ABSTRACT

Title of Document:	DISCRETE CHOICE MODELS FOR REVENUE MANAGEMENT
	Pratt Hetrakul, Doctor of Philosophy, 2012
Directed by:	Professor Cinzia Cirillo, Department of Civil & Environmental Engineering

In the transportation field, the shift of airline and railway industries toward web-based distribution channels has provided passengers with better access to fare information. This has resulted in passengers becoming more strategic to price. Therefore, a better understanding of passenger choice behavior is required in order to support fare strategies. Methods based on discrete choice (DC) analysis have recently been introduced in revenue management (RM). However, applications of DC models in railway ticket pricing are limited and heterogeneity in choice behavior across different categories of travelers has mostly been ignored. Differences in individual taste are crucial for the RM sector. Additionally, strategic passenger behavior is significant, especially in markets with flexible refund and exchange policy, where ticket cancellation and exchange behavior has been recognized as having major impacts on revenues.

This dissertation examines innovative approaches in discrete choice modeling to support RM systems for intercity passenger railway. The analysis, based on ticket reservation data, contributes to the existing literature in three main aspects. Firstly, this dissertation develops choice models of ticket purchase timing which account for heterogeneity across different categories of passengers. The methodology based on latent class (LC) and mixed logit (ML) model framework offers an alternative approach to demand segmentation without using trip purposes which are not available in the data set used for the analysis.

Secondly, this dissertation develops RM optimization models which use parameters estimated from the choice models and demand functions as key inputs to represent passenger response to RM policy. The approach distinguishes between leisure and business travelers, depending on departure time and day of week. The formulated optimization problem maximizes ticket revenue by simultaneously solving for ticket pricing and seat allocation. Strategies are subjected to capacity constraints determined on the basis of the railway network characteristics.

Finally, this dissertation develops ticket cancellation and exchange model using dynamic discrete choice model (DDCM) framework. The estimated model predicts the timing of ticket cancellations and exchanges in response to trip schedule uncertainty, fare, and refund/exchange policy of the railway service. The model is able to predict new departure times of the exchanged tickets and covers the full range of departure time alternatives offered by the railway company.

DISCRETE CHOICE MODELS FOR REVENUE MANAGEMENT

By

Pratt Hetrakul

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Cinzia Cirillo, Chair/Advisor
Professor Ali Haghani
Professor John Philip Rust
Professor Lei Zhang
Professor Paul M. Schonfeld

Acknowledgements

Firts of all, I would like to express my deepest gratitude to my dissertation advisor, Dr. Cinzia Cirillo for her encouragement and dedication to the success of my research. She has always been an excellent teacher I have relied on throughout my graduate studies and a shining example of a successful faculty. It has been my great pleasure working alongside with her. I would like to thank my advisory committee members, Dr. Ali Haghani, Dr. John Philip Rust, Dr. Lei Zhang, and Dr. Paul M. Schonfeld for their valuable comments and suggestions on this research. I also thank Dr. John Philip Rust for his course in computational economics which forms my background for much of this work.

I would also like to thank Dr. Fabian Bastin at the University of Montreal for his guidance and suggestion toward my model formulation and for all the help he provided when I was a visiting student in Montreal. My special thanks to my colleague, Dr. Renting Xu who inspired me and provided me tremendous help in C programming.

This dissertation would not have been possible without the support of my family. My utmost respect goes to my father who gives me the passion of pursuing Ph.D. I would like to thank my mother for understanding my passion for reseach and my brother for his great support.

I would also like to thank students in the UMD Transportation Program and my Thai friends in Maryland for making my stay an enjoyable experience.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction

The field of revenue management (RM) was established in 1978 after the US airline deregulation with the main objective of revenue maximization under demand uncertainty. The deregulation has established a strong competition among US domestic air carriers which drives airline companies to invest time and money to satisfy customer demand by optimally allocating the available resources (Glover et al., 1982; Davis, 1994). Analysts have also become increasingly interested in predicting passenger demand to support RM strategy. As a consequence, in recent years, discrete choice models have been integrated into RM systems. The approach supports RM decisions by providing insight on passenger level responses to product attributes based on behavioral preferences. Researchers have also begun to investigate how discrete choice models and passenger-level data can be integrated with optimization models at a system level.

More recently, the shift of airline and railway industries toward web-based distribution channels has provided passengers better access to fare information. This phenomenon has influenced passengers to become more strategic to price. In RM, the term “strategic passengers” generally refers to passengers who anticipate future sales and choose purchase timing and products to maximize their expected consumer surplus. In some circumstances, not only do strategic passengers decide when and where to purchase tickets, but they are also influenced to exchange and cancel their tickets. This is especially true in markets with flexible refund and exchange policy, where strategic passengers are inclined to buy tickets in advance and adjust ticket schedules when their travel plans or fare rates change. Reliable predictions in ticket cancellation and exchange decisions can support analysts to improve efficiency in capacity planning and refund/exchange policies. Indeed, despite the fact that small improvements in accuracy of the demand models can contribute to millions of dollars in annual revenue for an airline (Neuling, Riedel et al. 2004), the cancellation models used in practice are still fairly simplistic. Currently, many airlines forecast cancellation rates using time-series or averaging methods which do not capture or explain how individual passengers make decisions. To the extent that different types of passengers and itineraries exhibit distinct cancellation and exchange rates, current state of practice cannot make accurate predictions when the underlying passenger or itinerary mix changes.

While numerous RM studies have focused on airline and hotel industries, its applications to railway are relatively limited. A railway is a green alternative compared to other transportation modes and plays significant roles in emission reduction policy. In the US, $8 billion ($2009) has been made available under the American Recovery and Reinvestment Act of 2009 for rebuilding high speed rail links throughout the country. The Amtrak Northeast Corridor Infrastructure Master Plan 2010 (Amtrak, 2010) called for $52 billion ($2010) in investment to cover needed system repair, upgrades, and some capacity enhancements to accommodate passenger demand in 2030 which is projected to increase 60% from 2012 (Amtrak, 2012). To this extent, it is expected that the railway passenger traffic will almost certainly increase over the next decade. In term of revenue, Amtrak saw $2 billion ($2009) in 2009 (Amtrak, 2009) and combined revenues for all passenger railway operators in the UK for 2009 were in excess of £6 billion (£2009) (Office of Rail Regulation, 2009).

As in the airline industry, the goal of railway RM systems is to find the optimal number of passengers travelling along each leg in order to maximize the overall revenue.

This can be achieved by implementing ticket pricing or by limiting the availability of certain ticket class or certain market to passengers. However, the problem is very complex due to its probabilistic and dynamic nature. The probabilistic aspect is due to the uncertainty in forecasting passenger demand while the dynamic aspect is due to inter-temporal effects of passenger behavior such as trip rescheduling or ticket cancellation decisions. To support the development of railway RM strategies, a better understanding of the railway passenger choice behavior is required. It is also important that the model be capable of supporting a broad range of policy implications by utilizing the availability of disaggregate data. The approach is expected to offer a rich behavioral interpretation of passenger behavior and explore differences based on passenger and trip characteristics.

1.1 Problem Statement

While railway pricing and RM are based on the heterogeneity of choice behavior across different categories of passengers, the use of passenger choice models to directly support this aspect of railway planning applications has been relatively limited so far. Given that RM strategy relies on the premise that different passengers are willing to pay different amounts for a product, incorporating heterogeneous choice models in the RM strategy is expected to contribute to significant revenue improvements.

Meanwhile a number of studies have proposed different approaches to model ticket cancellation and exchange behavior, their primary focus is to predict the timing of cancellations and exchanges without predicting new departure times of the exchanged tickets. The prediction of new departure times can further support RM capacity planning. More importantly, previous studies have mostly ignored inter-temporal effects of strategic passengers in ticket cancellation and exchange behavior. This behavior is crucial in markets with flexible refund and exchange policy where ticket cancellation and exchange behavior has been recognized as having major impacts on revenues.

In this dissertation, we will then focus on how a railway operator can use existing data sources to develop passenger choice models that capture the characteristics previously mentioned and relax the limitations of previous studies.

1.2 Research Objectives and Scope

The first objective of this research is to develop passenger choice models of ticket purchase timing which account for heterogeneity across different categories of passengers and capable of supporting railway RM decisions. To demonstrate choice models' application in RM strategy, the estimated models are incorporated into RM optimization models which optimize ticket revenue based on pricing and seat allocation strategy. The second objective of this research is to explore the use of dynamic discrete choice model (DDCM) for an important aspect of passenger behavior, namely cancellation and exchange behavior. The developed model is expected to capture inter-temporal effects on individual behavior in ticket cancellation and exchange decisions that are usually treated in a static context.

The principal objectives can be summarized as follows:

1. Develop passenger choice models of ticket purchase timing which account for taste heterogeneity across different categories of passengers. The models should be able to capture passenger behavioral characteristics under railway RM policy as well as differences in taste preferences across segments. The model output is then integrated into the RM optimization model system.

2. Develop RM optimization models which optimize ticket revenue by simultaneously solving for pricing and seat allocation strategy under capacity constraints determined on the basis of the railway network characteristics. The optimization model should be formulated to allow passengers to realistically respond to RM policy based on purchase timing and variation in demand volume.
3. Assess the impact of heterogeneous choice models toward RM strategy and assess their performance in the RM optimization models.
4. Develop ticket cancellation and exchange model which captures inter-temporal behavior of strategic passengers based on a dynamic discrete choice model (DDCM) framework. The model should reflect exchange, refund, and fare policy of the railway operator. Its structure must be designed to be capable of predicting not only ticket cancellation and exchange time, but also the new departure times of the exchanged tickets. The prediction of new departure times is expected to further support RM capacity planning.
5. Develop an efficient algorithm to approximate dynamic programming problem in the DDCM.

Given the primary focus of this dissertation is on developing choice models using railway operator data sources, the analysis is subjected to certain data limitations which do not allow certain aspects of the problem to be investigated, which are:

Demand substitution across markets

In this research, consideration of changes in origin-destination will be ignored because it requires the definition of a choice set that is significantly different across passengers. Actually, no information is available to construct a realistic choice set for each passenger.

However, given that railway operators typically forecast demand based on historical booking independently for each origin-destination market, this assumption should have limited impact for most applications.

Demand substitution across transportation modes

Since a railway operator cannot get full access to its competitors booking data, the analysis will primarily focus on its existing demand without direct consideration of demand substitution across transportation modes. Given that this research is based on ticket reservation data which reflect the outcomes of the travelers' decisions on transportation mode and service, this assumption is consistent with our approach. On the other hand, the model development will focus on other elements of passenger decisions such as ticket purchase timing, and ticket cancellation/exchange. Moreover, to represent the variation in passenger demand volume corresponded to changes in RM policy, passenger demand functions are proposed to represent induced and lost demand.

Demand substitution across passenger classes

This railway service offers two passenger classes; first class and coach class. Given that coach class passengers constitute 92% of the total demand; this research primarily focuses on coach class passengers. The sample size of the first class passengers is also relatively small compared to coach class passengers. The small sample size of first class passengers poses difficulties in estimating choice models and the demand functions, thus incapable for the models to be integrated in the RM revenue optimization.

Revenue management objective

The primary goal of this research is to propose RM strategy which requires minimal changes from its existing operational routine. Therefore, the proposed RM optimization problem will not consider strategy which requires change in their train operation (e.g. adjusting number of car, distributing capacity across multiple passenger classes). This assumption results in the operation cost to remain unchanged from the existing condition, and therefore reasonable for the RM strategy to consider revenue maximization as their primary objective.

1.3 Contributions

This dissertation develops and estimates models of railway passenger choice that better reflect the characteristics of the choice environment in the railway industry. The dissertation contributes to the existing literature of discrete choice model in RM by: (1) accounting for passenger taste heterogeneity in ticket purchase timing, (2) incorporating heterogeneous passenger choice models in RM optimization strategy, and (3) explicitly modeling inter-temporal effects on individual decisions in ticket cancellation and exchange.

With respect to the first category, this dissertation develops choice models of ticket purchase timing which account for taste heterogeneity across different categories of passengers based on latent class and mixed logit model frameworks. We develop an alternative approach to segment railway demand without using trip purpose. Since trip purpose is not available in the ticket reservation data, other elements in the ticket reservation data such as the characteristics of the trip (departure schedule) are used. The proposed approach has several advantages over previous studies in the literature based on deterministic segmentation of the demand. This approach leads to a more intuitive segmentation of the market between time-sensitive business travelers and price-sensitive leisure travelers.

With respect to the second category, we develop RM optimization models which incorporate the proposed choice models into ticket revenue optimization process. The optimizations account for both passenger demand volume and purchase timing in response to RM policy and simultaneously solve for pricing and seat allocation. Strategies are subjected to capacity constraints determined on the basis of the railway network characteristics. In particular, the parameters estimated from the choice models and the demand functions are used as key inputs for passenger responses to RM policy.

Finally, this dissertation is the first study to model ticket cancellation and exchange behavior using a dynamic discrete choice model (DDCM) framework. The approach enables to account for inter-temporal behavior of strategic passengers who are considered to be forward-looking agents. A model is formulated as an optimal stopping problem to predict the timing of ticket cancellations and exchanges, as well as the new departure times of the exchanged tickets. The prediction of new departure times is expected to further support RM capacity planning.

1.4 Dissertation Overview

This dissertation is organized as follows:

Chapter 1 introduces the research background, objectives, and the expected contributions from this research. Chapter 2 provides a comprehensive review of discrete choice model used in RM focusing on its applications to airline and railway industries. Shortcomings in the existing studies are also identified and research directions for this dissertation are proposed. Chapter 3 presents methodologies for the passenger choice model applicable to disaggregate data including multinomial logit (MNL), latent class (LC), mixed logit (ML), and dynamic discrete choice model (DDCM). Chapter 4 presents descriptive statistics relative to the ticket reservation data used in this research. The analysis focuses on data related to passenger choice model development such as the distribution of demand, advance booking, and fare as well as ticket cancellation and exchange behavior. Chapter 5 proposes passenger choice models of ticket purchase timing which account for taste heterogeneity using methodologies presented in Chapter 3. In Chapter 6, the passenger demand functions are presented; they represent the demand volume of each market in response to RM policy based on linear and log-linear regression. In Chapter 7, the RM optimization models are developed for a single-leg problem and network problem. Both approaches account for passenger responses in term of purchase timing and demand volume developed in the previous two chapters. In Chapter 8, the choice model for ticket cancellation and exchange behavior is developed using dynamic discrete choice model (DDCM) framework. Finally, Chapter 9 presents a summary of major findings, contributions and suggested future research directions.

Chapter 2 : Literature Review

In this chapter, we review the existing literature in passenger choice model for revenue management (RM). First, a brief overview of RM problem is provided. Then, an extensive review on discrete choice model in RM is presented, focusing on its application in airline and railway industries as well as strength and weakness of each approach. Then, a comprehensive review in dynamic discrete choice model (DDCM) formulation is presented, focusing on the setting and application related to this research. Next, we review approaches used in ticket cancellation and exchange modeling. Finally, the review summary is presented, some of the shortcomings of the existing studies are identified, and the research directions for this dissertation are described.

2.1 Fundamentals of Revenue Management Problem

Revenue management (RM) is the analytical application that predicts consumer behavior and optimizes product availability and price to maximize revenue. To achieve RM objective, it is essential for the firm to understand customers' perception of product value and accurately align product prices, placement, and availability with each customer segment in a profitable manner.

In the railway industry, given the short term cost of the operation are largely fixed and the variable costs per passenger are small, the problem reduces to seeking booking policies that maximize revenues. The fundamental revenue management strategy consists of four major elements: forecasting, overbooking, seat inventory control, and pricing. Given that overbooking has not been widely practiced in railway industry, three aspects of RM will be discussed: demand forecast, seat inventory control, and pricing.

2.1.1 Demand Forecast

Forecasting is essential for RM in determining the number of reservations which should be accepted based on seat allocation policy. RM demand forecasting primarily relies on two major approaches; aggregate and disaggregate. Aggregate approach focuses on using short term booking data to predict the future demand where the focus is on the total number of passenger demand instead of individual passenger behavior. The approach typically relies on simple smoothing technique to incorporate partial booking data from related flights at different phases in their booking process (McGill and van Ryzin, 1999). Disaggregate approach involves the use of discrete choice model. The approach is based on discrete choice transportation demand modeling framework of Ben-Akiva (1987). Discrete choice analysis has been applied in several choice behavior of the revenue management industry, for instance, path preference models (Hopperstad, 1994), and a deterministic model of demand behavior under price changes which incorporate diversion and recapture of passenger in different booking classes (Gallego, 1996).

2.1.2 Seat Allocation

The problem of seat allocation (seat inventory control) has been studied by many researchers since 1972. The approach practiced in airline RM problem is generally based on expected marginal seat revenue (EMSR) which is a refinement of the displacement cost valuation technique. The seat inventory control problem ranges from single-leg seat inventory control which considers a single flight leg to network seat inventory control with the consideration of network effects. Due to hub-and-spoke nature of the airline industry since the 1980s, a network seat inventory control has been significantly considered to account for passenger making flight connection. Several approaches which deals with network revenue management have been studied such as mathematical programming, segment control, virtual nesting, and bid price methods. The detailed description of these techniques are in McGill and van Ryzin (1999) and Talluri and van Ryzin (2004a).

2.1.3 Pricing

Pricing is essential for RM process in several aspects. In quantity-based RM, the existence of differential pricing of airline seats is considered as given. In price-based RM, price is the most important determinant of passenger demand behavior. Pricing and seat inventory control are dual to one another (Gallego and van Ryzin, 1997), however, there exists limited number of studies on joint seat allocation and pricing problem. McGill and van Ryzin (1999) suggests that the detailed empirical studies of the behaviors of different passenger types in response to changes in fare is a promising research directions for improving accuracy in demand forecast.

2.1.4 Joint Pricing and Seat Allocation

In the traditional RM optimization problem, pricing and seat allocation have often been considered as two independent problems. On one hand, the pricing focused on demand segmentation and optimal fare regardless of any capacity constraint. On the other hand, seat allocation focuses on setting booking limits by fare products based on fare and capacity constraint. Nevertheless, the two problems are interrelated and complementary to one another; the prices charged influence demand where acceptable demand is determined by seat allocation, and should be jointly considered as a single optimization problem. As noted by McGill and van Ryzin (1999), the integration of pricing and inventory allocation decisions should receive more attention by analysts in RM. In this context, Weatherford (1997) emphasized the importance of considering prices as part of the optimization problem and suggested including them as decision variables in the seat allocation problem. The author considered a single flight leg with at least two fare products. The demand for each fare product was assumed to be normally distributed and represented by a linear demand function of the competing products' fare.

Kuyumcu and Garcia-Diaz (2000) studied joint pricing and seat allocation problems for an airline network using historical data. The optimization problem aimed at maximizing the total revenues within the network. The study assumed that demand is normally distributed and that there is no interaction of demand across fare classes, and markets (origin-destination pairs). Fare was assumed to be an exogenous variable for the passenger decision process, as no explicit hypothesis regarding the relationship between demand and fare or any other product characteristics was made.

Bertsimas and de Boer (2002) analyzed the joint problem of pricing and seat allocation in a network setting. The authors assumed that demand for each fare product was uncertain and that expectation of the product demand only depends on the product's price. The numerical experiment suggested that coordination of pricing and seat allocation policies in the network and accounting for demand uncertainty can lead to significant revenue gains. It was also demonstrated that the underlying optimization problem is convex for certain types of demand distributions, thus tractable for large instances.

Cote et al. (2003) proposed a model with the capability of jointly solving the pricing and seat allocation problem in a network with competitor. The approach was based on a bi-level programming framework: the airline was assumed to know how its competitor would react and this behavior was explicitly integrated in the decision process. The decision variables include fares with seat allocation considered as constraints. The main assumption was that the demand for each fare product and itinerary combination was assumed to be fully known without relationship between fare and demand.

Ongprasert (2006) studied the seat allocation problem for intercity high speed rail services in Japan. The analysis includes: revenue maximization, average passenger load factor (APLF), and the number of passenger rejection. The choice model was estimated using nested logit model where the upper level consists of two transportation alternatives (high speed rail and airlines) and the lower level consists of fare product alternatives. The passenger choice model was incorporated in the seat allocation optimization problem which accounted for shared capacity of the railway network. Results show that seat allocation accounting for passenger choice behavior contributes to revenue improvement by offering discounted fare in the off peak trip.

Chew et al. (2008) developed a joint optimization model of pricing and seat allocation for a single product with a two period lifetime. Product price was assumed to increase as the time it perishes approaches, while demand is expressed as a linear function of price. To maximize the expected revenue, a discrete time dynamic programming model was developed to obtain the optimal prices and the optimal inventory allocations. Based on the concave property of the objective function, the authors used an iterative procedure to find the optimal solution. The problem was also extended to multiple time periods, where the concavity property no longer holds; for this case, several heuristics were suggested to solve the problem.

Cizaire (2011) developed several approaches to solve the joint problem of airline optimal fare and seat allocation. The underlying demand volume is modeled as a function of fares. The analysis proposes both deterministic and stochastic approaches to model demand; in particular, heuristics were developed to solve the stochastic problem. The problem considers a multiple products, multiple time periods without network considerations.

2.2 Discrete Choice Model in Revenue Management

Discrete choice analysis (DCA) is an ongoing field of research in RM which assumes that passenger makes choice among list of product alternatives. The approach is capable of revealing factors that influence passenger decisions in term of product attributes, trip characteristics, and passenger profile. In recent years, DCA has been applied in several RM industries. Airlines are considered to be the industry where DCA has been most widely applied. However, the railway industry has also seen ongoing application of DCA for its RM system.

2.2.1 Airline Industry

One of the most recognizable studies in the discrete choice model application for airline industry are the work of Andersson (1998) and Algers and Besser (2001) who applied logit choice models to estimate buy-up and recapture factors at one of Scandinavian Airline Systems (SAS) hubs. Another line of research focusing on understanding choice behavior is the passenger origin and destination simulator (PODS) studies of Belobaba and Hopperstad (1999) which is a simulation model of passenger purchase behavior that accounts for airline preference, time preference, path preference, and price sensitivity.

