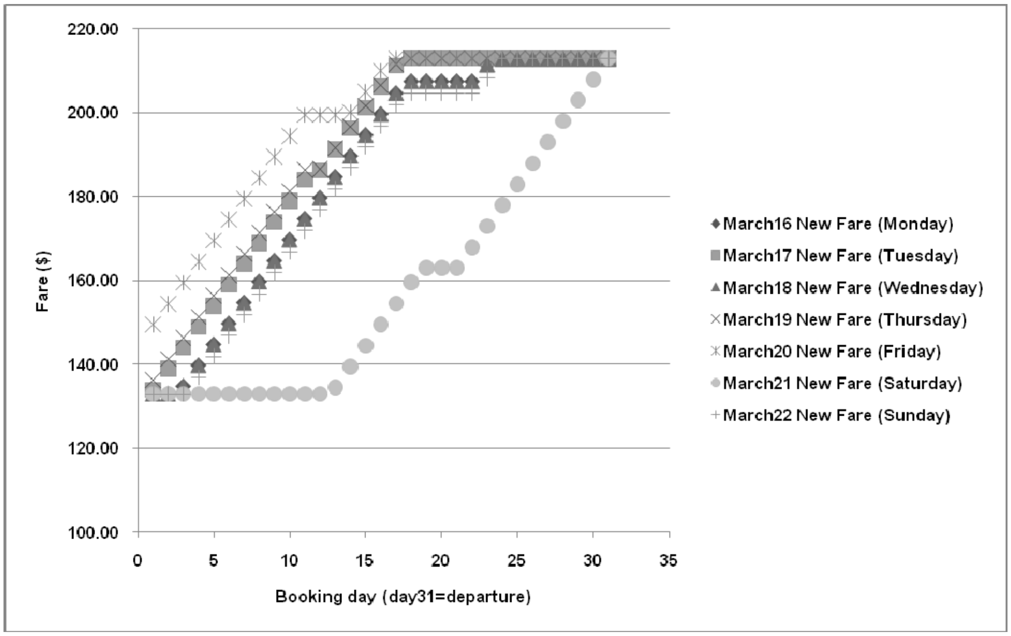

Figure 3 Station 3 Fare Strategy in March 16-22

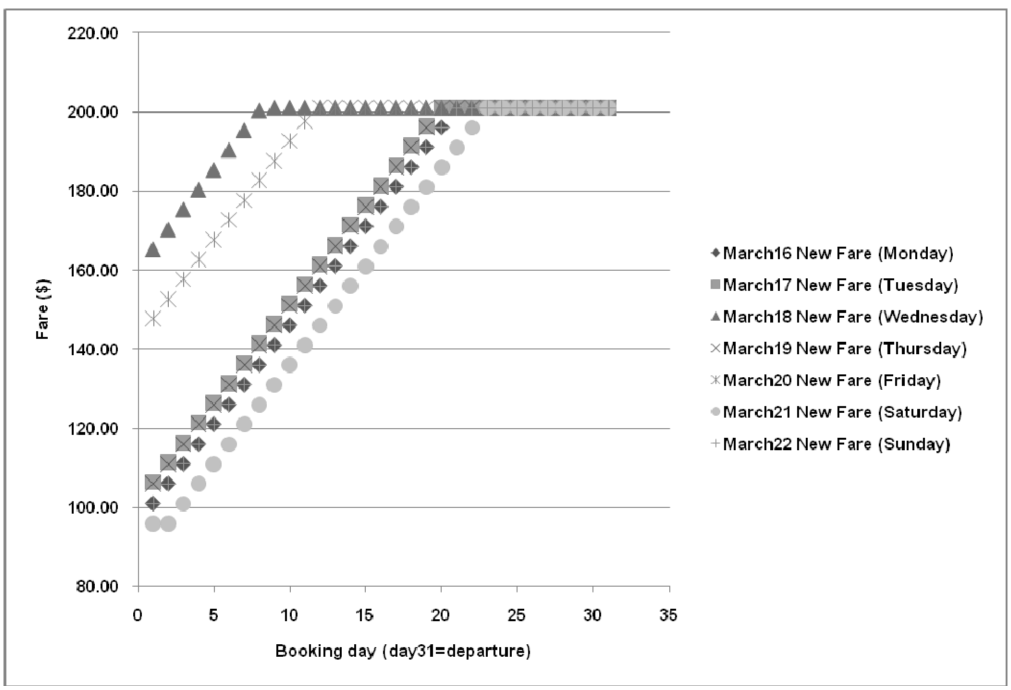

Figure 4 Station 4 Fare Strategy in March 16-22

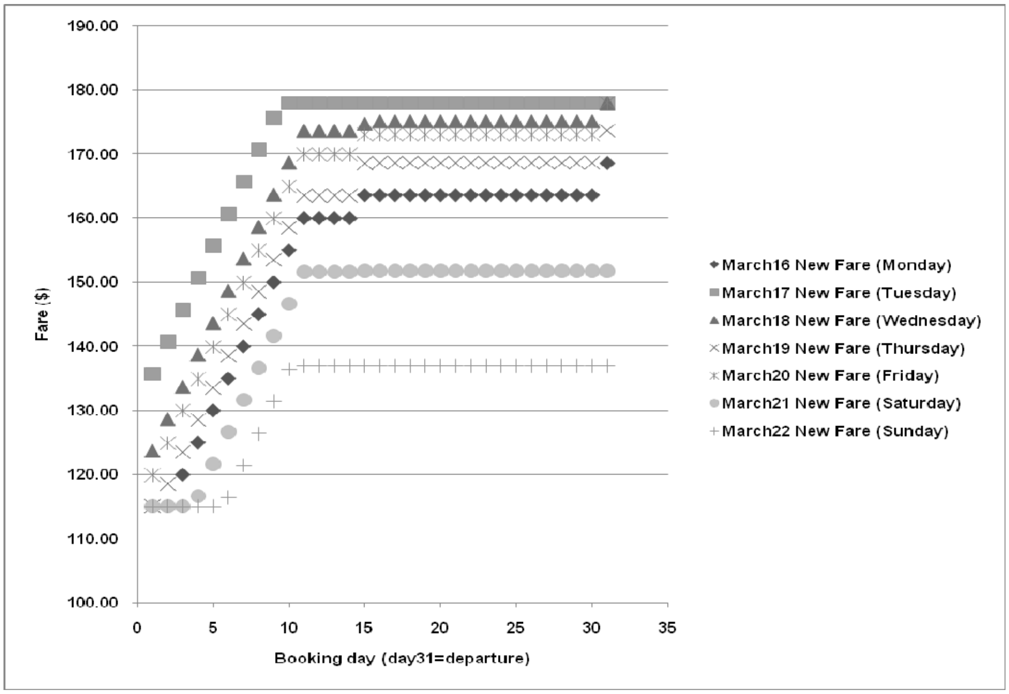

Figure 5 Station 5 Fare Strategy in March 16-22

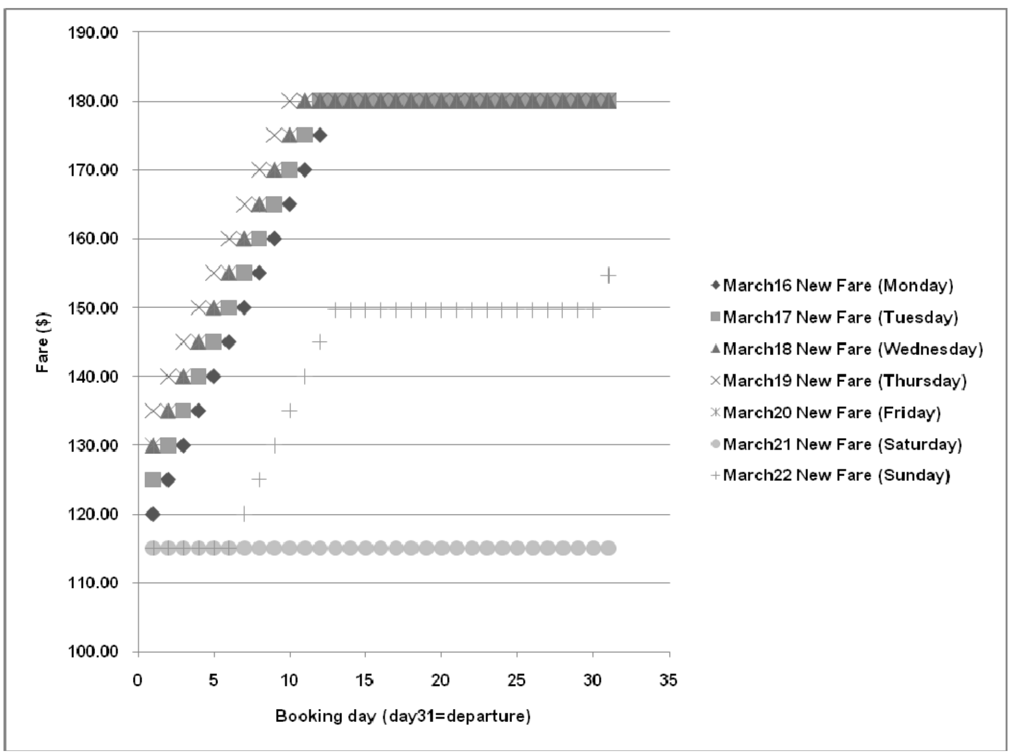

Figure 6 Station 6 Fare Strategy in March 16-22

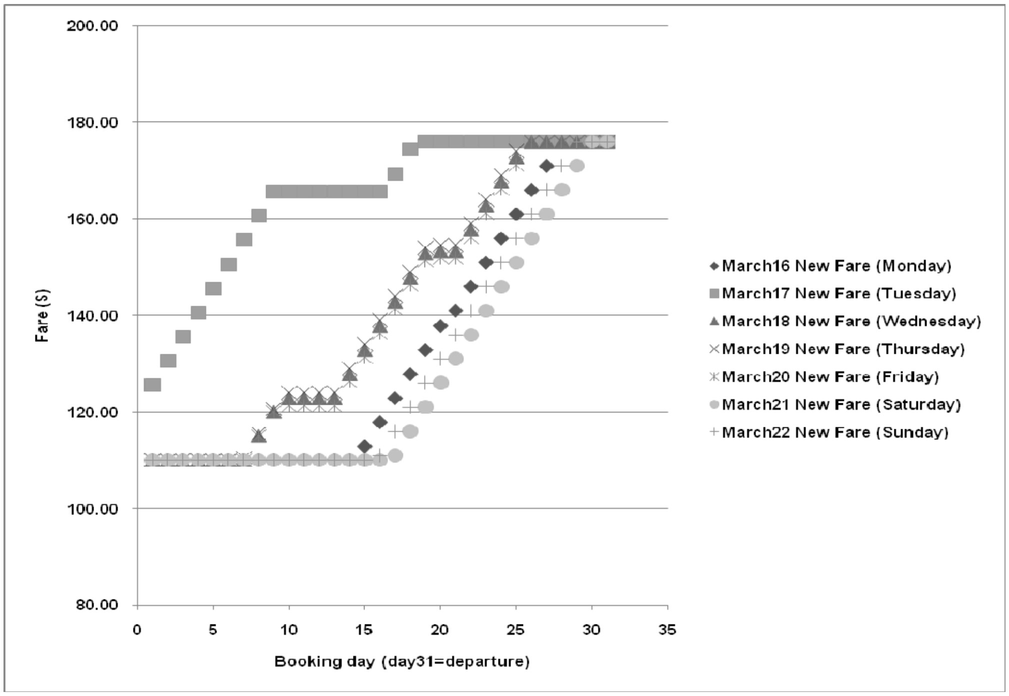

Figure 7 Station 7 Fare Strategy in March 16-22

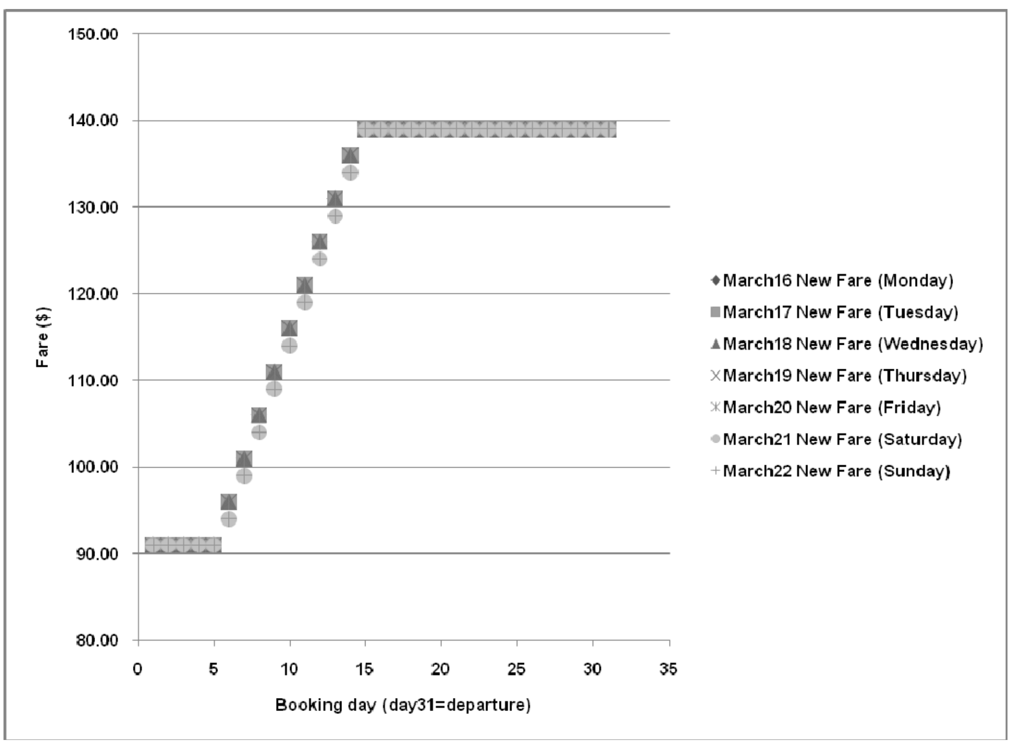

Figure 8 Station 8 Fare Strategy in March 16-22

We will discuss these results by station and 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 2. For instance, Table 2 shows that demand reaches the highest peak on Wednesday and the lowest on Saturday particularly in the first (Day 1-11), second (Day 12-20), and fourth (Day 26-29) booking period. The fare strategy in Figure 1 suggests that the fare should be charged at higher price on the day with high demand (Wednesday) and lower price on the day with low demand (Saturday) throughout the sale