Talluri and van Ryzin (2004) analyzed a single-leg RM problem where a consumer choice behavior is modeled by a general discrete choice model. Their choice model specifies the probability of purchasing each product as a function of the set of available fare products. Their approach is based on expectation maximization (EM) method which overcomes the limitation of incomplete data when only purchase transaction data are available but it is not possible to distinguish a period without an arrival from a period in with an arrival without customer making purchase. The model is applied to the buy-up and buy-down behavior of airline product to support RM decisions for the set of product offering.

The feasibility and benefit of the discrete choice analysis in RM was examined in Vulcano et al. (2008), which focused on buy-up, buy-down, and diversion of airline market. They used the same approach as Talluri and van Ryzin (2004) to address unobservable shopping data issue using variation of expectation maximization (EM) method. Their choice model was estimated with multinomial logit (MNL) model where the choice set consists of all the flights offered by multiple airlines on a given day between specific pair of airports. Their result indicates the revenue improvement of 1.4% to 5.3% in the tested markets. Their study suggests testing a model with unobservable segment (latent class) to allow for the model to predict more accurately.

2.2.2 Railway Industry

Railways have several characteristics different from airline. A railway trip generally consists of more legs due to more station stops. Each leg, defined by pair of stop, must be determined in terms of opportunity cost or capacity allocation. Walk up ticket is general for railway, where passenger purchases the ticket on the day of departure, especially for high speed service. Other RM policies used by the airlines such as overbooking, nested fare structure are typically not considered in railway RM.

Ciancimino et al. (1999) is the first published work which deals with passenger railway RM. They studied a single-fare multi-leg capacity allocation problem with an objective of allocating a specific number of seats for each origin-destination pair to maximize revenue for the entire journey. Their analysis involves deterministic and probabilistic approaches. The deterministic approach is based on linear programming while the probabilistic approach involves modeling the service demand with truncated normal distribution. Hood (2000) developed a logit choice model to incorporate demand estimation in time tabling and pricing decisions. Their result indicates similar demand estimation to the observed values with computational time burden.

Whelan and Johnson (2004), and Whelan et al. (2008) estimated a nested logit model to evaluate the impact of fare structure on train overcrowding. 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. Li et al. (2006) conducted simulation study to assess dynamic pricing policy for railway RM and overall network performance. They proposed passenger response model as an activity based model accounting for time shifting, mode choice, departure time choice, and route choice. Their model is based on two steps, first the demand is modeled with micro level behavior model to respond with parametric policies such as pricing and then the interaction is simulated by linking the demand and supply simulation by using sensitivity of demand information to optimize the policy parameters. The forms of dynamic pricing considered include time-based pricing (time of day, and day of week), regional-based, and route-based and other discount availability.

Sibdari et al. (2008) studied dynamic pricing policy for Amtrak Auto Train. They proposed discrete-time multi product dynamic pricing model which is updated on a daily basis. The choice model involves a multi-stage decision process similar to the model proposed by Whelan and Johnson (2004) and by Whelan et al. (2008); passengers make the decisions to buy or not to buy and whether to upgrade the accommodation or not. Their 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. Passenger demand is specified as Poisson random variable with specified mean and a passenger demand on a given day is a function of remaining time before departure, car accommodation price, and coach seat price.

Ongprasert (2006) studied seat allocation problem for the Japan intercity high speed rail service. The problem focuses on seat allocation problem with three objectives, revenue maximization, average passenger load factor (APLF), and the number of passenger rejection. The choice model was estimated with nested logit model where the upper level consists of two transportation alternatives; high speed rail and airlines, and the lower level consists of fare product alternatives. The passenger choice model is incorporated in the seat allocation optimization problem which accounts for shared capacity of the railway network. Results show that seat allocation accounting for passenger choice behavior contributes to revenue improvement by offering discounted fare in the off peak trip.

2.2.3 Choice Model Accounting for Taste Heterogeneity

Accounting for taste heterogeneity is essential for demand forecasting especially for railway RM because passenger preferences generally vary by departure time of day, day of week, and trip distance. A number of papers have investigated special classes of discrete choice model that accommodates taste heterogeneity. Bhat (1998) estimated an intercity travel mode choice model which accommodates variations in response to level of service measures due to observed and unobserved individual characteristics. The study emphasized the necessity to incorporate systematic and random variation in responsiveness to level-of-service variables. Greene and Hensher (2003) compared latent class (LC) with mixed logit (ML) model using stated preference data on long distance travel survey in 2000. Shen (2009) compared the difference between latent class and mixed logit models using two stated choice survey data sets from Osaka, Japan relative to mode choice using non-nested test to compare the model fits.

In the RM context, Carrier (2008) analyzed the choice of airline itinerary and fare product based on latent class (LC) model framework. In this model, passenger choice set was constituted from booking data, fare rules, and seat availability data. Instead of segmenting passenger by trip purpose, which is not available in booking data, the author utilizes variables such as frequent flyer membership, ticket distribution channel, and travel day of week for the class membership model. The approach is shown to provide a more distinct and intuitive segmentation across passengers. Teichert et al. (2008) applied the latent class model to explore preferences within airlines segments and analyzed respondents' profiles in terms of individual socioeconomic and trip characteristics. They concluded that the segmentation criterion currently applied by airlines does not adequately mirror the heterogeneity in customer's preference patterns. They suggested that product marketing be aligned to passenger attitudes and socio-demographic profile which are different across passenger segment. Wen and Lai (2010) used latent class model to identify airline passengers' potential segments and preferences for international air carriers using individual socioeconomic and trip characteristics as class membership variables. The latent class model is capable of representing heterogeneity across passenger segments which results in improved prediction accuracy over the multinomial logit model. Specifically, the willingness to pay for service attribute improvements is found to be substantially different across air routes and to vary by traveler segment.

Regardless of the number of research efforts on heterogeneity in choice behavior, the application of heterogeneous choice model in RM problem is still relatively limited. Most of the studies which incorporated choice models in RM problem have assumed that customers are homogeneous in taste preferences. Studies which rely on this assumption include Zhang and Adelman (2009); Topaloglu (2009); and Erdelyi and Topaloglu (2010) who incorporated customer choice models in the network RM pricing. In their setting, the price for each product is chosen from a discrete set, and the demand for each product depends on the price of the product only. However, given that RM relies on the premise that different customers are willing to pay different amounts for a product, accounting for passenger heterogeneity is expected to provide high yield toward RM strategy.

Recently, a limited number of studies which incorporate heterogeneous passenger choice model in RM problem have primarily focused on choice-based deterministic linear programming (CDLP) problem. CDLP is a class of revenue optimization which solves for sets of product to be made available to the customers at different points in time during the sales horizon. In this context, Rusmevichientong et al. (2012) analyzed a model that captures the substitution between the products and preference heterogeneity. Each customer is assumed to belong to a particular class and the demand from each customer class is governed by a multinomial logit choice model with class-dependent parameters. This problem considers a set of different products and maximizes the expected profit across all customer classes. Mendez-Diaz et al. (2012) specified LC model which divides customers into segments based on choice of product alternatives considered by each customer. Their demand model allows product to overlap across segments and the preference parameters for each product alternative in the logit model are assumed to be known in advance. The authors proved that the latent class logit assortment problem is NP-Hard, and solved the choice-based deterministic linear program (CDLP) using branch and cut approximation method. The procedure was tested in the context of both capacitated and un-capacitated retail assortment problems.

2.3 Dynamic Discrete Choice Model

Dynamic discrete choice models have been firstly developed in economics and applied to study a variety of problems that include fertility and child mortality Wolpin (1984), occupational choice Miller (1984), patent renewal Pakes (1986), and bus engine replacement Rust (1987). In general dynamic discrete choice structural models, agents are forward-looking and maximize expected inter-temporal payoffs; the consumers get to know the rapidly evolving nature of product attributes within a given period of time and different products are supposed to be available on the market. The timing of consumers' purchases is formalized as an optimal stopping problem where the agent (consumer) must decide on the optimal time of purchase.

Existing studies on choice modeling for revenue management (RM) have mostly ignored temporal effects in individual decision making. Although static models enable analysts to address the dependence of demand on the set of products offered by the provider, they are unable to model forward-looking agents who would typically wait and see before making the final decision. There is an emerging research effort toward dynamic frameworks that account for inter-temporal variability in choice modeling. Existing research on inter-temporal price variation that considers forward-looking consumers includes Stokey (1979), Landsberger and Meilijson (1985), and Besanko and Winston (1990). These papers are based on the assumptions that customers are present in the market throughout the entire season, and that the seller's inventory is practically unlimited. Customers purchase at most one unit during the season, and they optimally select the timing of their purchases so as to maximize individual surplus.

Su (2007) studied a model of strategic customer by identifying four customer classes, different from each other in two dimensions: high versus low valuations and strategic (i.e., patient) versus myopic (impatient) behavior. The price path is assumed to be predefined by the seller, and after the specific pricing policy is announced, strategic consumers can weigh the benefits of waiting for a discount (if any is offered). The paper demonstrates that the joint heterogeneity in valuations and in the degree of patience is crucial in explaining the structure of optimal pricing policies.

In RM, behavior of ticket cancellation and exchange is clearly influenced by demand uncertainty over time. Stokey (1979) showed that offering a single price can be optimal when inter-temporal differentiation is feasible, but assumes that consumers have perfect information on the future evolutions of their valuations. In Png's (1989), consumers face both uncertainty in their valuations as well as uncertainty about the capacity. Gale and Holmes (1992) examined advance purchase discounts where a monopoly firm offers two flights at different times and where consumers are assumed to not know their preferred flight in advance. In this study, advance purchase discounts are used to smooth the demand of the consumers with a low cost of time.

Gallego and Phillips (2004) used a similar approach in their work on flexible products. Dana (1998) showed that advance purchase discounts may improve the revenues of price-taking firms when consumer demand is uncertain. In this case, firms in competitive markets can improve profits by offering advance purchase discounts. Shugan and Xie (2000) developed an inter-temporal consumer choice model for advance purchase which distinguishes the act of purchasing and consumption. The model accounts for buyer's valuation of services that depends on buyer states at the time of consumption and assumes the product capacity to be unlimited. In a later paper, Xie and Shugan (2001) extended this analysis of advance selling to the finite-capacity case and introduced a refund option.

Ringbom and Shy (2004) proposed a model where consumers have the same deterministic valuation (maximum willingness to pay) for a certain service of product but different probabilities of showing up; capacity is assumed to be infinite and prices are endogenously given; results show that by adjusting partial refunds it is possible to endogenize the participation rates. Aviv and Pazgal (2008) considered an optimal pricing problem of a fashion-like seasonal good in the presence of strategic customers (forward-looking characteristics) with a time-varying valuation pattern. Customers have partial information about the availability of the inventory and their arrival is assumed to be time dependent. The system is characterized by a leader follower game under Nash equilibrium where customers select the timing of their purchase so as to maximize individual surplus while the seller maximizes expected revenue. Gallego and Sahin (2010) developed a model of customer purchase decision with evolution of trip schedule valuations over time. This analysis considers partial refundable fare based on a call option approach; each customer updates his/her valuation over time and decides when to issue and when to exercise options in a multi-period temporal horizon.

Meanwhile, a number of studies on demand uncertainty have focused on the supply chain management approach. To our knowledge, Spinler et al. (2002, 2003) are among the first in the operations management literature that incorporated consumer's uncertainty in valuations into revenue management, and the first to study partially refundable fares. Other studies on uncertain valuations for traditional revenue management problems include Levin et al. (2009), Yu et al. (2008), and Koenigsberg et al. (2006). There is also an emerging literature that deals with strategic consumers who develop expectations on future prices and product availability based on the observed history of prices and availabilities (e.g. Besanko and Winston 1990, Gallego et al., 2009, Liu and van Ryzin, 2005, Aviv and Pazgal, 2008).

Based on this comprehensive literature review, some of the most relevant studies for our dynamic discrete choice model formulation are described as follows.

2.3.1 Optimal Replacement of Bus Engine

Rust (1987) examined the replacement investment decisions at the level of individual agent (in this case is the bus manager). The problem considered the decision of how long the bus should be operated before the engine was replaced with a new or completely overhauled bus engine. This problem is represented as a discrete decision process with a state variable of cumulative mileage on a bus since last engine replacement, and control variable which is maintenance manager's decisions on bus engine replacement. The problem represents a regenerative process as the bus is considered as good as new when its engine is replaced. The problem is formulated as an optimal stopping problem with the mileage process specified as a regenerative random walk and estimated with maximum likelihood estimator using nested fixed point (NFXP) algorithm. The feature of this single individual level model enable for the model accuracy to be simply evaluated by checking with the individual agent whether the estimated utility function is reasonable. In this case, the study reveals that the model corresponded closely with the expectation of the decision maker.

2.3.2 Durable Goods with Technological Evolution

Melnikov (2000) developed estimation to analyze the impact of technological change on dynamic of consumer demand for durable product. The timing of consumer purchase was formalized as an optimal stopping problem where the solution to the problem defines a hazard rate of the product adoption. The empirical analysis of the study is based on the data on U.S. computer printer market. The empirical result support the hypothesis of forward long consumer's behavior enabling for better demand forecast.

2.3.3 New Durable Goods with Heterogeneous Consumer Taste

Gowrisankaran and Rysman (2009) developed a dynamic model of consumer preferences for new durable goods with heterogeneous consumer taste. The framework was applied to the new durable goods with infinite time horizon. The products are assumed to be infinitely durable and follow Markov process based on the firm optimizing behavior. The decision problem for consumer i at time t is to make decision between (1) choosing among one of among products available in period t and (2) choosing not to purchase any product in the current period. The consumer is then faced with similar but not identical decision problem at time $t+1$ and onward. The decision maker is assumed to make a decision which maximizes the sum of the expected discounted utility conditional on information at time t.

2.4 Ticket Cancellation and Exchange Model

In the context of ticket cancellation and exchange model, a number of papers have been published in the past decade. Garrow and Koppelman (2004a) proposed an airline cancellation and exchange behavior model based on disaggregate passenger data; airline travelers' no-show and standby behavior is modeled using a multinomial logit (MNL) model estimated on domestic US itineraries data. The approach enables the identification of rescheduling behavior based on passenger and itinerary characteristics and supports a broad range of managerial decisions. Variable used to identify passenger rescheduling behavior are traveler characteristics, familiarity to the air transportation system, availability of viable transportation alternatives, and trip characteristics. Garrow and Koppelman (2004b) extended their work by introducing a nested logit structure and demonstrated the benefit of directional itinerary information. The nested logit (NL) tree groups show, early standby, and late standby alternatives in one nest and no show alternative in another nest. The analysis emphasized the superiority of nested logit model structure over multinomial logit model and the importance of distinguishing between outbound and inbound itineraries.

Iliescu et al. (2008) further expanded the work of Garrow and Koppelman (2004a, 2004b) by proposing a discrete time proportional odds (DTPO) model to predict the occurrence of ticket cancellation and exchange based on the Airline Reporting Corporation (ARC) data. The cancellation probability is defined as a conditional probability that a purchased ticket will be canceled in a specific time period given it survived up to that point (hazard probability). Results show that the intensity of cancellation is strongly influenced by the time from the ticket purchase and the time before flight departure as well as by other covariates (departure day of week, market, group size, etc.). Specifically, higher cancellation is observed for recently purchased ticket and ticket which associated departure dates are near.

Graham et al. (2010) adopted discrete time proportional odds (DTPO) model to investigate when and why travelers make changes to their airline itineraries. Analysis is based on a nine month period panel data of university employees in Atlanta, US. The analysis focused on tickets issued less than 60 days before the outbound departure date. The use of panel data enabled the analysts to study how cancellation behavior differs by frequency of travel as well as by carrier. The deriving empirical analysis identifies the reasons why business travelers exchange their ticket, and concluded that differences exists between outbound and inbound itineraries, between exchange and cancellation rates for frequent and infrequent business travelers, across air carriers and timing when refund and exchange events occur. The results also indicate that the timing of cancellation exhibit a strong pattern, i.e., ticket changes are two to three time more likely to happen within the first week after purchase and are more likely to occur as the departure date approaches.

2.5 Review Summary and Research Directions

Based on the literature review, this section summarizes major studies related to this research and identifies promising research directions for this dissertation. Table 2.1 represents major studies related to passenger choice models in the RM applications. In term of model characteristics, it is said to be multi-fare if there are multiple fare offered for the same set of resources. A multi-leg indicates whether the network (in airline problem), or the sequential network nature (railway problem) is addressed in the problem. Dynamic pricing indicates whether the price of the product changes during the sale horizon via optimization step. Table 2.2 represents major studies of dynamic choice models where the framework is relevant to this research. Table 2.3 represents major studies of RM revenue optimization which incorporate passenger demand model in their framework focusing on the formulation of joint pricing and seat allocation problem. Consideration are given to the number of products and time periods accounted, the presence of a competitor, the number of legs in the network, the approaches used to model demand and to optimize the problem of pricing and seat allocation. These three tables also compare the studies proposed in this dissertation to the existing studies reviewed.

Table 2.1 Summary of static choice models

Authors	Application	RM Policy	Model	Characteristics
Andersson (1998); Algers and Besser (2001)	Buy-up and recapture factor for airline	Pricing	Logit	-
Talluri and van Ryzin (2004)	Developing model framework	-	Logit	Multi-fare, capacity allocation
Ongprasert (2006)	Fare discount strategy for Japan intercity high speed rail service	Pricing	Nested Logit	Multi-fare, multi-leg, capacity allocation
Vulcano et al. (2008)	Airline single-leg market	Pricing	Logit with expectation maximization (EM)	Multi-fare, capacity allocation
Carrier (2008)	Airline Itinerary and fare product choice	-	Latent Class (LC) model	Multi-fare, multi-leg, capacity allocation
Sibdari et al. (2008)	Amtrak auto train dynamic pricing	Dynamic Pricing	Poisson demand model	Capacity allocation
Whelan et al. (2008)	Fare structure on train overcrowding	-	Nested Logit	Multi-fare, multi-leg
This research	Ticket purchase timing	Joint pricing and seat allocation	MNL, LC, ML (parametric, non-parametric)	Single-product, multiple-time periods, heterogeneous passengers

Table 2.2 Summary of dynamic choice models

Authors	Application	Model	Data	Characteristics	Estimation Method
Rust (1987)	Bus engine replacement	Logit	10 years monthly data on bus mileage (104 buses)	Single-agent, homogeneous product attribute	Nested fixed point
Melnikov (2000)	Computer printer	Nested logit	Monthly sales data (27 manufacturers)	Homogeneous consumers with one purchase, differentiated durable products	Nested three step method
Gowrisankaran and Rysman (2007)	Digital camcorder demand	Logit	Monthly data of 378 models and 11 brands, number of units sold	Repeated purchase, heterogeneous consumers and differentiated products	Three levels of non-linear optimization
Iliescu et al. (2008)	Ticket cancellation/exchange (application in RM overbooking)	Discrete time proportional odd (DTPO) model	Airline ticketing data	Heterogeneous consumers by segmentation covariates, time varying covariates.	Conditional probability maximum likelihood (DTPO likelihood function)
Graham et al. (2010)	Ticket cancellation/exchange	Discrete time proportional odd (DTPO) model	Airline revealed preference ticketing data	Heterogeneous consumers	Conditional probability maximum likelihood (DTPO likelihood function)
This research	Ticket cancellation/ exchange	Logit	Railway ticket reservation data	Heterogeneous customers by segmentation covariates, single exchange/cancel decision	Two step look-ahead DP approximation

Table 2.3 Summary of RM revenue optimization framework

Authors	No. of product>1	Demand Model	No. of time period>1	Multiple legs	Competitor	Joint pricing/seat allocation	Simultaneous Optimization
Weatherford (1997)	√	Linear function of price with cross elasticities	×	×	×	√	√
Kuyumcu and Garcia-Diaz (2000)	√	Normally distributed demand	×	√	×	√	×
Bertsimas and de Boer (2002)	√	Function of price	√	√	×	√	×
Cote et al. (2003)	√	Constant	×	√	√	√	×
Ongprasert (2006)	√	Nested logit	√	√	√	√	×
Chew et al. (2008)	×	Linear function of price	√	×	×	√	×
Iliescu (2008)	√	Discrete time proportional odd (DTPO)	√	×	×	×	×
Cizaire (2011)	√	Function of price	√	×	×	√	√
This research	×	MNL, LC, ML choice models, log-linear demand functions	√	√	×	√	√

Based on the review summary in Table 2.1 to Table 2.3, the research directions addressed by this dissertation are described as follows:

2.5.1 Railway Network Characteristic

The simplified RM problem in airline is generally formulated as a single-leg problem without interdependencies of RM strategies between legs. However, this simplification is less appropriate for railway because majority of trips are composed of multiple legs and the capacity of each leg is shared across multiple markets. To account for this characteristic, the network revenue optimization approach considered in this dissertation will take this aspect into account.

2.5.2 Joint Pricing and Seat Allocation

Based on the literature, some studies on joint pricing and seat allocation have been made for the network setting. However, few studies have proposed a simultaneous optimization of pricing and seat allocation. Most of them relied on iterative approach where each element is optimized sequentially. To this aspect, we believe that the railway operator revenues could be further improved by developing a simultaneous RM optimization model of pricing and seat allocation which takes choice model parameters as inputs for the demand side response. The analysis of the combined impacts of pricing and seat allocation with realistic demand response for RM strategy is expected to improve the revenue optimization process.

2.5.3 Passenger Taste Heterogeneity

The review of choice model in airline RM provides several promising approaches applicable for railway problem. Review shows that the latent class approach can address the issue of unobservable trip purpose from the booking data by classifying passenger into classes using other elements of booking data such as trip characteristics. The applicability of the approach to this research will then be explored. More importantly, the choice model accounting for heterogeneity can be incorporated in the RM optimization models and allows for more realistic demand side response, which can be useful for supporting range of RM decisions (e.g. pricing, seat allocation).

2.5.4 Inter-Temporal Effects in Ticket Cancellation and Exchange Behavior

Ticket cancellation and exchange is influenced by passenger uncertainty of trip schedule which involves inter-temporal effects in the decision process. This effect is highly observed among strategic customers where studies have emphasized the impact of their behavior toward RM (e.g. Castillo et al., 2008; Zhang, 2005).This study of inter-temporal effects on passengers' decisions has been very limited in the RM literature especially for ticket cancellation and exchange. Moreover, to date none of the ticket cancellation and exchange models in RM is capable of predicting a new departure time for the exchanged ticket. Thus, this dissertation aims to fulfill this gap by developing dynamic discrete choice model (DDCM) for ticket cancellation and exchange behavior.

Chapter 3 : Methodologies

This chapter summarizes several methodologies in modeling passenger choice with an emphasis the revenue management (RM) application. The review focuses on methods feasible for the booking data including static and dynamic discrete choice model such as multinomial logit (MNL) model, latent class (LC) model, mixed logit (ML) model, and the dynamic discrete choice model (DDCM).

3.1 Multinomial Logit (MNL) Model

The logit model is the easiest and most widely used discrete choice model. This is because its choice probabilities take a closed-form solution, thus allowing the estimation to be done numerically. In a logit model, a decision maker labeled n faces J alternatives. The utility that the decision maker obtains from alternative j is decomposed into deterministic term V_{nj} and error term ε_{nj} which is treated by researcher as random:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \;\; \forall j \tag{3.1}$$

The error term, ε_{nj} is assumed to be independently, identically distributed extreme value type I (Gumbel distribution). The probability that the decision maker chooses alternative i is:

$$P_{ni} = \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \;\; \forall j \neq i) \tag{3.2}$$

$$= \text{Prob}(\varepsilon_{nj} \; < \varepsilon_{ni} + V_{ni} - V_{nj} \;\; \forall j \neq i) \tag{3.3}$$

If ε_{ni} is considered given, the expression is the cumulative distribution for each ε_{nj} evaluated at $\varepsilon_{ni} + V_{ni} - V_{nj}$ which according to Gumbel distribution (Type I extreme value) is $\exp(-\exp(-(\varepsilon_{ni}+V_{ni}-V_{nj})))$. Since ε's are independent, the cumulative distribution over all $j \neq i$ is the product of the individual cumulative distribution:

$$P_{ni}|\varepsilon_{ni} = \prod_{j \neq i} \exp(-\exp(-(\varepsilon_{ni}+V_{ni}-V_{nj}))) \tag{3.4}$$

With ε_{ni} not given, the choice probability is the integral of $P_{ni}|\varepsilon_{ni}$ over all values of ε_{ni} weighted by Gumbel distribution density:

$$P_{ni}|\varepsilon_{ni} = \int \left(\prod_{j \neq i} \exp(-\exp(-(\varepsilon_{ni}+V_{ni}-V_{nj}))) \right) \exp(-\varepsilon_{ni}) \exp(-\exp(-\varepsilon_{ni})) d\varepsilon_{ni} \tag{3.5}$$

This yields the result in closed form solution:

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_j \exp(V_{nj})} \tag{3.6}$$

where utility is usually specified to be linear in parameters:

$$V_{nj} = \beta' x_{nj} \tag{3.7}$$

where x_{nj} is a vector of attributes related to alternative j thus, the logit choice probability becomes:

$$P_{ni} = \frac{\exp(\beta' x_{ni})}{\sum_j \exp(\beta' x_{nj})} \tag{3.8}$$

3.2 Mixed Logit (ML) Model

Mixed logit is a highly flexible model capable of approximating any random utility model (McFadden &Train, 2000). It obviates three limitations of the standard logit model by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time. From fare pricing perspective, distribution imposed on individual's preference for an attribute enables an analyst to investigate the hypothesis that some individuals are more price conscious than other individuals, and therefore improve fare strategy.

Mixed logit probabilities are the integral of standard logit probabilities over a density of parameters (β). Choice probabilities of a mixed logit model can be expressed in the form:

$$P_{ni} = \int L_{ni}\,(\beta) f(\beta) d\,\beta \tag{3.9}$$

where $\int L_{ni}\,(\beta)$ is the logit probability evaluated at parameter β:

$$L_{ni}(\beta) = \frac{\exp(V_{ni}(\beta))}{\sum_{j=1}^{J} \exp(V_{nj}(\beta))} \tag{3.10}$$

and $f(\beta)$ is a density function. $V_{ni}(\beta)$ is deterministic term observed by the analyst, which depends on the parameters β. Usually, the utility is linear in β, thus $V_{ni}(\beta) = \beta' x_{ni}$. The mixed logit probability then takes its usual form:

$$P_{ni} = \int \left(\frac{\exp(\beta' x_{ni})}{\sum_{J} \exp(\beta' x_{nj})} \right) f(\beta) d(\beta) \tag{3.11}$$

Mixed logit model can be viewed as a mixture of the logit function evaluated at different β's with $f(\beta)$ as the mixing distribution.

3.3 Latent Class (LC) Model

Latent class model is based on discrete segmentation which assumes that heterogeneity in passenger behavior is likely to be driven by specific elements. The approach generally group observations into meaningful segments which have similar needs, constraints, and preferences. For instance, passengers can be segmented to business and leisure travelers with different preferences parameters defined.

In latent class model, segmentation is done through probabilistic approach which link explanatory variable such as trip's characteristics and passenger profile into class membership model when assigning passenger into classes. The class membership model is combined with choice model enabling the model to account for differences in choice behavior between different segments of the market.

The structure of the latent class (LC) passenger choice model could be described as follows. Let i represents alternative from $1, \dots, J_b$ in the choice set C of booking b. The model form can be written as:

$$P(i|X_M, X_C) = \sum_{s=1}^{S} P(s|\, X_M) P(i|X_C, s) \;\; \forall i \in C \tag{3.12}$$

where

s	is class index; {1, 2, …, S }
X_M	is class membership explanatory variable
X_C	is class specific choice models explanatory variable

The utility function of alternative *i* given the customer is in the class *s* can be written as:

$$U_{ib} = X_{Cib}\beta_C + \varepsilon_{ib} \tag{3.13}$$

where

X_{Cib}	is a (*1xK*) vector of choice model explanatory variable
β_C	is a (*Kx1*) vector of parameters
ε_{ib}	is a random disturbance (i.i.d. extreme value)

The class specific choice probability of travel option *i* can be expressed as:

$$P(i|X_{Cib}, s) = \frac{\exp(X_{Cib,s}\beta_{C,s})}{\sum_{j=1}^{J_b} \exp(X_{Cjb,s}\beta_{C,s})} \quad \forall s \in S \quad \forall i \in C \tag{3.14}$$

where $\beta_{C,s}$ are the unknown parameters of the class-specific choice model. The probability of belonging to the latent class *s* can be written as:

$$P(s|X_{Mb}) = \frac{\exp(X_{Mb,s}\beta_M)}{\sum_{s=1}^{S} \exp(X_{Mb,s}\beta_M)} \tag{3.15}$$

where β_M are the unknown parameters for class membership model.

3.4 Dynamic Discrete Choice Model (DDCM)

3.4.1 Structural Estimation of Markov Decision Process

Markov decision process has a rich framework in modeling problem choices made over time under uncertainty. The framework is generalized as sequential decision making with two types of variable, state variables s_t and control variables d_t both of which indexed by time $t = 0,1,\dots,T$. The decision maker is presented by set of primitives (u, p, β). $u(s_t, d_t)$ is a utility function representing agent's preference at time t. $p(s_{t+1}|s_t, d_t)$ is Markov transition probability representing agent's belief about uncertain future state. $\beta \in [0,1)$ is the rate at which the agent discount utility in future period. Agents are assumed to be rational behaving according to an optimal decision rule $d_t = \delta(s_t)$ which solves

$$V_{t=0}^{T}(s) \equiv \max_{\delta} E_{\delta}\left\{\sum_{t=0}^{T} \beta^t u(s_t, d_t)|s_0 = s\right\} \tag{3.16}$$

where E_δ denotes expectation with respect to controlled stochastic process $\{d_t, s_t\}$ induced by the decision rule δ.

3.4.2 General Case of DDCM Problem

In discrete decision process, the agents are forward-looking which choose a decision rule $d_t = \delta(x_t, \varepsilon_t, \theta)$ to maximize their expected discounted utility over the time horizon with discount factor $\beta \in [0,1)$. This could be represented as the inter-temporal optimization problem in which the solution is given recursively by Bellman's equation:

$$V_\theta(x, \varepsilon) = \max_{d \in D(x)} \left[u(x, d, \theta) + \varepsilon(d) + \beta \int V_\theta(x', \varepsilon') \pi(dx', d\varepsilon' | x, \varepsilon, \theta) \right] \qquad (3.17)$$

The function $V_\theta(x, \varepsilon)$ is the maximum expected discount utility obtainable by the agent when the state variable is (x, ε). We then define the expected value function, EV_θ by

$$EV_\theta(x, \varepsilon, d) = \int V_\theta(x', \varepsilon') \pi(dx', d\varepsilon' | x, \varepsilon, \theta) \qquad (3.18)$$

This allows us to write Bellman's equation in a slightly simpler notation as:

$$V_\theta(x, \varepsilon) = \max_{d \in D(x)} [u(x, d, \theta) + \varepsilon(d) + \beta EV_\theta(x, \varepsilon, d)] \qquad (3.19)$$

The conditional independence (CI) enables us to simplify the above problem by providing simple formula for the likelihood function. CI implies:

$$\pi\left(x_{t+1}, \varepsilon_{t+1} \middle| x_t, \varepsilon_{t,} d_t, \theta\right) = p(x_{t+1} | x_t, d_t, \theta) q(\varepsilon_t | x_t, \theta) \qquad (3.20)$$

By assuming further that q is a multivariate extreme value distribution, the full likelihood function for the sample data could be shown by the function:

$$L(\theta) \equiv L(x_1, \dots, x_T, d_1, \dots, d_T | \vartheta) = \prod_{t=2}^{T} P(d_t | x_t, \theta) \, p(x_t | x_{t-1}, d_{t-1}, \theta) \qquad (3.21)$$

Dynamic discrete choice modeling (DDCM) involves estimation of the unknown parameter vector θ. The estimation procedure which is maximum likelihood in the case of DDCM has nested problem structure. It involves the estimation of θ that maximizes the full likelihood function or the partial likelihood function as:

$$L_P(\theta) \equiv \prod_{t=1}^{T} P(d_t|x_t, \theta) \tag{3.22}$$

Subject to the constraint that the function EV_θ is given by the unique fixed point to the contraction mapping $T_\theta EV_\theta = EV_\theta$ defined by:

$$EV_\theta(x, d) = T_\theta EV_\theta(x, d)$$

$$\equiv \int \log\left[\sum_{d\in D(x')} \exp\{u(x', d', \theta) + \beta EV_\theta(x', d')\}\right] p(x', d', \theta) \tag{3.23}$$

where the conditional choice probability $P(d\,|\,x, \theta)$ is given by classical multinomial logit formula:

$$P(d|x, \theta) = \frac{\exp\{u(x, d, \theta) + \beta EV_\theta(x, d)\}}{\sum_{d'\in D(x')} \exp\{u(x, d', \theta) + \beta EV_\theta(x, d')\}} \tag{3.24}$$

3.5 Conclusions

This chapter reviews passenger choice model based on several methodologies from static model (MNL, ML, and LC) to dynamic model (DDCM). The chapter provides methodological foundation for the development of ticket purchase timing model accounting for taste heterogeneity (in Chapter 5) and dynamic discrete choice model of ticket cancellation and exchange (in Chapter 8).

Chapter 4 : Data Analysis

This chapter analyzes the data used in this research. The analysis focuses on elements of ticket reservation data useful for developing the passenger choice models in Chapter 5 and Chapter 8, and demand functions in Chapter 6. The analysis focuses on distribution of passenger demand, advance purchase, fare, and behavior of ticket cancellation and exchange.

4.1 Analysis of Passenger Demand

Ticket reservation data of the US intercity passenger railway collected over two months period in 2009 are used for the analysis. Data contain information in terms of trip origin, destination, fare class, reservation date, departure date, departure time, arrival time, fare, and accommodation charge. Data analysis focuses on the first month data which contains the total of 406,422 observations.

This railway service under consideration consists of two ticket classes: first class, and coach class. This research focuses on coach class passengers which are the predominant demand of this railway service (accounting for 92% of the total demand) and only on reservations which are eventually confirmed and paid. To reduce the problem size, the analysis focuses on the northbound trip. This results in a final data set of 110,828 observations. Overview of data statistics in term of distribution of passenger demand, advance booking, and fare are as follows:

4.1.1 Demand by Trip Schedule

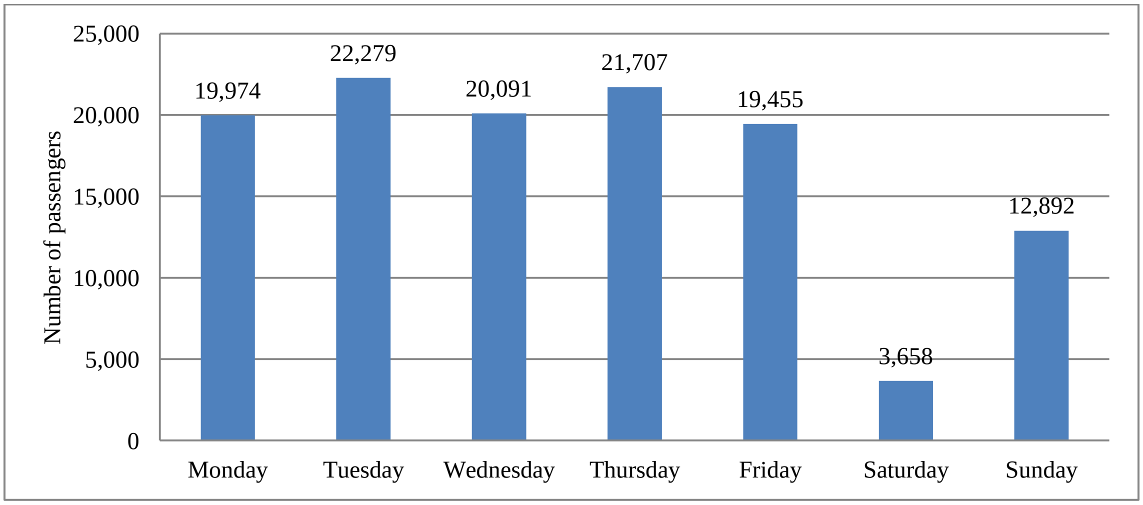


Figure 4.1 Monthly passenger demand by day of week

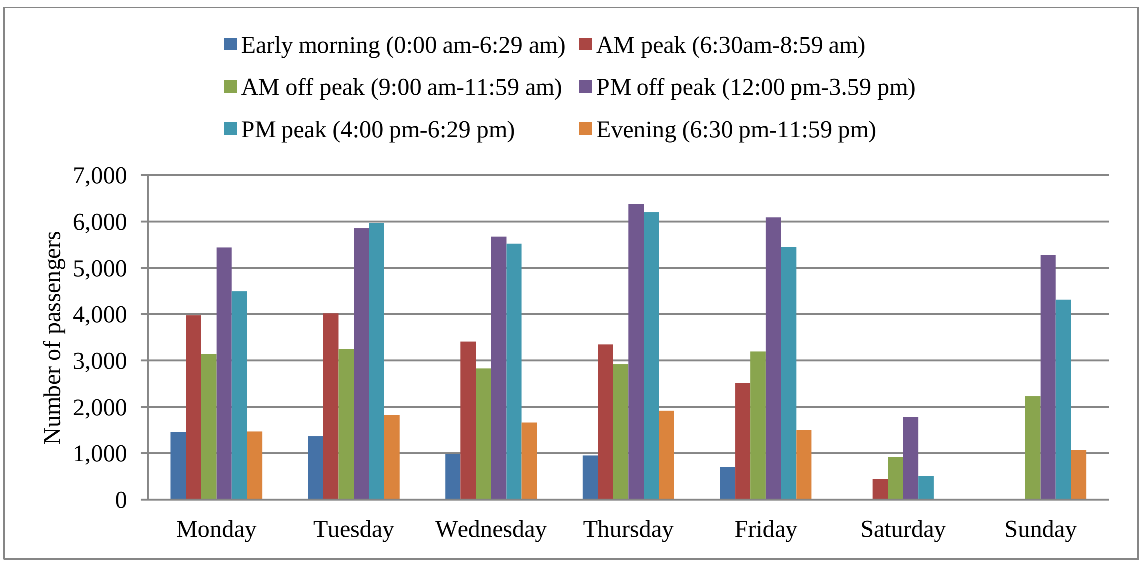


Figure 4.2 Monthly passenger demand by day of week and time of day

Figure 4.1 shows passenger demand distribution by day of week which indicates higher demand on weekday rather than weekend. Saturday departure shows least demand compared to other days of week. Figure 4.2 shows passenger demand distribution by day of week and time of day[1] which indicates high demand for passengers departing during PM off peak (12:00-3:59 PM.) and during PM peak (4:00-6:29 PM.).

4.1.2 Advance Purchase Behavior

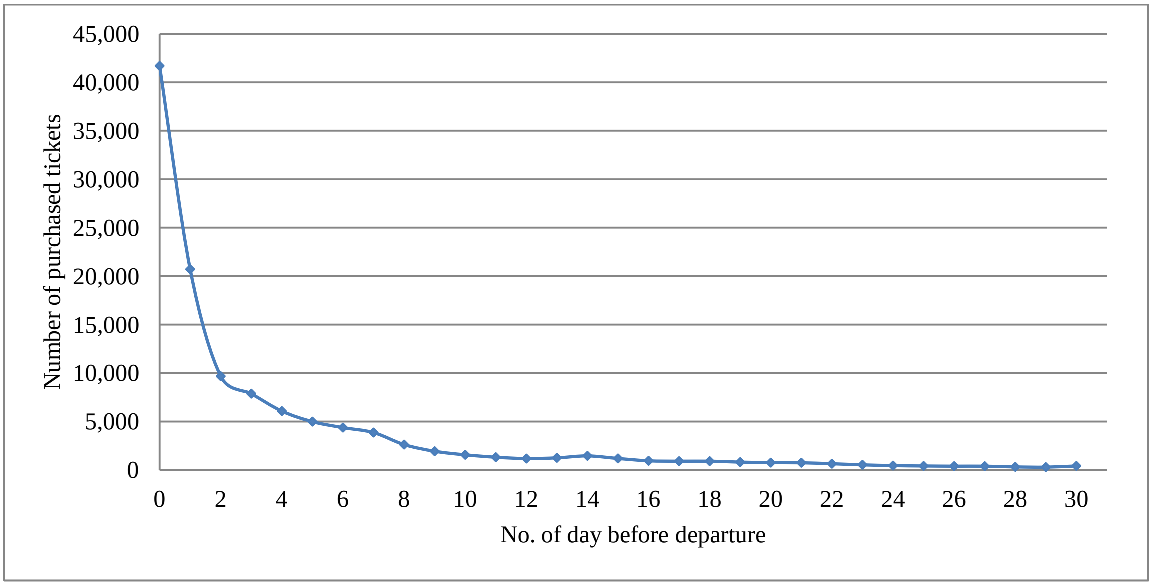


Figure 4.3 Number of purchased tickets prior to departure

Figure 4.3 shows number of purchased tickets by number of day before departure. It indicates that about 98% of the passengers purchased the tickets no earlier than 30 days before departure. The majority of the passengers (82.56%) purchased the ticket no earlier than one week before departure and a significant number of passengers (60.01%) purchased tickets no earlier than 2 days before departure.

[1] Departure time is grouped using time of day period of the intercity trip (Jin, 2007).

4.1.3 Fare Distribution

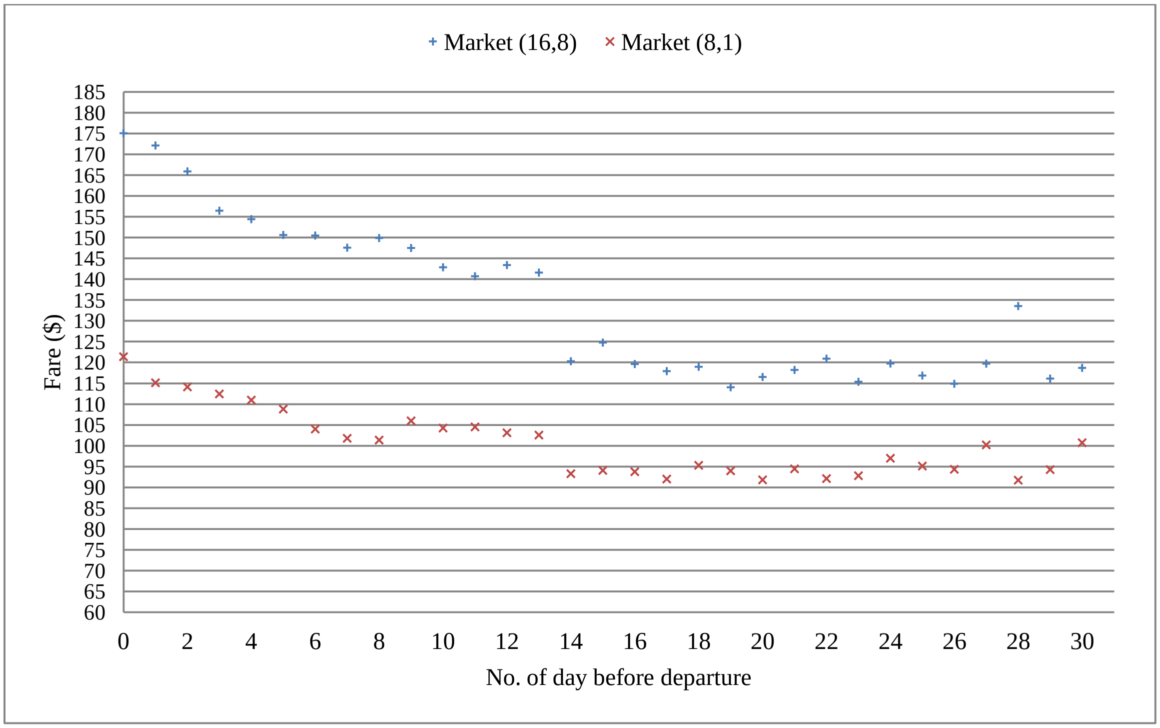


Figure 4.4 Average fare by booking time of major markets

Figure 4.4 represents fare distribution by number of day before departure in major markets (Station 16 to Station 8, and Station 8 to Station 1). These two markets account for more than one third of the passenger demand in the northbound direction. It shows that fares primarily increase as time approaches departure. The same fare pattern is also observed in other markets.