horizon. The changes of day of week effect from booking period to booking period in the demand function also influence the price strategy, for example when we compare between Monday and Wednesday of station 1, in booking period 2 (Day 12-20) Wednesday is shown to have greater demand compared to Monday, thus its corresponded fare price in Figure 1 in the beginning of the booking period 2 (Day 12-20) is lower than Wednesday. However, in the booking period 3 (Day 21-25), Monday shows higher passenger than Wednesday, thus the optimization takes this into account and the fare price for Monday start increasing before the end of booking period 2 (Day 17) and finally match Wednesday fare at the beginning of the booking period 3 (Day21). Other stations' fare strategies exhibits a similar trend to the one occurred for station1.

For station 2, departure days with high demand obtained from the associated demand function are Monday, Wednesday, and Friday. According to our result, 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 shown to be 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 the highest fare on Monday, Wednesday, and Friday with the lowest fare on Saturday.

For station 3, departure day with high demand obtained from the associated demand function is Friday with high demand in booking period 3 to 5 (Day 21-31). According to our result, the lowest demand is shown to be 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 the highest fare on Friday with the lowest fare on Saturday.

For station 4, departure day with high demand obtained from the associated demand function is Wednesday with high demand in booking period 2 (Day 12-20) and 4 (Day 26-29). According to our result, the lowest demand is shown to be Saturday in booking period 1 (Day 1-11), 3 (Day 21-25), 4 (Day 26-29), and 5 (Day 30-31). Thus, the fare strategy by day of week in Figure 4 suggests the highest fare on Friday with the lowest fare on Saturday.

For station 5, departure day with high demand obtained from the associated demand function in Table 3 is Tuesday with high demand in booking period 2 (Day 12-20) and 5 (Day 30-31). The lowest demand is shown to be Saturday in all booking periods. Thus, the fare strategy by day of week in Figure 5 suggests the highest fare on Tuesday with the lowest fare on Saturday.

For station 6, departure day with high demand obtained from the associated demand function is Thursday with high demand in booking period 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 shown to be Saturday from booking period 2 to 5 (Day 12-31). Thus, the fare strategy by day of week in Figure 6 suggests the highest fare on Thursday with 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) according to Table 4. The lowest demand is shown to be Saturday in all booking periods. However, due to its relatively low fare price of this station, the allowable fare gap of this destination imposed in the optimization problem is smaller compared to its nearby station, therefore the day of week effect is not sufficient to influence the fare strategy of this station in Figure8 to significantly vary by day of week.