4.2 Analysis of Ticket Exchange and Cancellation

The data set used for the analysis has been extracted from intercity railway ticket reservation records registered in March 2009. We focus on coach class[2] passengers traveling in the north bound direction. This data set contains 155,175 individual

[2] The coach class of this railway service accounts for 92% of the total demand.

transactions expressed in terms of ticket purchase, cancellation, and exchange over time prior to departure. Ticket exchange decision is defined as the exchange of the original ticket for a new one and the payment of an additional cost depending on the operator's exchange policy. In our case study, passengers are not charged with exchange fee, but have to pay the difference between the new and the old ticket fare. In the case of ticket exchange, passenger either obtains a new ticket right away or after several time periods (repurchase). Ticket cancellation is defined as the final cancellation of the ticket with the passenger obtaining ticket refund depending on the operator's refund policy.

Table 4.1 shows the descriptive statistics derived from the data set in use. Ticket exchange and cancellation account for 18.22% and 29.75% of the sample respectively.

Table 4.1 Overview of ticket cancellation and exchange statistics

Ticket exchange	No. of reservations	Percent of exchange	Percent of total
1. Total exchange	28,280	100.00%	18.22%
1.1 Number of exchange			
Exchange (one time)	22,857	80.82%	14.73%
Exchange (one or two times)	27,088	95.79%	17.46%
Exchange (more than 2 times)	1,193	4.22%	0.77%
1.2 Type of exchange			
Change OD (a)	4,773	16.88%	3.08%
No change (either OD or departure)	7,001	24.76%	4.51%
Reschedule departure day (b)	1,406	4.97%	0.91%
Reschedule departure time	13,565	47.97%	8.74%
Reschedule departure day and time (c)	1,539	5.44%	0.99%
Ticket Cancellation	**No. of reservations**	**Percent of cancel**	**Percent of total**
2. Total final cancellation	46,158	100.00%	29.75%
2.1 Final cancellation after exchanged	3,506	7.60%	2.26%
Total (Northbound, March 2009, Coach Class)	155,175		100.00%
Effective Sample (Total - (a) - (b) - (c))	147,457		95.03%

Single exchange and no more than two exchanges account for 80.82% and 95.79% of the exchange ticket respectively (14.73% and 17.46% of the sample) with the illustration shown in Figure 4.5.

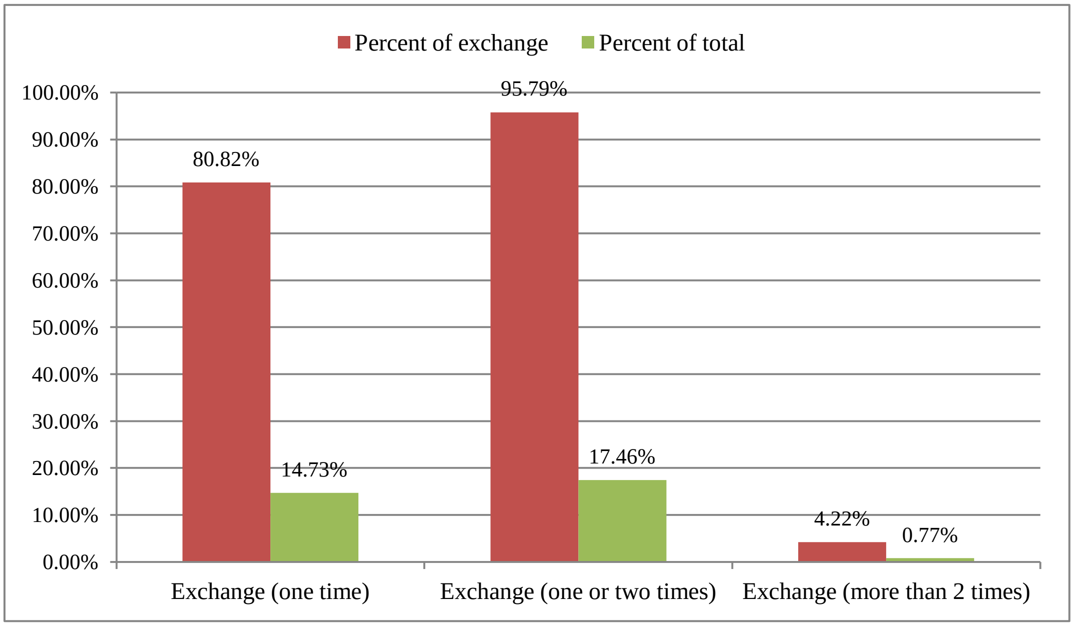


Figure 4.5 Number of exchange

We observe that only 2.26% of the sample make an exchange prior to ticket cancellation; thus for the model in Chapter 8, we assume that passenger make ticket adjustment no more than once (either exchange or cancel). Based on this assumption, data are constructed to model the first exchange decision in case of multiple exchange, and model final cancellation in case passenger both exchange and cancel. We do not consider passengers who change origin-destination or reschedule departure day because the share of these population is relatively low accounting for 3.08% and 1.90% (0.91% + 0.99%) of the sample respectively with the illustration shown in Figure 4.6.

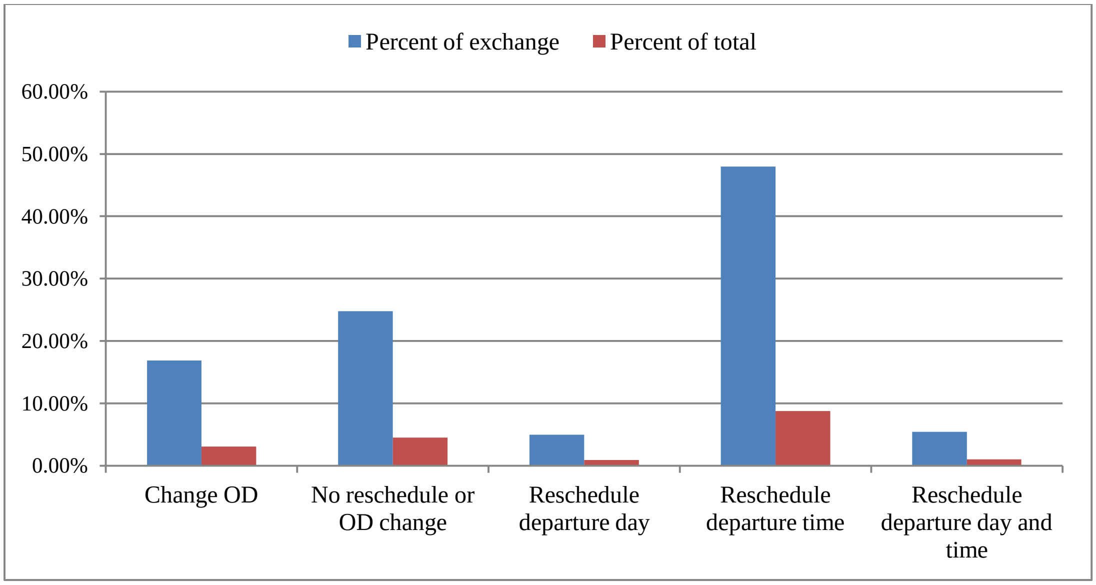


Figure 4.6 Type of exchange

Consideration of changes in origin-destination and departure day decisions requires the definition of a choice set that is significantly different across passengers and no information is available to construct a realistic choice set for each passenger. This results in the focused sample population to be composed of entire sample (155,175) subtracted by passengers with origin-destination change and departure day change (a, b, and c in Table 4.1) which results in 147,457 individual ticket reservation records of the sample.

4.3 Conclusions

This chapter provides statistical overview of data focusing on the aspects related to the model development in the following chapters. The analysis of passenger demand provides an insight on distribution of demand, advance booking, and fare which are useful for setting up the structure and constructing the choice set of the models in Chapter 5. The analysis also complements the specification of the demand functions in Chapter 6.

The analysis of ticket cancellation and exchange provides an insight for setting up the structure, and constructing the choice set for the model in Chapter 8.

Chapter 5 : Choice Models Accounting for Taste Heterogeneity

Most of the empirical studies in railway revenue management (RM) reported so far do not account for heterogeneity across different categories of travelers. Passenger taste heterogeneity is considered a major characteristic of the railway industry since passenger preferences generally vary by distance (short haul or long haul) and by time of day. Given that RM relies on the premise that different customers are willing to pay different amounts for a product, accounting for passenger heterogeneity is expected to provide high yield toward RM strategy. More specifically, Garrow (2010) suggests that calibrating models by segments to distinguish between time-sensitive and price-sensitive customers can highly impact demand prediction accuracy and contribute to significant RM system performance.

This chapter presents an application of advanced econometric techniques by exploiting railway ticket reservation data to develop choice model accounting for taste heterogeneity across different categories of passengers (Hetrakul and Cirillo, forthcoming). To this scope, based on the data analysis in the previous chapter, the methodologies in Chapter 3 based on multinomial logit, latent class, and mixed logit models are applied in this chapter to model ticket purchase timing decisions in three market segments. In Section 5.1, selected sample used for model estimation is described. Section 5.2 to Section 5.4 are dedicated to the choice set generation, and estimation results. These sections investigate heterogeneous characteristics of passenger behavior and quantify the impact of fare, advanced booking, and departure schedule toward purchase timing decision. To evaluate models' prediction capability, their validation is presented in Section 5.5 and their performance is compared in Section 5.6. Later, in Chapter 7, these choice models will be incorporated in the RM optimization.

5.1 Sample Selection

Ticket reservation data of US intercity passenger railway in 2009 are used in this analysis. The analysis focuses on coach[3] class passengers traveling northbound with the confirmed ticket contributing to actual revenue. This results in the total of 110,828 observations.

This railway service consists of 16 stations total, resulting in 119 origin-destination pairs for the north bound trip. However, some origin-destination pairs have relatively small sample sizes which are insufficient for the model estimation. To alleviate this problem and also reduce the number of models, the 16 stations are aggregated into 4 station groups. The estimated models corresponded to each station groups are shown in Table 5.1.

[3] The coach class of this railway service accounts for 92% of the total demand.

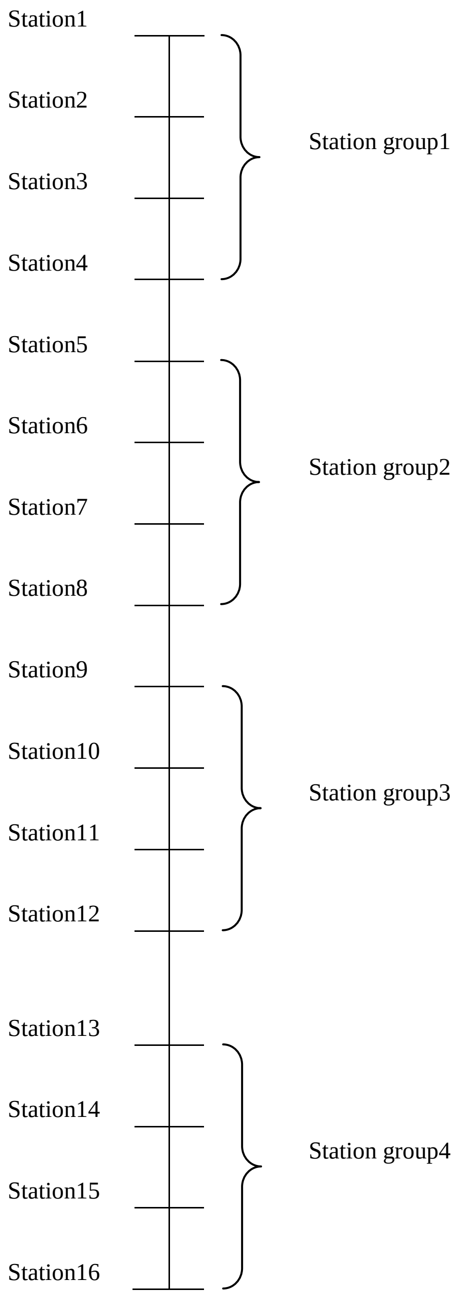


Figure 5.1 Aggregate station groups

Table 5.1 Choice model number corresponded to station group

Origin/Destination	Station group1	Station group2	Station group3	Station group4
Station group1	Model 7	-	-	-
Station group2	Model 6	Model 8	-	-
Station group3	Model 4	Model 5	Model 9	-
Station group4	Model 1	Model 2	Model 3	Model 10

All the models in Table 5.1 are estimated (see appendix A), however due to space limitation, only 3 models for representative markets are focused in this chapter as follows:

1. Long Distance: trip from station group 4 to station group 2 (Model 2)
2. Medium Distance: trip from station group 2 to station group 1 (Model 6)
3. Short Distance: trip within station group 4 (Model 10)

5.2 Choice Set Generation

The fare of this railway service varies depending on departure day of week, departure time, how early the reservation is made in advance, and passenger demand for each departure. Different passenger groups are also subjected to different discount policy such as seniors, children, military, and group travel. Based on this fare variation over the sale horizon, passengers are assumed to make the choice of when to purchase the ticket. Data analysis indicates that 98 percents of the tickets were purchased no earlier than 30 days before departure. Thus, it is assumed that passenger makes the purchase timing decision on the choice set of 31 alternatives, from 30 days before departure (booking day 1) to departure day (booking day 31). In reality, although it is possible that the choice set can vary across individuals depending on when the travel decision is made. However, this approach captures the maximum number of choice alternatives considered by the passenger, which results in passengers who actually consider a smaller choice set (less than 30days before departure) to have significantly lower probability of purchase on the alternative not being considered. This approach has been widely used in discrete choice analysis especially for the mode choice models, in the circumstance that decision maker does not necessarily consider all the possible choices when making the decision (e.g. transit not in the vicinity area being included in the mode choice).

Based on the data set, for each individual, we can only observe fare on the purchase day but not on other days in the sale horizon. To accommodate choice modeling, fares on other days in the sale horizon have been approximated from the actual data by averaging over the observed fares within the same booking day over the month. Different choice models have been estimated for the three market segments identified: (1) multinomial logit (MNL), (2) mixed logit (ML), (3) latent class (LC), and (4) multinomial logit (MNL) with socioeconomic information.

5.3 Estimation Results and Model Fit Comparison

5.3.1 Estimation Results

The estimation results derived from multinomial logit, mixed logit, and latent class model across three market segments are reported in Tables 5.2 to 5.4. The multinomial and mixed logit models are estimated with AMLET (Another Mixed Logit Estimation Tool) (Bastin, 2011). The latent class model is estimated with Latent Gold Choice 4.5, a software package by Statistical Innovations specifically designed for latent class choice modeling (Vermunt and Magdison, 2005).

For the model fit comparison, we use adjusted rho-square to account for penalty in including variables which are statistically insignificant. The approach encourages balanced specifications by trading off the improvement in the log likelihood function against the inclusion of additional variables. Given that our model specification does not include alternative-specific constants, it is appropriate to measure the goodness of fit of a model with respect to zero; $LL(0)$. We have chosen the adjusted rho-square ($\bar{\rho}_0^2$) provided by Koppelman and Bhat (2006) for the model fit comparison given by:

$$\bar{\rho}_0^2 = \frac{LL(\beta) - K - LL(0)}{LL(*) - LL(0)} = 1 - \frac{LL(\beta) - K}{LL(0)} \tag{5.1}$$

where K is the number of parameters used in the model.

Table 5.2 Long distance choice model result

MNL				ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.184	42.872	*	-0.370	141.708	*
price.period1	-0.006	3.634	*	-5.318	77.001	*
price.period2	-0.012	9.173	*	(2.830)	(60.153)	*
price.period3	-0.011	10.783	*			
price.period4	-0.010	10.380	*			
price.period5	-0.005	5.287	*			
price.period6	-0.003	3.366	*			
wknd.period1	1.240	25.231	*	1.897	20.459	*
wknd.period2	1.099	27.923	*	0.681	11.363	*
wknd.period3	0.407	13.106	*	0.347	8.765	*
wknd.period4	0.466	14.830	*	0.041	0.977	
wknd.period5	-0.472	13.477	*	-0.483	10.257	*
wknd.period6	0.260	9.369	*	0.516	11.436	*
No. of observations		37,373			37,373	
Rho-squared:		0.2970			0.3397	
Adjusted rho-squared:		**0.2969**			**0.3396**	
Log-likelihood at optimal		-90,226			-84,742	
Log-likelihood at zero		-128,338			-128,338	
LL at constant		-90,487			-90,487	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.139	-21.091	*	-0.899	-18.162	*
price.period1	0.000	0.111		0.014	1.560	
price.period2	-0.003	-1.684		-0.082	-2.069	*
price.period3	-0.005	-2.778	*	-0.049	-5.557	*
price.period4	-0.001	-0.737		-0.069	-7.931	*
price.period5	0.003	1.789		-0.077	-8.052	*
price.period6	-0.010	-4.348	*	-0.059	-6.941	*
Class Model	**Class1**			**Class2**		
Class Size	0.619			0.381		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.181	4.845	*	-0.181	-4.845	*
Monday	-0.402	-14.288	*	0.402	14.288	*
Tuesday	-0.338	-11.586	*	0.338	11.586	*
Wednesday	-0.375	-11.908	*	0.375	11.908	*
Thursday	-0.286	-9.502	*	0.286	9.502	*
Friday	-0.213	-7.690	*	0.213	7.690	*
Saturday	-0.019	-0.433		0.019	0.433	
Early morning	1.085	19.204	*	-1.085	-19.204	*
AM peak	0.985	22.737	*	-0.985	-22.737	*
AM off peak	0.474	15.353	*	-0.474	-15.353	*
PM off peak	0.096	3.491	*	-0.096	-3.491	*
PM peak	0.113	4.036	*	-0.113	-4.036	*
No. of observations					37,373	
Rho-squared:					0.3034	
Adjusted rho-squared:					**0.3032**	
Log-likelihood at optimal					-89,402	
Log-likelihood at zero					-128,338	
LL at constant					-90,487	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.248	43.103	*
price.adult	-0.024	10.139	*
price.child (2-15)	-0.004	0.279	
price.senior (62+)	-0.023	9.401	*
price.unacc child (8-11)	0.001	0.072	
price.student advantage	-0.010	2.838	*
price.adultAAA member	-0.010	3.720	*
price.childAAA member	-	-	-
price.military adult	-0.028	7.638	*
price.disabled adult	-0.024	4.788	*
price.others	-0.043	17.059	*
price.period6	-0.019	8.249	*
wknd.period1	2.012	40.426	*
wknd.period2	0.878	34.139	*
wknd.period3	-0.044	1.492	
wknd.period4	0.050	1.751	
wknd.period5	-0.133	4.703	*
wknd.period6	0.236	7.935	*
No. of observations		37,373	
Rho-squared:		0.2904	
Adjusted rho-squared:		**0.2903**	
Log-likelihood at optimal		-91,070	
Log-likelihood at zero		-128,338	
LL at constant		-90,487	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table 5.3 Medium distance choice model result

Variable	MNL Est	MNL T-Stat		ML Est	ML T-Stat	
advbk	-0.160	30.287	*	-0.392	100.986	*
price.period1	-0.018	6.775	*	-5.421	35.523	*
price.period2	-0.024	9.831	*	(4.252)	(29.595)	*
price.period3	-0.021	9.799	*			
price.period4	-0.018	8.377	*			
price.period5	-0.012	5.593	*			
price.period6	-0.008	4.074	*			
wknd.period1	1.163	21.812	*	4.299	57.340	*
wknd.period2	1.203	29.228	*	1.747	43.787	*
wknd.period3	0.751	19.745	*	0.168	4.589	*
wknd.period4	0.537	17.042	*	-0.664	20.754	*
wknd.period5	-0.149	3.310	*	-1.187	27.385	*
wknd.period6	-0.504	13.923	*	-1.353	30.550	*
No. of observations		29,514			29,514	
Rho-squared:		0.2779			0.3500	
Adjusted rho-squared:		**0.2778**			**0.3499**	
Log-likelihood at optimal		-73,182			-65,877	
Log-likelihood at zero		-101,351			-101,351	
LL at constant		-73,807			-73,807	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.084	-10.870	*	-0.672	-16.043	*
price.period1	0.017	3.864	*	-0.014	-0.917	
price.period2	0.018	3.851	*	-0.130	-2.310	*
price.period3	0.018	3.826	*	-0.073	-4.953	*
price.period4	0.022	4.696	*	-0.087	-5.805	*
price.period5	0.029	6.196	*	-0.093	-6.019	*
price.period6	0.010	1.792		-0.077	-5.371	*
Class Model	**Class1**			**Class2**		
Class Size	0.566			0.434		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.952	8.602	*	-0.952	-8.602	*
Monday	-1.212	-13.484	*	1.212	13.484	*
Tuesday	-1.061	-12.027	*	1.061	12.027	*
Wednesday	-1.161	-12.777	*	1.161	12.777	*
Thursday	-1.196	-13.013	*	1.196	13.013	*
Friday	-0.859	-11.172	*	0.859	11.172	*
Saturday	-0.837	-9.565	*	0.837	9.565	*
Early morning	0.977	11.940	*	-0.977	-11.940	*
AM peak	0.852	14.129	*	-0.852	-14.129	*
AM off peak	0.428	8.998	*	-0.428	-8.998	*
PM off peak	0.109	2.949	*	-0.109	-2.949	*
PM peak	-0.076	-2.027	*	0.076	2.027	*
No. of observations					29,514	
Rho-squared:					0.2816	
Adjusted rho-squared:					**0.2814**	
Log-likelihood at optimal					-72,809	
Log-likelihood at zero					-101,351	
LL at constant					-73,807	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.293	69.420	*
price.adult	-0.064	27.225	*
price.child (2-15)	-0.071	5.034	*
price.senior (62+)	-0.063	24.982	*
price.unacc child (8-11)	-0.063	10.836	*
price.student advantage	-0.054	19.977	*
price.adultAAA member	-0.057	23.248	*
price.childAAA member	-	-	-
price.military adult	0.396	0.786	
price.disabled adult	-0.063	7.736	*
price.others	-0.088	34.916	*
price.period6	-0.058	24.431	*
wknd.period1	1.582	30.329	*
wknd.period2	0.855	23.879	*
wknd.period3	0.389	12.255	*
wknd.period4	0.398	12.808	*
wknd.period5	0.296	6.590	*
wknd.period6	-0.520	11.507	*
No. of observations		29,514	
Rho-squared:		0.2760	
Adjusted rho-squared:		**0.2758**	
Log-likelihood at optimal		-73,383	
Log-likelihood at zero		-101,351	
LL at constant		-73,807	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table 5.4 Short distance choice model result

	MNL			ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.534	20.090	*	-0.602	58.485	*
price.period1	0.056	7.863	*	-6.945	25.872	*
price.period2	0.011	2.760	*	(3.371)	(20.846)	*
price.period3	-0.007	2.626	*			
price.period4	-0.019	11.389	*			
price.period5	-0.016	12.133	*			
price.period6	-0.008	8.293	*			
wknd.period1	2.701	9.230	*	3.390	6.153	*
wknd.period2	1.560	7.202	*	1.483	3.920	*
wknd.period3	-0.268	1.183		-0.255	0.875	
wknd.period4	-0.639	4.205	*	-1.129	4.460	*
wknd.period5	-0.758	4.708	*	-1.108	3.839	*
wknd.period6	0.404	4.267	*	0.618	2.516	*
No. of observations		4,454			4,454	
Rho-squared:		0.5637			0.5493	
Adjusted rho-squared:		**0.5628**			**0.5487**	
Log-likelihood at optimal		-6,674			-6,894	
Log-likelihood at zero		-15,295			-15,295	
LL at constant		-6,478			-6,478	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-1.164	-6.080	*	-0.327	-10.734	*
price.period1	-0.378	-1.451		-0.066	-2.898	*
price.period2	-0.518	-1.953		-0.083	-3.591	*
price.period3	-0.448	-2.712	*	-0.095	-3.954	*
price.period4	-0.606	-3.504	*	-0.089	-3.597	*
price.period5	-0.595	-3.399	*	-0.081	-3.202	*
price.period6	-0.426	-2.788	*	-0.090	-4.031	*
Class Model	**Class1**			**Class2**		
Class Size	0.595			0.406		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.462	4.444	*	-0.462	-4.444	*
Monday	0.135	1.595		-0.135	-1.595	
Tuesday	0.031	0.379		-0.031	-0.379	
Wednesday	0.118	1.432		-0.118	-1.432	
Thursday	-0.062	-0.760		0.062	0.760	
Friday	-0.037	-0.440		0.037	0.440	
Saturday	0.079	0.585		-0.079	-0.585	
Early morning	-1.117	-5.476	*	1.117	5.476	*
AM peak	-0.922	-9.202	*	0.922	9.202	*
AM off peak	-0.297	-3.406	*	0.297	3.406	*
PM off peak	-0.083	-1.007		0.083	1.007	
PM peak	-0.287	-3.548	*	0.287	3.548	*
No. of observations					4,454	
Rho-squared:					0.5845	
Adjusted rho-squared:					**0.5828**	
Log-likelihood at optimal					-6,356	
Log-likelihood at zero					-15,295	
LL at constant					-6,478	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.674	32.119	*
price.adult	-0.215	22.716	*
price.child (2-15)	-	-	-
price.senior (62+)	-0.209	17.927	*
price.unacc child (8-11)	-	-	-
price.student advantage	-0.099	5.940	*
price.adultAAA member	-0.158	11.881	*
price.childAAA member	-	-	-
price.military adult	-1.084	18.308	*
price.disabled adult	-0.227	6.498	*
price.others	-0.238	19.380	*
price.period6	-0.204	21.406	*
wknd.period1	5.381	18.684	*
wknd.period2	1.900	9.105	*
wknd.period3	-0.674	3.033	*
wknd.period4	-1.646	11.372	*
wknd.period5	-1.577	8.975	*
wknd.period6	-0.384	3.444	*
No. of observations		4,454	
Rho-squared:		0.5611	
Adjusted rho-squared:		**0.5599**	
Log-likelihood at optimal		-6,713	
Log-likelihood at zero		-15,295	
LL at constant		-6,478	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

5.3.2 Statistical Test to Compare Non-nested Models

Since our models cannot be written as a restricted version of one another, the model fit can be statistically compared using the non-nested hypothesis test proposed by Horowitz (1982). The null hypothesis associated with the non-nested hypothesis is:

$$H_0\text{: Model1 } (\bar{\rho}_H^2) = \text{Model2}(\bar{\rho}_L^2) \tag{5.2}$$

The decision rule used to express significance of the test is:

$$\text{Reject } H_0 \text{ } if \text{ } \Phi\left[-(-2(\bar{\rho}_H^2 - \bar{\rho}_L^2) \times LL(0) + (K_H - K_L))^{\frac{1}{2}}\right] < \alpha \tag{5.3}$$

where:

$\bar{\rho}_H^2$ is the larger adjusted rho-square value,

$\bar{\rho}_L^2$ is the smaller adjusted rho-square value,

K_H is the number of parameters in the model with the larger adjusted rho-square,

K_L is the number of parameters in the model with the smaller adjusted rho-square,

Φ is the standard normal cumulative distribution function,

$LL(0)$ is the log likelihood at zero (associated with equally likely model),

α is the significance level.

The summarized measure of fit for the models are briefly shown in Table 5.5, and the parameters used for the non-nested model hypothesis testing is shown in Table 5.6.

Table 5.5 Results comparison

Long Distance	**MNL**	**ML**	**LC**	**MNL-Socio**
LL(0)	-128,338	-128,338	-128,338	-128,338
LL at optimal	-90,226	-84,742	-89,402	-91,070
Adjusted Rho-square	0.2969	**0.3396**	0.3032	0.2903
No. of estimated parameters	13	9	26	17
Medium Distance	**MNL**	**ML**	**LC**	**MNL-Socio**
LL(0)	-101,351	-101,351	-101,351	-101,351
LL at optimal	-73,182	-65,877	-72,809	-73,383
Adjusted Rho-square	0.2778	**0.3499**	0.2814	0.2758
No. of estimated parameters	13	9	26	17
Short Distance	**MNL**	**ML**	**LC**	**MNL-Socio**
LL(0)	-15,295	-15,295	-15,295	-15,295
LL at optimal	-6,674	-6,894	-6,356	-6,713
Adjusted Rho-square	0.5628	0.5487	**0.5828**	0.5599
No. of estimated parameters	13	9	26	15

Table 5.6 Hypothesis testing parameters

Market	**H-L***	$\bar{\rho}_H^2$	$\bar{\rho}_L^2$	K_H	K_L	$LL(0)$	x	$\Phi(x)$
Long	ML-MNL	0.3396	0.2969	9	13	-128,338	-104.75	0.0000 E+00
	LC-MNL	0.3032	0.2969	26	13	-128,338	-40.44	0.0000 E+00
	ML-LC	0.3396	0.3032	9	26	-128,338	-96.63	0.0000 E+00
	MNL-MNL-socio	0.2969	0.2903	13	17	-128,338	-41.15	0.0000 E+00
Medium	ML-MNL	0.3499	0.2778	9	13	-101,351	-120.89	0.0000 E+00
	LC-MNL	0.2814	0.2778	26	13	-101,351	-27.07	1.1107 E-161
	ML-LC	0.3499	0.2814	9	26	-101,351	-117.82	0.0000 E+00
	MNL-MNL-socio	0.2778	0.2758	13	17	-101,351	-20.21	3.9974 E-91
Short	MNL-ML	0.5628	0.5487	13	9	-15,295	-20.88	4.9783 E-96
	LC-MNL	0.5828	0.5628	26	13	-15,295	-24.97	6.4759 E-138
	LC-ML	0.5828	0.5487	26	9	-15,295	-32.55	1.0217 E-231
	MNL-MNL-socio	0.5628	0.5599	13	15	-15,295	-9.30	7.0223 E-21

*H= Model with the larger adjusted rho-square, L= Model with the smaller adjusted rho-square.

5.4 Result Discussion

5.4.1 Multinomial Logit (MNL) Model

A. MNL Specification

The independent variables that enter the final models are advance booking (number of day before departure), fare ($), and weekend departure dummies. Fare and advance booking variables are aimed to capture passenger tradeoff behavior between early booking with cheaper fare and late booking with higher fare. The model specification allows for price coefficient to take different values across booking periods to accommodate the assumption that passengers have different price sensitivities over the sale horizon. The booking periods are grouped such that booking days within the same booking period have approximately the same number of reservations. These six booking periods 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 weekend dummy variables assume different values across booking periods to capture the effect of departure day of week toward ticket purchase timing. Based on this specification, the utility of passenger i booking the ticket on day j which falls within the booking period k can be expressed as:

$$U_i(j,k) = \left(\beta_{advbk} \times advbk_j\right) + \left(\beta_{fare}^{k} \times fare_j\right) + \left(\beta_{wknd}^{k} \times wknd_i\right) + \varepsilon_{ij} \qquad (5.4)$$

where the independent variables and their associated index are:

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

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

$advbk_j$ = Number of day before departure of booking day j

$fare_j$ = Fare of booking day j ($)

$wknd_i$ = Weekend dummy (1 if departure is on weekend, 0 otherwise)

ε_{ij} = A mutually independent noise term of individual i on choice j following a Gumbel distribution

B. MNL Result

Results obtained with the MNL indicate that disutility is associated with advance booking and that the lack of flexibility to change travel plan is negatively perceived. Passengers generally prefer to hold their purchase and pay for the product as late as possible. The price sensitivities in all the booking periods have negative sign and are statistically significant at the 5% significance level in all the three markets. The decreasing magnitude of price sensitivity as booking period approaches departure is in line with the expectation. Passengers are expected to be most sensitive to fare at the beginning of the sale horizon. As time approaches departure, passengers become less sensitive to fare especially on the departure day and more concerned about seats availability. The weekend dummies show the expected pattern; the value decreases as the booking period approaches departure indicating that passenger traveling on weekend generally purchase ticket earlier in advance compared to passenger traveling on weekday. It is also expected that weekend travelers are primarily leisure travelers.

5.4.2 Mixed Logit (ML) Model

A. ML Specification

The price coefficients in the MNL are found to vary across booking periods, thus the specification of mixed logit model is expected to capture variation in price sensitivities across population. For simplicity a parametric approach is used in mixed logit model specification; price is specified as a log-normally distributed random coefficient. The log-normal distribution is selected over the normal distribution because it provided a better fit and because it ensures a negative value of the parameter for the entire population. Based on this specification, the utility of passenger i booking the ticket on day j which falls within the booking period k can be expressed as:

$$U_i(j,k) = \left(\beta_{advbk} \times advbk_j\right) + \left(\exp[\ln(\tilde{\beta}_{fare})] \times fare_j\right) + \left(\beta^k_{wknd} \times wknd_i\right) + \varepsilon_{ij} \tag{5.5}$$

The index follows equation 5.4. In this specification $\ln(\tilde{\beta}_{fare})$ is distributed as normal. The price coefficient of mixed logit in Table 5.2-5.4 represents the mean (μ) and standard deviation (σ) of the log of price coefficient.

B. ML Result

The mixed logit model accounting for heterogeneity in price sensitivity with log-normal distribution[4] shows a better model fit compared to the MNL and the LC model in long and medium distance markets based on the adjusted rho-square value (shown in Table 5.5). However, in short distance model, the MNL has a better fit compared to the ML; it should be noted that these two models are not nested. This could be due to the fact that price coefficient segmented by time period reflects the observed data better than one random coefficient parameter for price specified with a log-normal distribution.

When comparing the ML and the MNL, based on the null hypothesis that that both models perform equally well, the non-nested hypothesis test indicates that the

[4] The price parameters of mixed logit in Table 5.2-5.4 are log of the price coefficients with the underlying normal distribution.

probability of the MNL performing as well as the ML in long and medium distance market are 0.0000 and 0.0000 respectively (shown in Table 5.6) which result in the rejection of null hypothesis at the 0.03% significance level (the ML outperforms the MNL). As for short distance model, the probability of the ML performing as well as the MNL is $4.9783 \times 10 - 96$ which results in the rejection of null hypothesis at the 0.03% significance level (the MNL outperforms the ML).

The ML model results show the expected sign for advance booking, fare and weekend dummies as observed in the MNL models. The estimates are all statistically significant at the 5% significance level except for the weekend dummy at booking period 4 (long distance market) and booking period 3 (short distance market).

5.4.3 Latent Class (LC) Model

A. LC Specification

In latent class model specifications, we aim at segmenting passenger behavior by trip characteristics because we believe that passengers traveling at particular periods are relatively homogeneous in their characteristics. The explanatory variables of the choice model include fare ($) and advance booking (number of day before departure). Similar to the MNL model, the price coefficient is allowed to take different values across booking periods. Based on this specification, the utility of choice j for a specific class s in the class specific choice model[5] can be written as:

$$U_i(j,k,s) = \left(\beta^{s}_{advbk} \times advbk_j\right) + \left(\beta^{k,s}_{fare} \times fare_j\right) + \varepsilon_{ij} \tag{5.6}$$

[5] Based on latent class (LC) model methodology in Chapter 3, Equation 3.14.

The index follows equation (5.4). In addition to the explanatory variables of the choice model, other elements of the booking data are extracted to segment demand and capture heterogeneity across different categories of passengers in class membership model which are:

Departure day of week: Dummy variables are used to indicate whether the trip is taken on a particular day of week; this results into six dummy variables for the class membership model, one for each day of the week (except Sunday).

Departure time of day: Dummy variables are used to indicate whether the trip is taken on a particular time of day. We use six departure times as suggested by Jin (2007) for the intercity trip which are (1) early morning (0:00 am-6:29 am), (2) a.m. peak (6:30am-8:59 am), (3) a.m. off-peak (9:00 am-11:59 am), (4) p.m. off-peak (12:00pm-15:59 pm), (5) p.m. peak (16:00 pm-18:29 pm), and (6) evening (18:30 pm-23:59 am). Five departure times of day (except evening) are used for the class membership model.

Based on this specification, the utility that customer i belongs to class s in the class membership model[6] has the form:

$$U_i(s) = C_s + \sum_{t=1}^{5}(\beta_{TOD}^{t} \times TOD_i) + \sum_{d=1}^{6}\left(\beta_{DOW}^{d} \times DOW_i\right) + \varepsilon_{is} \qquad (5.7)$$

where C_s is class specific constant.

B. LC Result

In this context, the main advantage of the latent class specification over multinomial logit model is the ability to identify distinct group of passengers' behavior on ticket purchase

[6] Based on latent class (LC) model methodology in Chapter 3, Equation 3.15.

timing. We assume two latent classes of passenger, business and leisure trip which are segmented by the class membership model. Passengers are assumed to have different preferences about trip schedule which results in different willingness to pay (WTPs) for delaying ticket purchase across classes.

When comparing the LC and the MNL, the non-nested hypothesis test indicated that the probability of the MNL performing as well as the LC in long, medium, and short distance markets are 0.0000, $1.1107 \times 10 - 161$, and $6.4759 \times 10 - 138$ respectively which result in the rejection of null hypothesis at the 0.03% significance level (the LC outperforms the MNL). On the other hand, when we compare between the LC and the ML, the non-nested hypothesis test indicated that the probability of the LC performing as well as the ML in long and medium distance market are 0.0000 and 0.0000 respectively which result in the rejection of null hypothesis at the 0.03% significance level (the ML outperforms the LC). This concludes that the ML has the best statistical fit in long and medium distance markets. As for short distance model, the probability of the ML performing as well as the LC is $1.0217 \times 10 - 231$ which results in the rejection of null hypothesis at the 0.03% significance level (the LC outperforms the ML). This concludes that the LC has the best statistical fit in short distance market.

In this setting, the variables used to specify class membership model resulted in a set of 42 underlying scenarios of booking, also called covariate pattern. For each scenario, the likelihood that a booking belongs to a latent class can be calculated as a binary logit probability corresponded to class membership parameters. Figure 5.2 to

Figure 5.4 represent the probability of passenger belonging to class 1 for the 42 different covariate patterns.

The results obtained with the latent class model are coherent with those given by the multinomial logit and mixed logit models except that in the medium distance market, passengers in class 1 have positive price coefficient meaning that they are insensitive to price. Given that the magnitude of the coefficients within the same model could not be compared between different classes due to scale parameter (Carrier, 2008), the ratio of advance booking coefficient to the price coefficient is calculated to represent the willingness to pay (WTP) to delay the ticket purchase for one day and shown in Table 5.7. We report the minimum, maximum, and the average WTPs obtained from six price coefficients of each passenger class across three markets. We assume that the leisure passengers generally know their travel plans earlier in advance while the business travelers generally book their tickets closer to departure date. Therefore, the passenger class with higher WTP is assumed to be business passenger.

Table 5.7 Willingness to pay (WTPs) for one day delay in ticket purchase

	WTP Long distance ($/day)		**WTP Medium Distance ($/day)**		**WTP Short Distance ($/day)**	
	Class1	**Class2**	**Class1**	**Class2**	**Class1**	**Class2**
Minimum	13.92	11.03	*	5.19	1.92	3.44
Maximum	99.43	18.38	*	46.68	3.08	5.00
Average	**46.17**	**13.86**	*****	**6.36**	**2.42**	**3.96**

*Denotes price insensitive (price coefficient with positive sign)

For the long distance market, Table 5.7 shows higher WTP for purchase delay in class 1 (average $46.17 per day) than class 2 (average $13.86 per day) for the majority of the booking periods, indicating that passengers from class 1 are willing to pay more for the possibility to change their travel plans. Based on our assumption, passengers in class 1 are believed to be business travelers which accounts for 61.9% of this market (see Table 5.2). More specifically, the class membership model indicates that passengers departing from early morning to AM peak are predominantly class 1 passenger (see Figure 5.2).

For the medium distance market results in Table 5.7, WTP for class 1 (price insensitive) is higher than for class 2 (average $6.36 per day) for most of the booking periods considered. The results indicate that passenger in class 1 are predominantly business travelers which accounts for 56.6% of this market (see Table 5.3). More specifically, the class membership model indicates that passenger class 1 (business travelers) predominantly depart from early morning to AM off peak especially on Sunday as shown in Figure 5.3.

Based on the short distance market results in Table 5.7, WTP for purchase delay of class 1 (average $2.42 per day) is lower than for class 2 (average $3.96 per day). The result indicates that passengers in class 1 are predominantly leisure travelers which accounts for 59.5% of this market (see Table 5.4). More specifically, the class membership model indicates that passenger class 1 (leisure travelers) predominantly depart from AM off peak until evening as shown in Figure 5.4.