These findings conclude 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 where the fare price coefficient accounts for passenger price sensitivity. 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 because the decrease in passenger demand due to higher fare in this case is outweighed by the higher demand based on day of week effect, thus charging higher fare price on this day compared to other days within the week contribute to higher daily revenue.

The revenue improvements for each departure day in March (March 16th to March 22nd) and April (April 20th to April 26th) representative week are shown in Table 5 and 6. The total revenue improvement ranges from 2.06 to 14.64 percent and from 0.70 to 11.60 percent for March and April representative departure periods respectively. From Table 5 and 6, it is indicated that Monday, Tuesday, and Saturday are the days which revenues could be significantly improved by applying our proposed pricing scheme.

Table 5 Revenue Improvement in March representative week

Station	Mach16 Monday	Mach17 Tuesday	Mach18 Wednesday	Mach19 Thursday	Mach20 Friday	Mach21 Saturday	Mach22 Sunday
	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)%)
1. Station1	19.1	20.29	17.84	10.26	-1.01	15.08	9.23
2. Station2	30.63	16.05	24.08	13.03	27.58	71.28	-0.63
3. Station3	51.63	16.74	20.23	9.8	6.71	23.57	4.17
4. Station4	18.03	2.8	8.71	2.83	2.62	15.35	13.98
5. Station5	10.36	11.46	5.52	1.32	5.5	13.01	-0.71
6. Station6	20.68	13.85	4.99	2.31	3.7	-8.77	0.98
7. Station7	16.35	17.23	-0.5	1.9	-2.23	18.98	12.23
8. Station8	20.41	11.77	6.81	2.95	5.41	45.44	25.5
Revenue improvement per day	14.64	12.28	5.64	2.06	4.63	16.18	2.95

Table 6 Revenue Improvement in April representative week

Station	April 20 Monday	April 21 Tuesday	April 22 Wednesday	April 23 Thursday	April 24 Friday	April 25 Saturday	April 26 Sunday
	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)	Improved (%)
1. Station1	28.83	11.20	41.09	5.36	5.79	10.00	17.41
2. Station2	23.61	13.11	18.81	20.66	-5.46	-22.78	-3.57
3. Station3	39.70	27.78	9.07	17.17	4.31	9.56	10.42
4. Station4	20.98	9.64	19.54	-5.03	-0.35	4.31	2.56
5. Station5	5.40	8.59	4.49	1.25	3.76	12.36	-4.03
6. Station6	14.75	9.51	2.20	-0.38	1.37	-15.33	4.35
7. Station7	10.93	9.40	-2.47	-3.16	-3.97	23.08	22.38
8. Station8	11.41	13.53	9.34	8.02	3.50	27.46	33.21
Revenue improvement per day	8.98	9.95	5.42	2.28	2.94	11.60	0.70

6. CONCLUSIONS

In this paper, we have illustrated a fare pricing strategy for the Acela Express service operated by Amtrak. The RM method proposed is based on passenger's preference and products' attributes. Using sales data, a MNL model has been calibrated; the 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 in 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 booking data. The accuracy of the estimates depends on the market size; for instance, the model produces good results for station5 market which is the predominant market for Acela Express. Overall, we show that the proposed model in this paper is promising and can potentially lead to increase in revenue. It was demonstrated that the pricing strategy which accounts for choice behavior could potentially increase the revenue from 2.06 to 14.64 percent and 0.70 to 11.60 percent per day within the respective weeks of March and April. However, it should be noted that, as with any academic work, the model is based on some simplifying assumptions which might not fully comply with the real world problem. For example, Amtrak pricing strategy is more complicated than what presented in this paper. We did not account for cancellation behavior, various discounts, guest reward program, special fare plans or competition with non Acela trains or other modes of transportation. Also, the choice model is not tested independently to show if it accurately reflects customers' choice behavior in the market. So there is significant room to improve or extend this research.

Several research extensions 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 sales 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 model calibrated handles deterministic heterogeneity only. Mixed logit models could be adopted to address random heterogeneity in customer behavior. Both latent classes and random parameters logit models have the potential to improve the accuracy of the customer choice model.

To conclude, our booking data can be used to study cancellation behavior for high quality rail 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).

7. 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.

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.

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

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*.

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

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.