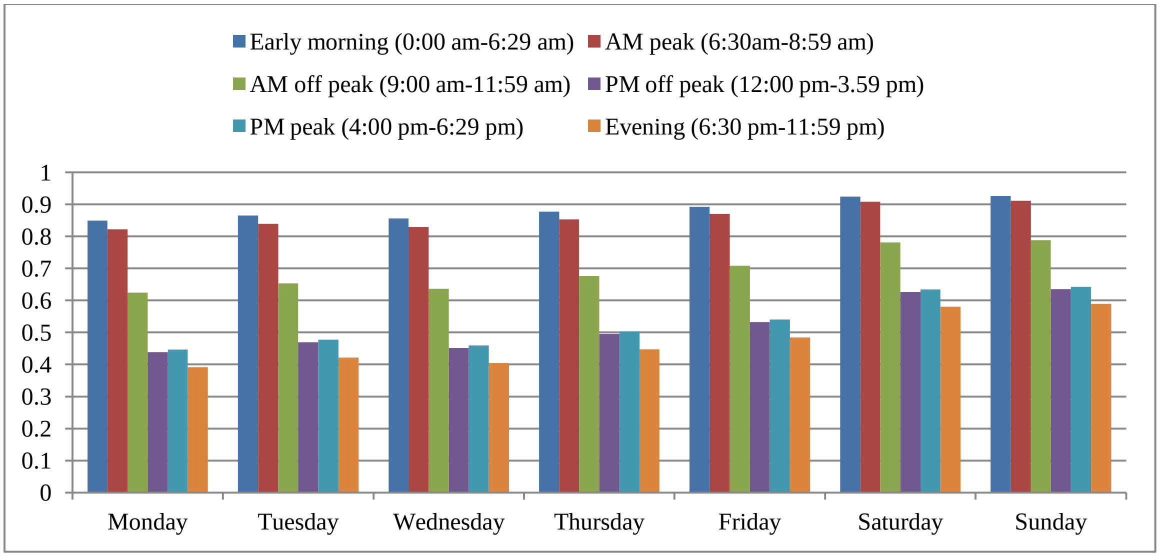


Figure 5.2 Probability of passenger belonging to class 1 in long distance market

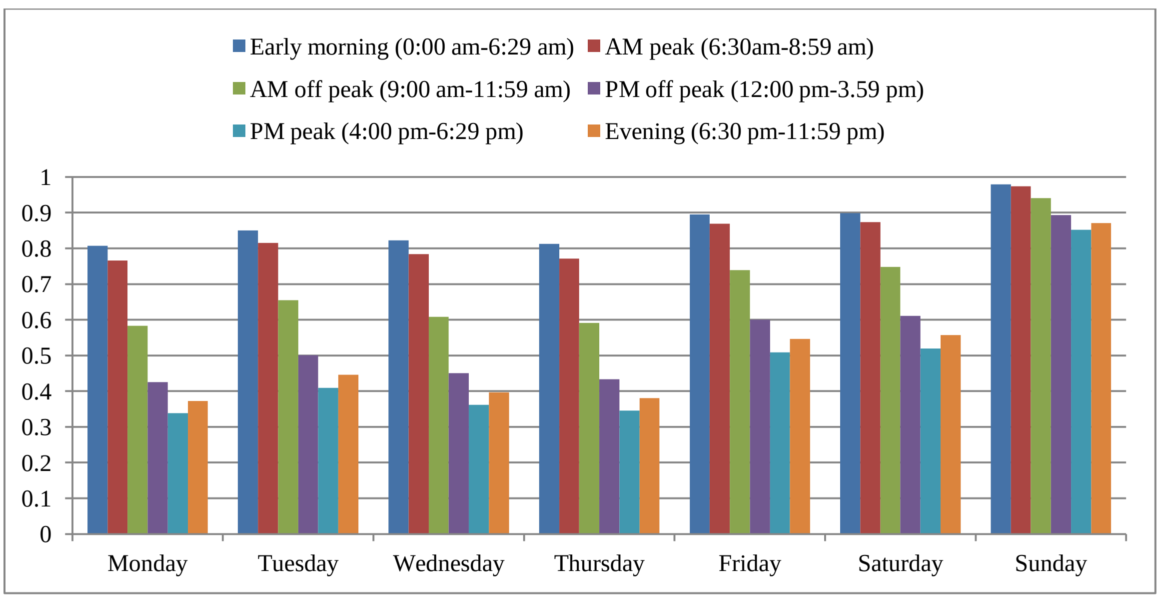


Figure 5.3 Probability of passenger belonging to class 1 in medium distance market

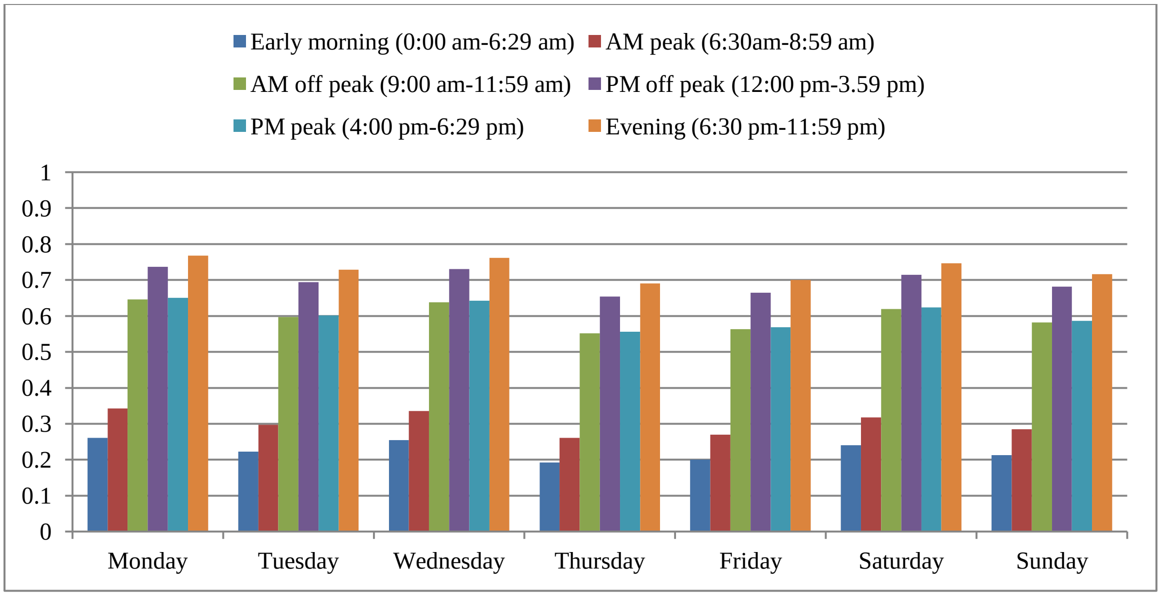


Figure 5.4 Probability of passenger belonging to class 1 in short distance market

5.4.4 MNL Model with Socioeconomic

A. MNL with Socioeconomic Specification

The first group of multinomial logit models estimated are based on the assumption that the booking data do not contain socioeconomic information. However, given that different passenger groups for this railway service are subjected to different discount rate such as seniors, children, military, and group travel, we could use ticket discount information to identify passenger types in term of socioeconomic characteristics. We have been able to identify nine passenger types from the data set. Unidentified passengers are grouped into passenger type ten.

In the model specification, we specify price coefficients differently across passenger types to accommodate the assumption that price sensitivities vary by passenger type. However, on the departure day (booking day31; period 6), given that passengers generally become insensitive to price and behave similarly across passenger types, price sensitivity of booking period 6 is specified separately and assumed to be the same across all passenger types. Based on the specification, the utility of passenger i of type p booking the ticket on day j which falls within the booking period k can be expressed as:

$$U_i(j,k,p) = (\beta_{advbk} \times advbk_j) + k_j^{j\neq 31}(\beta_{fare}^{p} \times fare_j) + k_j^{j=31}(\beta_{fare}^{k=6} \times fare_j) + (\beta_{wknd}^{k} \times wknd_i) + \varepsilon_{ij} \quad (5.8)$$

The index follows equation (5.4). k_j^c is a dummy equal to 1 when booking day j satisfies condition c and 0 otherwise.

B. MNL with Socioeconomic Result

Finally we report the results obtained from the estimation of a multinomial logit model accounting for deterministic taste heterogeneity (Table 5.2 to Table 5.4). The data set contains a limited number of observations including passenger types. This makes impossible the estimation of price sensitivity for a certain number of class and results into statistically insignificant coefficients for some of the classes considered (e.g. child and unaccompanied child for the long distance market).

When comparing the MNL with socioeconomic and the MNL, the non-nested hypothesis test indicated that the probability of the MNL with socioeconomic performing as well as the MNL in long, medium, and short distance markets are 0.0000, $3.9974 \times 10-91$, and $7.0223 \times 10-21$ respectively which results in the rejection of null hypothesis at 0.03% significance level (the MNL outperforms the MNL with socioeconomic).

For the long distance market, the results are in line with the expectations. The price sensitivity in the last booking period shows relatively low magnitude (-0.019) indicating that passengers become less sensitive to price. The unidentified passenger type is shown to be the most price sensitive (-0.043); passengers in this group are subjected to special discounts, thus it is reasonable that they are highly price sensitive. Military adult is the second most price sensitive passenger type (-0.028). The adult passenger type subjected to full fare amount shows price sensitivity equal to disabled adult (-0.024). Passengers subjected to student advantage discount and adult passenger with AAA membership have equal price sensitivity (-0.010) which is the lowest price sensitivity among all the passenger types.

For the medium distance market, the price sensitivity in the last booking period shows relatively low magnitude (-0.058) as expected. The unidentified passenger type are shown to be the most price sensitive (-0.088) among all passenger types. Child is the second most price sensitive passenger type (-0.071) followed by adult passenger (-0.064). Senior (-0.063), unaccompanied child (-0.063), and disabled adult (-0.063) are slightly less price sensitive than adult. Passengers associated with student advantage (-0.054) and adult with AAA membership (-0.057) are the least price sensitive.

For the short distance market, the estimation results show a slightly different pattern compared to long and medium distance market. The result indicates that the military group is the most price sensitive (-1.084). They are followed by unidentified passenger (-0.238), disabled adult (-0.227), adult (-0.215), senior (-0.209), and adult with AAA membership (-0.158) respectively. The passengers subjected to student advantage discount are the least price sensitive (-0.099). The weekend dummies indicate that passenger departing on weekend generally purchase their ticket ahead of time.

5.5 Model Validation

To compare the prediction capabilities of the models, we perform the out of sample validation for the long distance market. The data (37,373 observations) are divided into two groups. The first group consists of approximately 80 percents of the data (30,125 observations) containing passenger traveling from the 1st day to 25th day of the month. Departure day was chosen as a cut point because we assumed that data from the first group is obtained prior to the second group. The model is re-estimated using data of the first group and the results are applied to predict the decision of the second group which contains passengers who traveled on day 26th to 31st of the month (7,248 observations). The prediction capabilities of the four models are compared in Table 5.8. The root mean square error (RMSE) is used as the measure of error. The predictions from four models are compared in Figure 5.5 to Figure 5.8.

Table 5.8 Out of sample validation for long distance market

Choice*	Actual	MNL	MNL-socio	LC	ML
Day1	42	8	5	11	0
Day2	21	10	6	13	0
Day3	25	12	8	15	0
Day4	23	15	9	18	0
Day5	32	18	11	20	0
Day6	42	22	14	24	0
Day7	45	27	17	28	1
Day8	41	33	20	32	1
Day9	60	41	24	38	1
Day10	50	50	30	45	2
Day11	50	61	36	57	3
Day12	58	30	29	37	2
Day13	66	36	36	43	3
Day14	95	45	44	50	5
Day15	61	54	53	58	7
Day16	86	64	64	67	11
Day17	103	75	78	77	15
Day18	86	89	95	89	22
Day19	73	105	116	102	31
Day20	88	123	141	118	44
Day21	133	119	142	103	60
Day22	143	139	172	123	85
Day23	148	166	210	154	123
Day24	254	199	255	204	178
Day25	257	238	310	298	257
Day26	261	330	384	337	349
Day27	459	395	467	402	504
Day28	608	473	569	490	728
Day29	503	560	692	617	1,045
Day30	1,196	1,344	821	1,390	1,419
Day31	2,139	2,368	2,391	2,188	2,352
Total	**7,248**	**7,248**	**7,248**	**7,248**	**7,248**
RMSE		62	95	54	125

*Choice indicates booking day.

Based on Table 5.8, the LC model results in the least error with the root mean square error (RMSE) of 54 compared to RMSE of the MNL, the MNL with socioeconomic, and the ML which are 62, 95, and 125 respectively. The detailed prediction capability is elaborated in the next section.

5.6 Comparison on Model Performance

5.6.1 Model Fit

To conclude, for the long distance market, the mixed logit provides the best statistical fit with the adjusted rho-square highest among all other models. Both mixed logit and latent class models outperform multinomial logit models (with and without socioeconomics) as expected. In medium distance market, mixed logit provide the best statistical fit; it outperforms both multinomial logit specifications (with and without socioeconomics). In short distance model, latent class provides the best statistical fit while in this case the multinomial logit models (with and without socioeconomics) outperform mixed logit model. This is because in this market, the price coefficient segmented by time period and passenger type is capable of representing the observed behavior better than assuming log-normal distribution for the price coefficient. The overall results also indicate that segmenting the passengers' price sensitivity by booking period (the MNL) appears to be more appropriate than segmenting by socioeconomic (the MNL with socioeconomic) information for all the three markets.

5.6.2 Prediction Capabilities

The validation in Table 5.8 is grouped into periods and shown in Figure 5.5 to Figure 5.8. All the models under predict purchase decision on day 1 to day 9. This is because the fluctuation in passengers' decision in these days cannot be captured by the fare variable and because the pattern of observed purchases is not mono-tone, thus cannot be explained by the advance booking variable.

On day 10 and 11 the MNL and the LC models perform relatively well. From day1 to11, the MNL with socioeconomic performs the worst which could possibly be influenced by specific passenger type price sensitivity within these days (this model allows for different price sensitivity by passenger type). This confirms the aforementioned conclusion that segmenting price sensitivity by time period is more suitable than by passenger type in this empirical study. On day 12 to 20, the LC model performs the best and slightly better than the MNL, while the MNL with socioeconomic does not do well on day 11, 19, and 20 which should be influenced by the same reason given for day 1 to 11. Throughout day 1 to day 20, the prediction from the ML drastically under predicts purchase decision.

On day 21 to 28, the LC performs close to the MNL. On day 29 to 31 when the observed purchase decisions are significantly high, the LC performs almost as well as the MNL on day 29 and 30. On day 31, the LC model shows notable performance. Its predicted number of purchase is 2,188 which is very close to the observed value (2,139) where the MNL and the MNL with socioeconomic over predict the value (2,368 and 2,391 respectively). The ML model provides prediction for day 21 to day 31 relatively similar to those obtained with other models, although its performance is not very good.

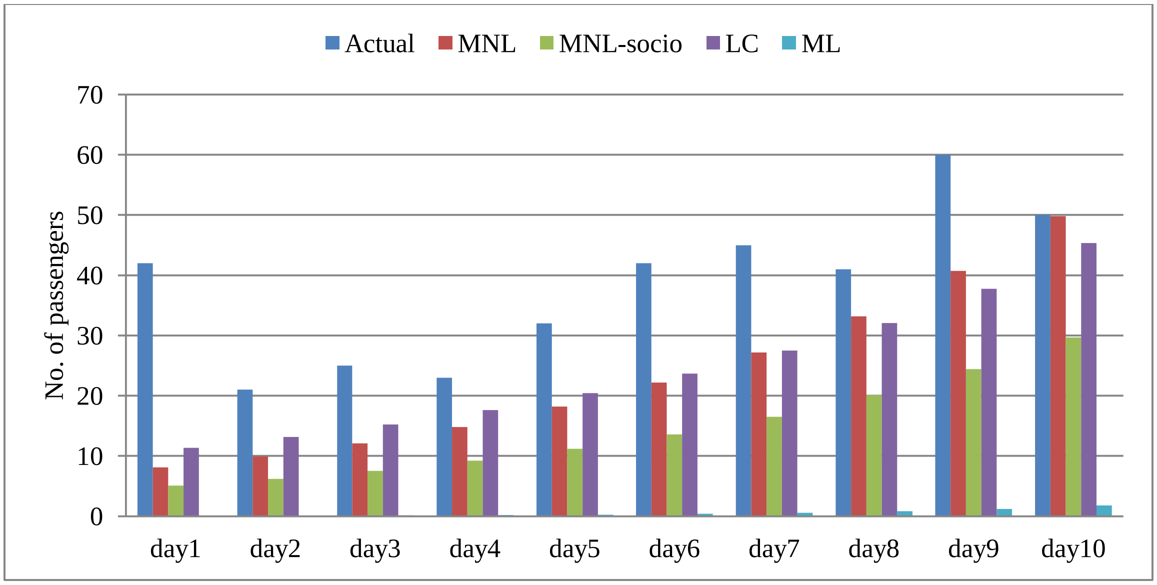


Figure 5.5 Booking day 1-10 prediction

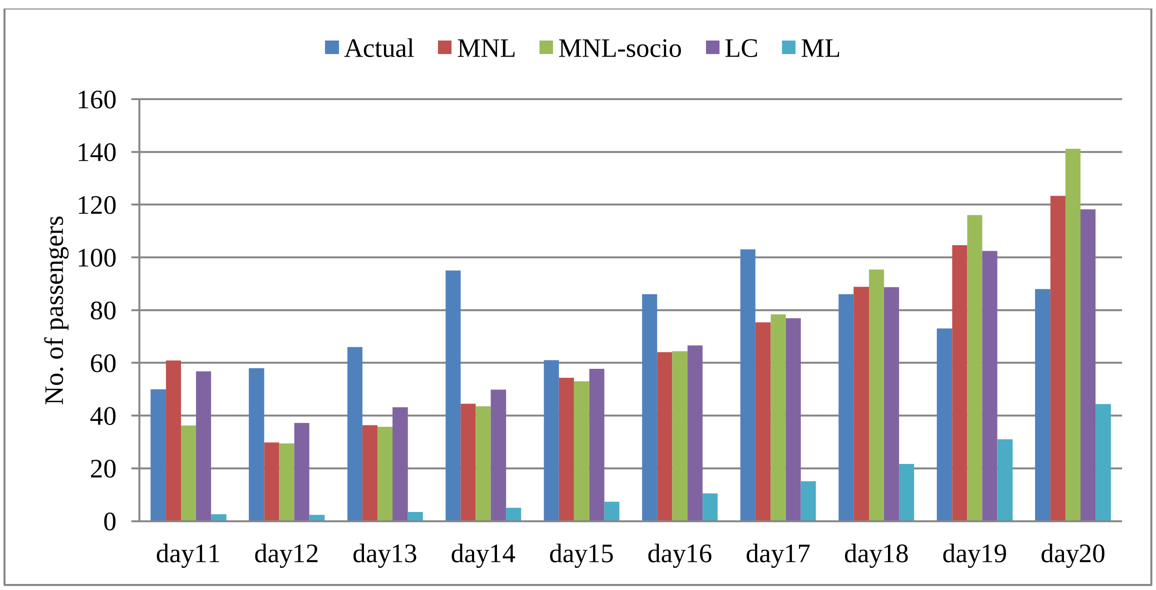


Figure 5.6 Booking day 12-20 prediction

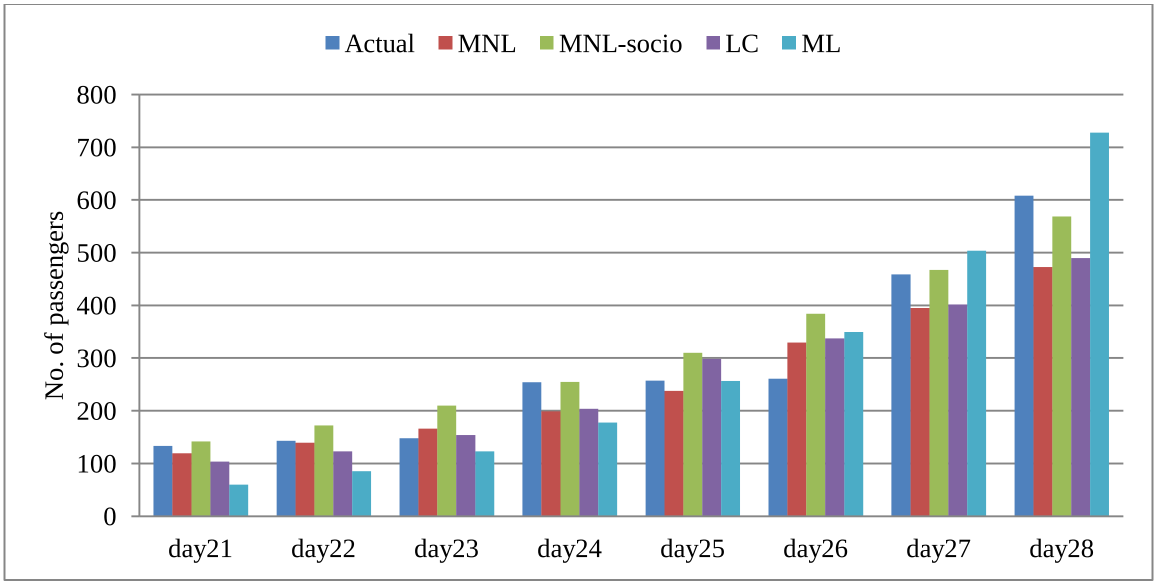


Figure 5.7 Booking day 21-28 prediction

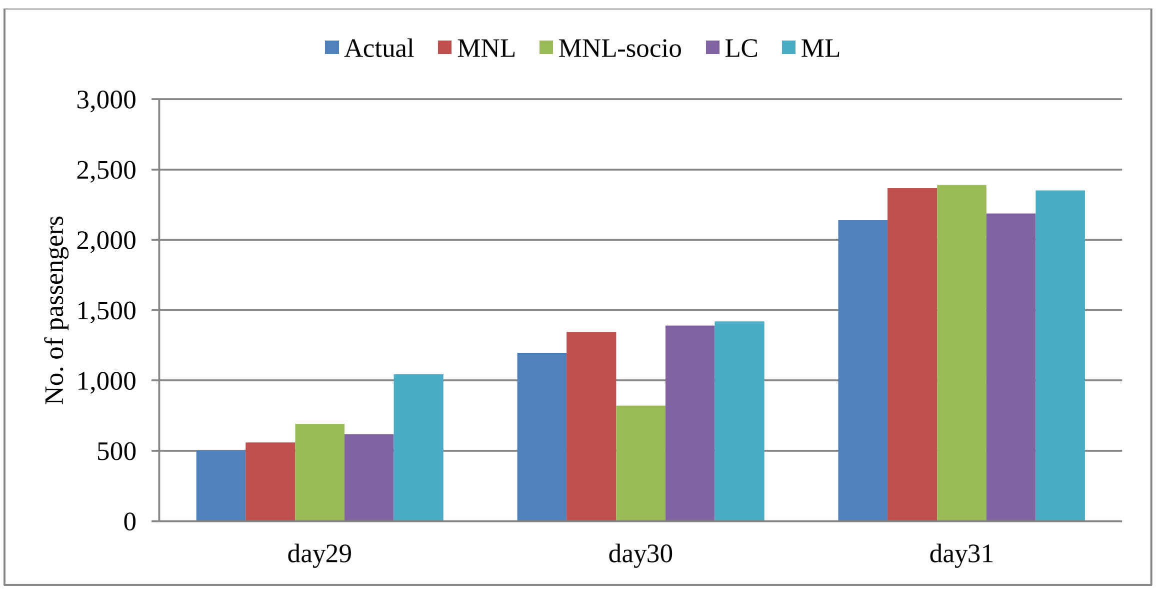


Figure 5.8 Booking day 29-31 prediction

5.7 Conclusions

In this chapter, advanced demand modeling approaches are proposed to study railway intercity passenger booking decision and to segment preferences; all the models are calibrated on ticket reservation data. Modeling formulations considered include multinomial logit, mixed logit, and latent class models; markets are segmented on trip distances: long, medium, and short. The results show that the following variables: fare, advance booking (number of day before departure), and departure day of week, can be used as determinants affecting ticket booking. Mixed logit and latent class models are then applied to account for taste heterogeneity. In term of statistical goodness of fit, the results indicate that the mixed logit model provides the best fit for the long distance and medium distance markets, while the latent class model provides the best statistical fit for the short distance market. In term of prediction capability, the out of sample validation for the long distance market indicated that the LC model provides the best prediction with least error based on root mean square error (RMSE) while the ML does not perform well in prediction. Interestingly, our analysis shows that models with better goodness of fit are not necessarily those with better prediction capabilities. Results also indicate that segmenting passengers by booking period provides better fit than segmenting passengers by socioeconomic characteristics.

Results from this study show that advanced demand models can be estimated on ticket reservation data and that market segmentation can be obtained even with limited knowledge of socio-demographic characteristics of the population.

Chapter 6 : Passenger Demand Functions

In the previous chapter, the passenger choice models are proposed. The models represent passenger response to fare by means of purchase timing decision. However, they do not account for the possibility of passenger leaving the service or induced demand. In this chapter, this assumption is relaxed by proposing demand functions which predict demand volume for each market corresponded to fare and other departure characteristics using linear and log-linear regression. This approach adopts the framework from Sibdari et al. (2008).

Although the demand functions provide similar level of aggregation to the choice models presented in the previous chapter, the estimation procedure is different. In choice models, each observation refers to an individual booking. Thus, for each booking, it is reasonable to assume dependency between choices since only one booking day can be selected. In contrast, each observation in the demand function refers to an aggregate demand volume on each booking day. Thus, demand on each booking day predicted from demand function does not necessarily influence demand of other booking days. The demand volume is then computed from the summation of predicted demand on each booking day over the sale horizon.

Two specifications of demand functions are presented in this chapter. In Section 6.1, the linear demand function is presented. The approach is applied to markets of single-origin multiple-destinations where the estimated parameters are later incorporated in the single-leg revenue optimization in Section 7.1. In Section 6.2, the log-linear demand function is presented. The approach is applied to network of selected stations where the estimated parameters are later incorporated in the network revenue optimization in Section 7.2 and Section 7.3.

6.1 Linear Demand Function

The monthly ticket reservation data from US intercity passenger railway in 2009 are used for the analysis. The analysis focuses on coach class passengers traveling from south end station (station 16) to other stations in the corridor; only reservations that were not cancelled and eventually contributed to the actual revenue are focused in this analysis. This results in the total of 44,847 reservation records.

6.1.1 Model Specification

In the linear demand function, other than the constant, the independent variables included in the final model are advance booking square, departure day of week dummies, fare, 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 results in a model with a positive price coefficient. This is because, unlike the classical demand model, in this case fare is not a completely independent variable. The revenue management team periodically changes fare in response to the demand to maximize ticket revenue.

To address this problem, the booking day is grouped into 5 periods and the demand function for each period is estimated independently. 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. With this approach, it is assumed that, in each booking period, fare 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 the fare of this railway operator changes in piecewise manner. In other words, fares 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 continuously and instantaneously.

This approach enables the analysis to compare the demand of the booking period subjected to different fare throughout the month and to obtain a service demand that is sensitive to price within the booking period. In the demand function, the passenger demand on booking day j which falls into booking period k' for the trip departure on day of week d takes the form:

$$\begin{aligned} &Demand(j, k', d) \\ &\quad = \alpha_0^{k'} + \left(\alpha_{advbksq}^{k'} \times advbk_j{}^2\right) + \sum_{d=1}^{6}\left(\alpha_{DOW}^{k'} \times DOW_d\right) \\ &\quad + \sum_{j \in k'}\left(\alpha_{bkday_j}^{k'} \times bkday_j\right) + \varepsilon \end{aligned} \tag{6.1}$$

where:

$advbk_j$	=	Advanced booking (number of day before departure)
DOW_d	=	Departure day of week dummies
$bkday_j$	=	Booking day specific dummies
ε	=	Error term

The daily passenger demand on day of week d for each origin-destination pair is the summation of passenger reservation on each booking day j over the sale horizon:

$$Demand(d) = \sum_{j=1}^{31} Demand(j, k', d) \tag{6.2}$$

Due to space limitations, demand function results are shown only for 3 representative markets in Table 6.1 to Table 6.3. Note that the predicted demands obtained from the demand functions are unconstrained, meaning that they are not bounded by a particular value; therefore induced demand is allowed. With this approach, in periods where historically demand was at capacity, the predicted demand obtained from this approach does not have such restriction.

6.1.2 Results and Interpretations

The models are estimated with Stata 9.0, data analysis and statistical software (Stata Corp). The price coefficients have the expected sign for the majority of the models. The square of the advance booking 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 destination which shows the best model fit is station 8; 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 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 6.1 Linear demand function result (origin16-destination1)

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)		
Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	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	-	-	AdvbkSq	-	-	AdvbkSq	-0.040	-1.213	AdvbkSq	-	-
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 6.2 Linear demand function result (origin16-destination8)

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)		
Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	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	-	-	AdvbkSq	-0.757	-8.406	AdvbkSq	-3.492	-3.700	AdvbkSq	-	-
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 6.3 Linear demand function result (origin16-destination12)

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)		
Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	T-Stat	Variable	Est	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	-	-	AdvbkSq	-	-	AdvbkSq	-0.832	-2.892	AdvbkSq	-	-
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

The validation of the demand functions on a hold-out sample extracted from the period March 16 to March 22, 2009 is reported in Table 6.4. Results indicate that the volumes of ticket sales predicted are comparable to the actual demand.

Table 6.4 Demand function out of sample validation

Station	Demand (No. of passengers)	Monday March 16	Tuesday March 17	Wednesday March 18	Thursday March 19	Friday March 20	Saturday March 21	Sunday March 22
1	Actual	38	39	50	28	36	16	32
	Predicted	35	38	31	27	34	11	56
2	Actual	4	9	15	10	11	5	5
	Predicted	6	7	12	10	12	5	5
6	Actual	43	35	38	34	57	22	35
	Predicted	34	33	39	43	46	24	44
7	Actual	36	34	57	56	76	8	49
	Predicted	36	43	55	63	70	11	39
8	Actual	1,432	1,584	1,845	1,684	1,776	466	1,302
	Predicted	1,527	1,734	1,903	1,952	1,862	478	1,136
9	Actual	251	227	316	351	314	39	114
	Predicted	246	313	358	381	341	53	97
10	Actual	204	235	288	271	231	36	47
	Predicted	148	262	298	306	278	41	81
12	Actual	546	633	559	739	664	97	212
	Predicted	468	684	667	745	587	81	204

6.2 Log-Linear Demand Function

The log-linear-demand function is broadly used in econometric studies and has several desirable theoretical and practical properties. By taking the log of the demand as dependent variable, the estimated demand is bounded by zero. This also allows for the estimation using linear regression.

Data used for this model are based on monthly ticket reservation data of intercity passenger railway in the U.S. in 2009. Data focus on the coach class passenger traveling on the northbound trip in which the ticket is confirmed. This results in the total of 110,828 reservation records.

6.2.1 Model Specification

For each departure day over the entire month, the data are prepared by summing the number of passenger demand on each booking day. This number is transformed to log of passenger demand. The average fare paid for each booking day is used as the fare variable. The independent variables which enter the final model include the intercept, the square of the time until departure ($advbk^2$), fare ($fare$), weekend dummy ($wknd$) indicating whether the departure day is in weekend, and booking period specific intercepts. The square of the advance booking is motivated from the analysis of the booking data where the non-linear relationship between advance booking and number of passenger booking is observed. The weekend dummy is included to account for the seasonality of train departing on different day which is motivated by the demand function in Sibdari et al. (2008). The specification of the log-linear demand function can be expressed as follows:

$$\log[D_{od}(j)] = \alpha_0 + \left(\alpha_{advbksq} \times advbk_j{}^2\right) + \left(\alpha_{fare} \times fare_j\right) + (\alpha_{wknd} \times wknd) + \sum_{k=1}^{6}(\alpha_k \times bkperiod_j) + \varepsilon \tag{6.3}$$

$$D_{od}(j) = \exp\left\{\alpha_0 + \left(\alpha_{advbksq} \times advbk_j{}^2\right) + \left(\alpha_{fare} \times fare_j\right) + (\alpha_{wknd} \times wknd) + \sum_{k=1}^{6}(\alpha_k \times bkperiod_j) + \varepsilon\right\} \tag{6.4}$$

where

$D_{od}(j)$ = Demand on booking day *j* for origin *o*, destination *d*

$advbk_j^2$ = Square of advance booking for booking day *j*

$fare_j$ = Fare ($) of booking day *j*

$wknd$ = Weekend dummy of departure day

$bkperiod_j$ = Booking period dummy (1 if $j \in k$, 0 otherwise)

α_0 = Intercept

ε = Error term

The daily passenger demand for each origin-destination pair is the summation of passenger reservation on each booking day *d* over the sale horizon:

$$D_{od} = \sum_{j=1}^{31} D_{od}(j) \tag{6.5}$$

6.2.2 Results and Interpretations

The models are estimated with Stata 9.0, data analysis and statistical software (Stata Corp). There are total of 119 origin-destination pair for the north bound trip which results in the total of 119 demand function estimated. However, in this chapter, only demand functions of major markets are shown in this report in Table 6.5 to Table 6.9.

Table 6.5 Log-linear demand function result (origin 16-destination 1)

Variable	Est	Std Err.	T-Stat	P>|t|	[95% confidence	Interval]
advbksq	0.0003	0.0004	0.86	0.389	-0.0004	0.001
fare	-0.0003	0.0009	-0.34	0.732	-0.002	0.0014
wkndmy	0.0314	0.0673	0.47	0.641	-0.1011	0.1639
period 1	(dropped)					
period 2	-0.0531	0.1649	-0.32	0.748	-0.3776	0.2715
period 3	0.0899	0.2138	0.42	0.675	-0.331	0.5107
period 4	0.3009	0.23	1.31	0.192	-0.1519	0.7537
period 5	0.5023	0.2502	2.01	0.046	0.0099	0.9947
period 6	1.1967	0.2502	4.78	0	0.7043	1.6891
constant	0.2092	0.2597	0.81	0.421	-0.3021	0.7204
No. obs	297				R-squared	0.277
F(8, 288)	13.81				Adj R-squared	0.257
Prob>F	0				Root MSE	0.504

Table 6.6 Log-linear demand function result (origin 16-destination 8)

Variable	Est	Std Err.	T-Stat	P>|t|	[95% confidence	Interval]
advbksq	-0.0018	0.0003	-6.46	0	-0.0024	-0.0013
fare	-0.005	0.0015	-3.36	0.001	-0.0079	-0.0021
wkndmy	-0.2318	0.0664	-3.49	0.001	-0.3622	-0.1014
period 1	-3.3826	0.2444	-13.84	0	-3.8623	-2.9029
period 2	-3.2735	0.1811	-18.08	0	-3.6289	-2.9181
period 3	-2.8051	0.1696	-16.54	0	-3.138	-2.4723
period 4	-1.8583	0.169	-10.99	0	-2.1901	-1.5265
period 5	-0.7223	0.2119	-3.41	0.001	-1.1381	-0.3065
period 6	(dropped)					
constant	6.4649	0.3018	21.42	0	5.8726	7.0572
No. obs	886				R-squared	0.657
F(8, 877)	209.65				Adj R-squared	0.654
Prob>F	0				Root MSE	0.834

Table 6.7 Log-linear demand function result (origin 16-destination 12)

Variable	Est	Std Err.	T-Stat	P>|t|	[95% confidence	Interval]
advbksq	-0.0008	0.0003	-2.57	0.01	-0.0015	-0.0002
fare	-0.0003	0.0015	-0.17	0.864	-0.0032	0.0027
wkndmy	-0.7216	0.0763	-9.46	0	-0.8713	-0.5718
period 1	-2.4704	0.2537	-9.74	0	-2.9685	-1.9723
period 2	-2.46	0.1666	-14.76	0	-2.7872	-2.1328
period 3	-1.8049	0.1537	-11.74	0	-2.1067	-1.5032
period 4	-1.2621	0.1527	-8.27	0	-1.562	-0.9623
period 5	(dropped)					
period 6	1.1766	0.1926	6.11	0	0.7984	1.5548
constant	3.6227	0.2353	15.4	0	3.1607	4.0848
No. obs	652				R-squared	0.68
F(8, 643)	170.39				Adj R-squared	0.676
Prob>F	0				Root MSE	0.758

Table 6.8 Log-linear demand function result (origin 16-destination 13)

Variable	Est	Std Err.	T-Stat	P>|t|	[95% confidence	Interval]
advbksq	0.0002	0.0007	0.27	0.786	-0.0011	0.0015
fare	-0.0052	0.002	-2.54	0.012	-0.0092	-0.0012
wkndmy	-1.0837	0.1186	-9.14	0	-1.3171	-0.8503
period 1	-3.3299	0.4431	-7.51	0	-4.2021	-2.4577
period 2	-3.122	0.2142	-14.57	0	-3.5437	-2.7004
period 3	-2.899	0.1663	-17.43	0	-3.2264	-2.5716
period 4	-2.1572	0.1547	-13.95	0	-2.4616	-1.8527
period 5	-1.0922	0.188	-5.81	0	-1.4623	-0.7221
period 6	(dropped)					
constant	4.1615	0.2464	16.89	0	3.6766	4.6465
No. obs	296				R-squared	0.681
F(8, 287)	76.45				Adj R-squared	0.672
Prob>F	0				Root MSE	0.727

Table 6.9 Log-linear demand function result (origin 16-destination 14)

Variable	Est	Std Err.	T-Stat	P>|t|	[95% confidence	Interval]
advbksq	0.0004	0.001	0.34	0.733	-0.0017	0.0024
fare	-0.0108	0.0049	-2.2	0.029	-0.0204	-0.0011
wkndmy	-0.5392	0.101	-5.34	0	-0.7388	-0.3396
period 1	(dropped)					
period 2	-0.0652	0.4432	-0.15	0.883	-0.9408	0.8103
period 3	0.0353	0.5758	0.06	0.951	-1.1022	1.1728
period 4	0.3672	0.6143	0.6	0.551	-0.8463	1.5807
period 5	1.3027	0.6307	2.07	0.041	0.0567	2.5486
period 6	3.3476	0.6314	5.3	0	2.1002	4.595
constant	0.5958	0.6452	0.92	0.357	-0.6789	1.8705
No. obs	162				R-squared	0.873
F(8, 153)	131.73				Adj R-squared	0.867
Prob>F	0				Root MSE	0.473

The estimated demand functions show goodness of fit from 0.277 to 0.873. Most of the parameters estimates have the expected sign, and are statistically significant at the 5% significance level. The square of advance booking and fare coefficients have expected sign in majority of the models shown in the examples. The sign of the square of advance booking is in line with the data analysis which indicates that number of reservation increase non-linearly as time approaches departure. The weekend dummy is statistically significant at the 5% significance level in most of the models. The negative coefficient of weekend is in line with the expectation since based on the data analysis, the passenger demand on Saturday and Sunday are significantly lower than weekday. The booking day period intercept is statistically significant and shows the expected pattern with increasing value as departure period approaches departure.

6.3 Conclusions

In this chapter, the demand functions are proposed to represent variation of demand volume in response to RM policy while taking into account trip departure characteristics. Demand functions estimated with linear and log-linear regression are presented, both approaches provides intuitive results which are in line with the expectation and capable of representing demand response to RM policy.

Chapter 7 : Revenue Optimization Framework

In the previous two chapters, discrete choice models and demand functions are developed. The choice models represent the passenger purchase timing decisions, while the demand functions represent passenger demand volume in response to fare and departure characteristics. In this chapter, the revenue optimization problem that incorporates both the choice models and the demand functions are developed to derive a fare strategy. Two optimization approaches are developed; the first is based on a single-leg approach, while the second proposes a network approach.

The single-leg approach is presented in Section 7.1, the problem assumes that revenue optimization is derived independently for each market without consideration of capacity redistribution among markets. The formulation proposed optimizes ticket revenue for each daily departure. The formulation considers markets with single-origin and multiple-destinations by incorporating the choice models in Section 7.1.2 and the demand functions in Section 6.1.

The network approach is presented in Section 7.2 and Section 7.3, the problem accounts for the shared capacity among multiple markets by allowing for seat capacity redistribution. The problem is formulated to jointly solve for pricing and seat allocation and optimizes ticket revenue for each train trip. The problem considers a network of selected stations by incorporating the choice models presented in Chapter 5. In Section 7.2, the heterogeneous choice models based on discrete segmentation (latent class model) is incorporated into the problem that considers nine stations in the network. Section 7.3 incorporates the heterogeneous choice model based on continuous segmentation (mixed logit model) with parametric and non-parametric distribution to a reduced network including five stations.

7.1 Single-Leg Revenue Optimization

7.1.1 Optimization Framework and Procedure

The optimization framework incorporates both passenger choice models and the demand functions (Cirillo et al., 2011). A two step process is proposed to model passenger demand. In the first step, passenger choice model of purchase timing is estimated using a multinomial logit model. In the second step, a linear regression determines passenger demand in response to fare. The proposed models are incorporated into RM optimization model to maximize expected revenue. With this framework, fare strategies that can be updated on a daily basis over the sale horizon are strategically calculated from this optimization problem.

In this setting, it is assumed that the railway operator aims to maximize ticket revenue for each daily departure. In this specific case study, the fares over a representative week in March (March 16 to March 22, 2009) are optimized. The fare strategy resulting from the optimization process is tested by comparing "model" revenue to the actual revenue registered from March 16 to March 22, 2009. This new fare strategy is then applied to a representative week in April (April 20 to April 26, 2009). This is mainly to show how fare strategies derived from choice models estimated for a specific month performs in the subsequent month.

7.1.2 Choice Model

A. Sample Selection

The same data set used for choice model estimation in Chapter 5 is used in the analysis. Data on coach class passengers traveling from south end station (station 16) to other stations in the corridor are selected. This results in 44,847 valid observations.

B. Model Specification

The specification of the final model includes the following explicative variables: advance booking (number of days before departure), fare ($), destination specific dummies, and long distance dummy. Some parameters are allowed to take different values in each booking period. The 31 days sale horizon is divided into 6 booking periods denoted by k as: (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 advance booking coefficient enters the model as a generic parameter while the fare, 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 model results adopting this approach. In this study, destination markets are selected 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 of 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: Station 1, Station 8, and Station 12. The effect of long distance is motivated by Whelan et al. (2008). In this study, it was found that leisure trips are in general long distance trips. In this context, by taking into account long distance trip variable the model is expected 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 j which falls within the booking period k can be expressed as:

$$U_i(j,k) = \left(\beta_{advbk} \times advbk_j\right) + \left(\beta^k_{fare} \times fare_j\right) + \left(\beta^k_{ST1} \times ST1_i\right) + \left(\beta^k_{ST8} \times ST8_i\right) + \left(\beta^k_{ST12} \times ST12_i\right) + \left(\beta^k_{LONG} \times LONG_i\right) + \varepsilon_{ij} \quad (7.1)$$

where the independent variables and their associated index are:

j = Booking day, $j \in \{1, \dots, 31\}$

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

$advbk_j$= Advance booking of booking day j

$fare_j$ = Fare of booking day j ($)

$ST1$ = Destination dummy (1 if destination is Station 1, 0 otherwise)

$ST8$ = Destination dummy (1 if destination is Station 8, 0 otherwise)

$ST12$ = Destination dummy (1 if destination is Station 12, 0 otherwise)

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

ε_{ij} = A mutually independent noise term of individual i on choice j following a Gumbel distribution

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

$$\Pr\left(bkday_j\right) = \frac{\exp[U_i(j,k)]}{\sum_{m=1}^{31} \exp[U_i(m,k)]} \tag{7.2}$$

C. Result Interpretation

The results obtained from the choice model calibration are reported in Table 7.1; most of the parameters estimated have the expected sign and statistically significant. The price coefficients show high statistical significance and the expected sign. The monotonically increasing value of price coefficients 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 coefficient indicates that passengers are insensitive to price on the day of departure. This result is reasonable for this railway service, given that the majority of the passengers are business travelers who book the ticket close to the departure date and are not very sensitive to fare. The smaller magnitude of the price coefficient 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.

Station 1 destination coefficients are coherent with the expected pattern. The monotonically decreasing value of the coefficients with respect to booking time implies that it is preferred to book the ticket for Station 1 as early as possible to ensure the availability of the seat. Station 1 is a high demand long distance market, with travel time by railway comparable to travel time by plane. Station 8 destination coefficients show an opposite pattern to Station 1, the increasing value of the coefficients 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. Station 12 destination coefficients follow the same pattern as Station 1, the most preferred booking periods being periods 2 and 1 respectively.

Long distance coefficients show statistically significant coefficients except for booking period 3. The long distance coefficients pattern is in line with the expectation; the earlier booking periods being more preferred by 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 coefficient indicates that it is generally more preferred 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 7.1 Passenger choice model result

I. Alternative specific estimates	Price	T-stat	Station1	T-stat	Station8	T-stat	Station12	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	-0.1100	-0.4	-0.1939	-0.8
Booking period4 (Day26-29)	-0.0132*	-23.1	-0.6941*	-2.2	-0.2819*	-2.2	-0.4220	-1.4	-0.6633*	-2.8
Booking period5 (Day30)	-0.0056*	-9.9	-0.9606*	-2.8	0.0700	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					Log-likelihood at zero			-154,004
							Log-likelihood at optimal			-107,192
							Number of observations			44,847

*Statistically significant at the 5% significance level.

7.1.3 Model Validation

To illustrate the choice model prediction capability, the model is validated on the major destination markets which will be included in the revenue optimization in the next section.

The prediction using the model estimated from the entire data set (within sample validation) is shown in Table 7.2. The prediction using the model estimated from the data excluding hold out sample (March 16 to March 22, 2009) (out of sample validation) is shown in Table 7.3

Results indicate that the choice model performs better for stations with large sample size such as Station 8, 9, 10, and 12 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 7.2 Choice model within sample validation

Booking Day	No. of reservations	Station								Total
		1	**2**	**6**	**7**	**8**	**9**	**10**	**12**	**(15 Stations)**
16	Actual	10	2	7	11	275	41	27	59	450
	Predicted	10	1	6	4	247	32	16	56	391
17	Actual	18	2	10	16	375	36	43	64	595
	Predicted	15	1	4	5	341	50	32	61	526
18	Actual	7	3	11	16	270	45	33	59	463
	Predicted	8	1	9	14	267	46	39	57	466
19	Actual	11	2	12	17	272	43	25	54	457
	Predicted	13	1	8	10	322	80	52	63	575
20	Actual	10	3	17	17	257	42	42	68	487
	Predicted	15	1	9	14	426	78	51	97	721
21	Actual	14	1	17	14	315	57	37	74	557
	Predicted	9	1	12	17	400	78	57	101	724
22	Actual	7	2	16	27	401	73	57	109	737
	Predicted	12	2	12	16	457	90	66	122	838
23	Actual	20	5	20	31	529	97	82	163	984
	Predicted	14	3	14	21	569	116	78	158	1,050
24	Actual	19	3	28	39	813	165	131	211	1,478
	Predicted	17	3	17	24	740	143	90	184	1,315
25	Actual	10	2	28	31	1,029	134	133	241	1,693
	Predicted	18	3	26	30	903	186	117	230	1,635
26	Actual	27	6	37	44	1,082	211	117	249	1,867
	Predicted	19	4	28	39	1,190	207	138	284	2,085
27	Actual	20	6	49	57	1,355	224	163	327	2,345
	Predicted	23	5	34	35	1,429	256	161	342	2,498
28	Actual	20	2	42	60	1,765	292	158	433	2,953
	Predicted	28	5	43	44	1,731	296	192	429	3,036
29	Actual	31	9	60	66	2,076	345	200	539	3,541
	Predicted	29	7	47	57	1,936	336	214	496	3,432
30	Actual	45	14	70	100	4,380	665	284	1,242	7,415
	Predicted	45	14	93	113	4,410	541	376	1,249	7,560
31	Actual	106	9	97	145	7,539	1,221	875	3,373	15,686
	Predicted	106	30	217	260	7,509	1,080	722	3,365	15,052
Total	**(31 Booking days)**	**462**	**90**	**623**	**756**	**24,645**	**3,916**	**2,593**	**7,668**	**44,847**

Table 7.3 Choice model out of sample validation

Booking Day	No. of reservations	Station								Total
		1	**2**	**6**	**7**	**8**	**9**	**10**	**12**	**(15 Stations)**
16	Actual	3	0	2	3	65	2	4	11	97
	Predicted	2	0	1	1	51	7	4	13	83
17	Actual	5	2	3	7	74	7	10	17	133
	Predicted	3	0	1	1	71	11	7	14	113
18	Actual	3	2	1	1	65	6	8	13	102
	Predicted	2	0	2	3	56	10	9	13	100
19	Actual	4	1	3	7	52	8	6	14	101
	Predicted	3	0	2	2	68	18	12	15	125
20	Actual	4	0	2	7	60	10	14	17	121
	Predicted	4	0	2	3	91	17	12	23	158
21	Actual	4	0	2	3	66	16	7	18	124
	Predicted	2	0	3	4	84	17	13	22	156
22	Actual	3	0	3	2	83	18	13	28	161
	Predicted	3	1	3	3	97	20	15	28	183
23	Actual	4	3	2	8	117	14	12	36	207
	Predicted	3	1	3	5	123	26	18	36	231
24	Actual	7	0	3	6	160	23	24	51	290
	Predicted	4	1	4	5	161	32	21	42	292
25	Actual	1	1	7	7	186	27	30	54	331
	Predicted	4	1	6	7	199	42	28	53	368
26	Actual	7	1	7	12	213	62	34	49	408
	Predicted	5	1	6	8	245	43	30	63	437
27	Actual	6	2	10	10	281	52	38	76	494
	Predicted	6	1	7	7	298	53	35	76	530
28	Actual	4	0	13	15	394	57	33	118	676
	Predicted	7	1	10	9	365	62	42	97	652
29	Actual	5	3	15	11	472	84	58	132	829
	Predicted	7	2	10	12	412	71	48	113	742
30	Actual	13	3	20	14	894	148	59	291	1,578
	Predicted	10	4	21	24	934	117	86	282	1,636
31	Actual	15	2	22	31	1,554	261	194	750	3,336
	Predicted	30	8	49	55	1,621	237	167	783	3,344
Total	**(31 Booking days)**	**113**	**24**	**141**	**160**	**5,227**	**841**	**590**	**1,759**	**9,760**

7.1.4 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:

$$\max_{fare_j} Rev = \left[\min\left\{\sum_{j=1}^{31} Demand(j), Capacity\right\}\right] \times \left[\sum_{j=1}^{31}\left\{fare_j \times \Pr(day_j)\right\}\right] \quad (7.3)$$

The first term represents the acceptable demand volume on a particular departure day. The acceptable demand is the minimum between the predicted demand and the train capacity; this is to ensure that acceptable demand can be accommodated with the train capacity. The predicted demand is obtained from the demand function in Section 6.1. The train capacity allocated for each destination market is approximated with historical data by assuming that the actual demand was at 80 percent of the allocated capacity. The capacity redistribution is not accounted in this analysis because fare optimization for each origin-destination pair is treated as an independent problem.

The second term in the objective function represents the expected fare expressed as the booking day specific fare ($fare_j$) weighted by the probability that passengers purchase the ticket on that booking day($\Pr(day_j)$). The probability is obtained from the choice model parameters in Section 7.1.2. Thus, the overall formulation represents the expected revenue per day for each destination with the booking day specific fare($fare_j$) as decision variables. It is assumed that the railway operator fare strategy is subjected to predetermined fare bound restriction and that the fares can only increase monotonically as time approaches departure. The incremental amount of fare from one booking day to the next is assumed to be within a certain value C (assumed to be \$ 5.00). The assumptions used in this research do not necessarily represent this railway operator RM policy. The constraints for this optimization problem corresponded to our RM control assumptions are:

$$fare_{lb} \leq fare_j \leq fare_{ub} \tag{7.4}$$

$$fare_j \leq fare_l \ \forall \ l > j \tag{7.5}$$

$$fare_l - fare_j \ \leq C \ (\$) \ \forall \ l = j + 1 \tag{7.6}$$

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

$$fare_j = \text{Real decision variable} \in \ R_+^{31} \tag{7.7}$$

$$fare_{lb} = \text{Lower bound on fare (\$) for each destination} \tag{7.8}$$

$$fare_{ub} = \text{Upper bound on fare (\$) for each destination} \tag{7.9}$$

7.1.5 Optimization Result

The optimization problem 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 exponential function of the MNL choice probability. The methodology is applied to 8 major destination markets out of the 15 destinations. The resulting fares in March representative week (March 16 to March 22) for 8 representative destinations are shown in Figure 7.1 to Figure 7.8.

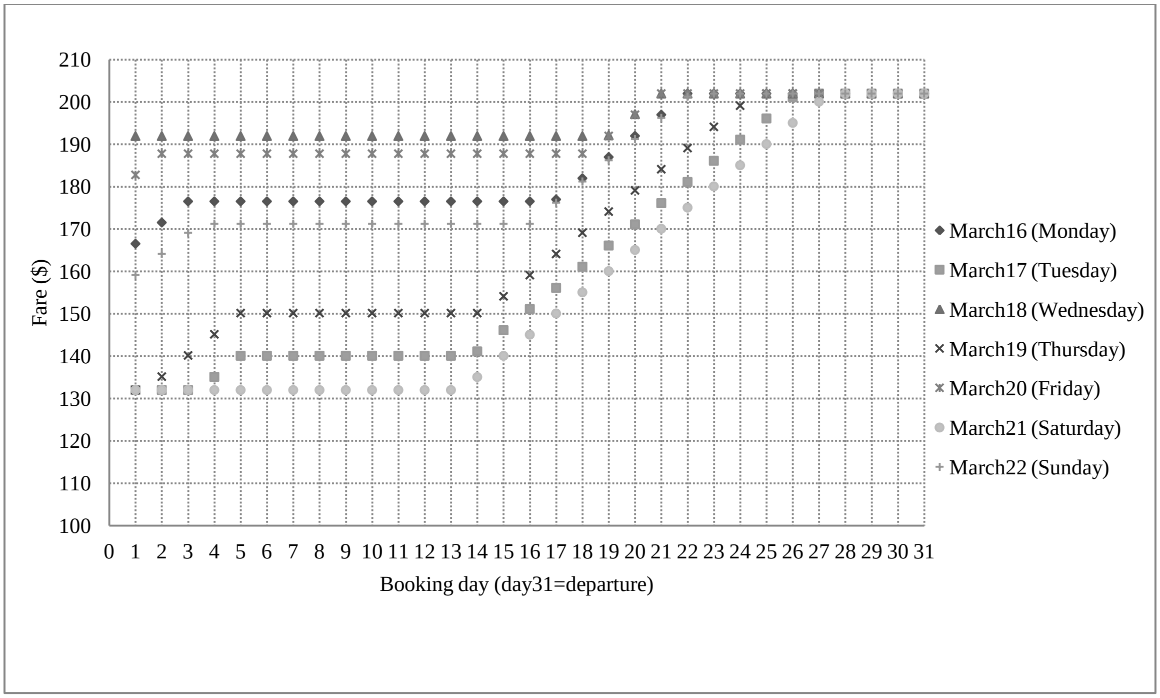


Figure 7.1 Fare strategy (origin 16-destination 1)

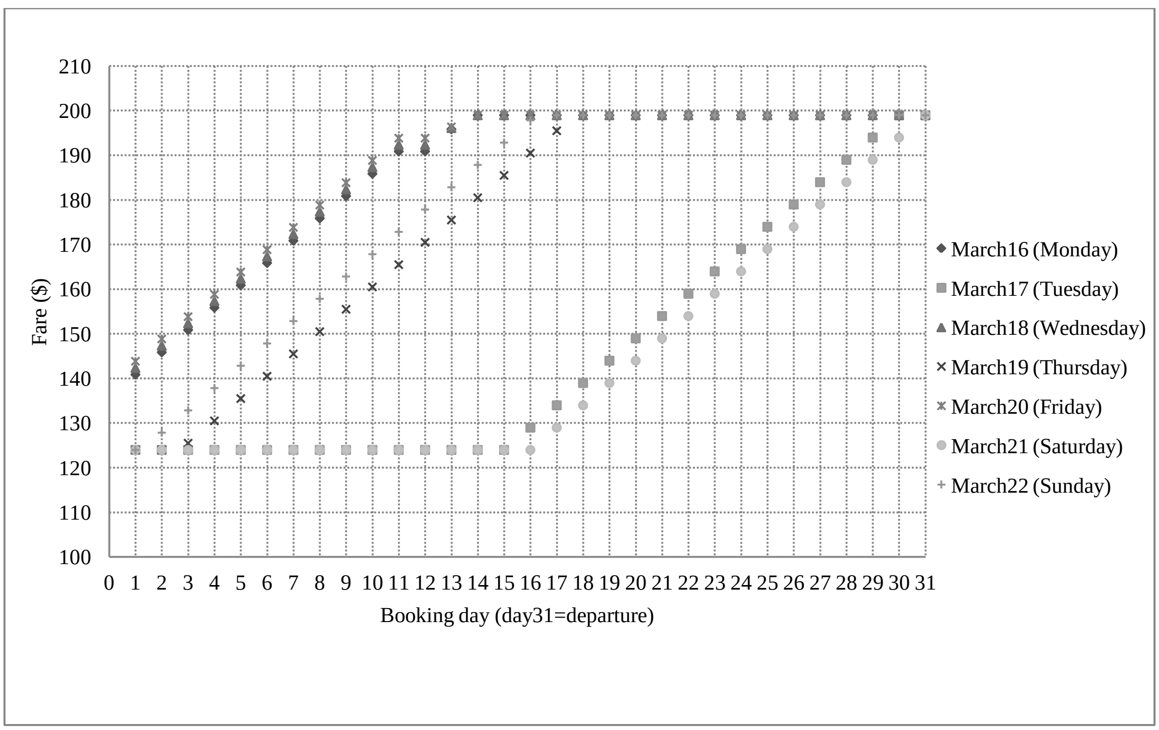


Figure 7.2 Fare strategy (origin 16-destination 2)

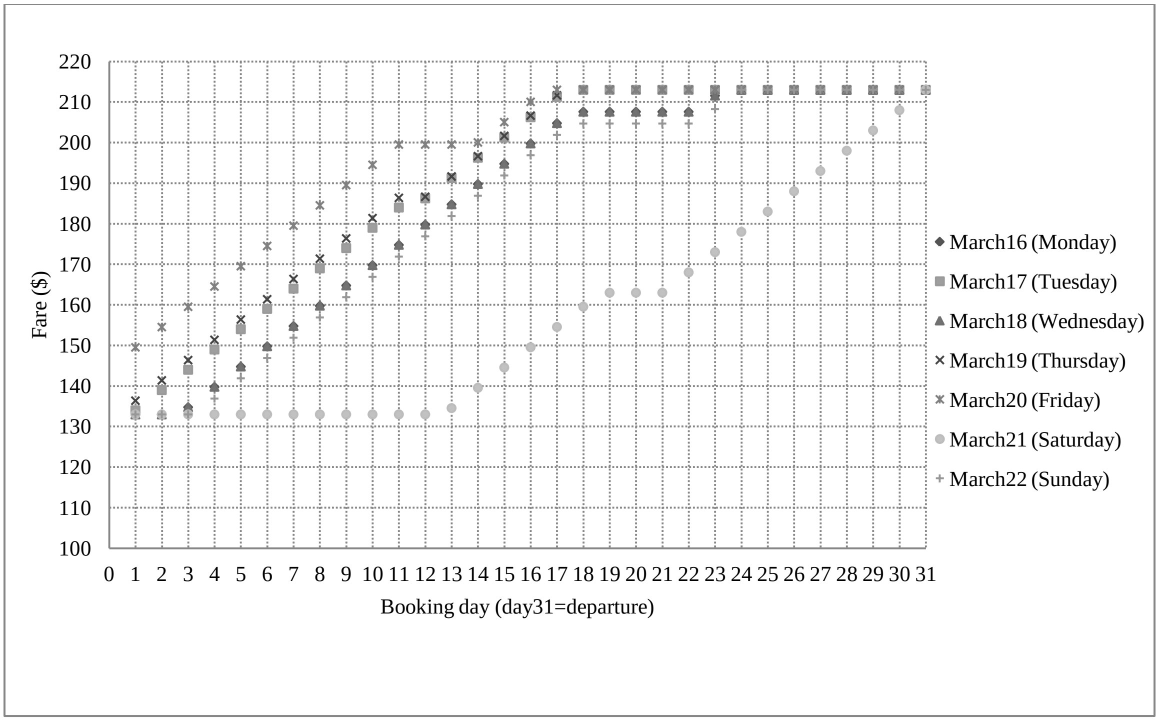


Figure 7.3 Fare strategy (origin 16-destination 6)

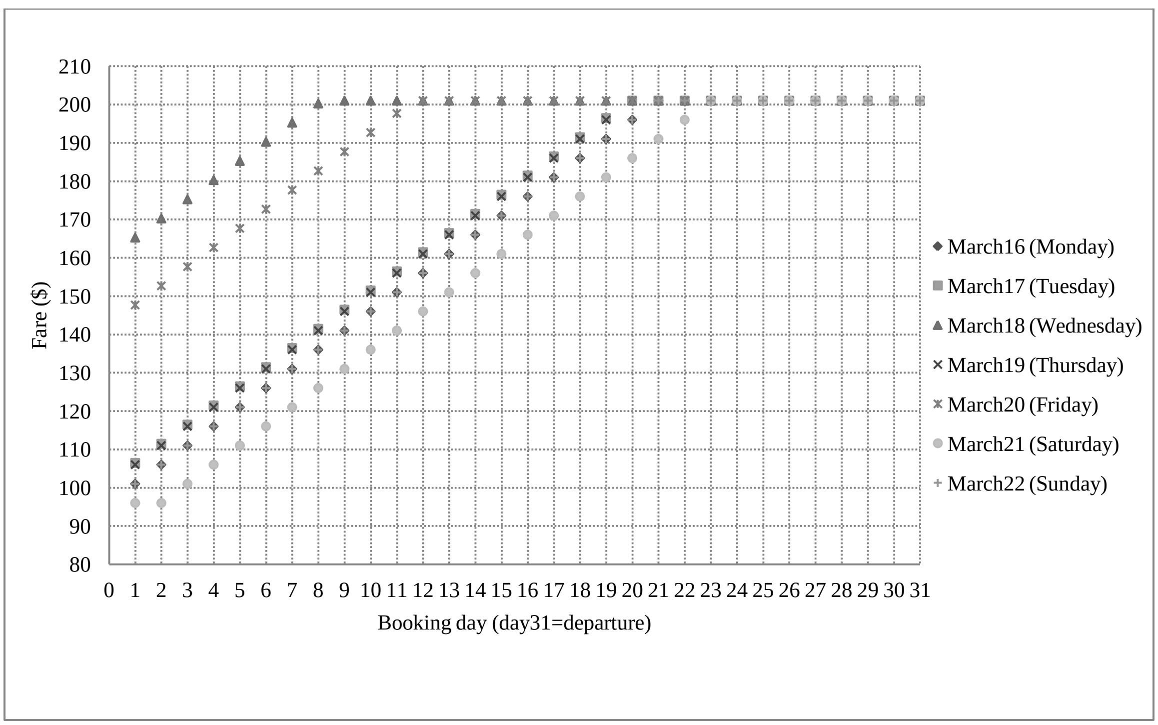


Figure 7.4 Fare strategy (origin 16-destination 7)

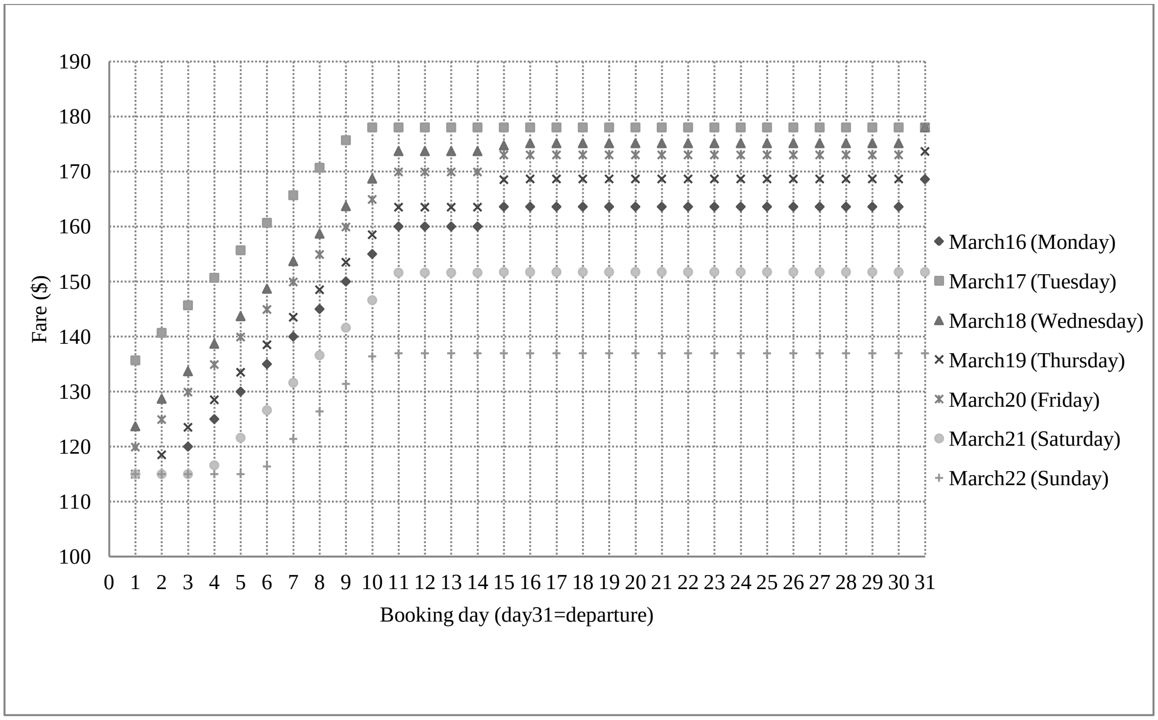


Figure 7.5 Fare strategy (origin 16-destination 8)

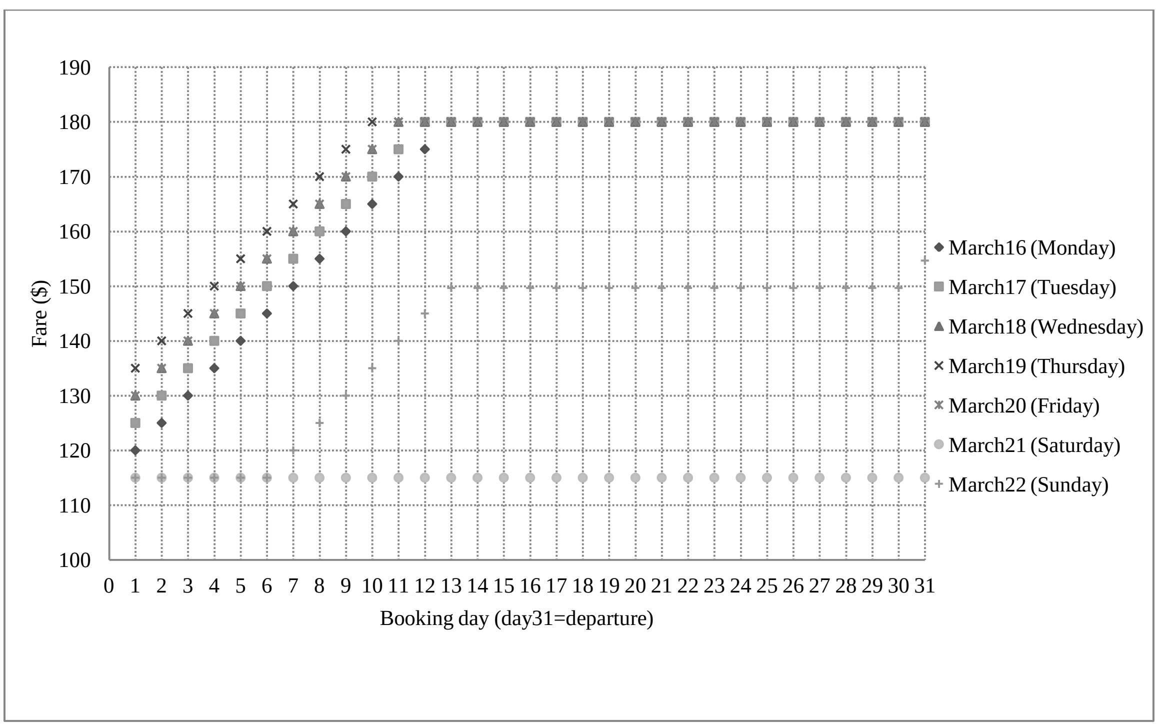


Figure 7.6 Fare strategy (origin 16-destination 9)

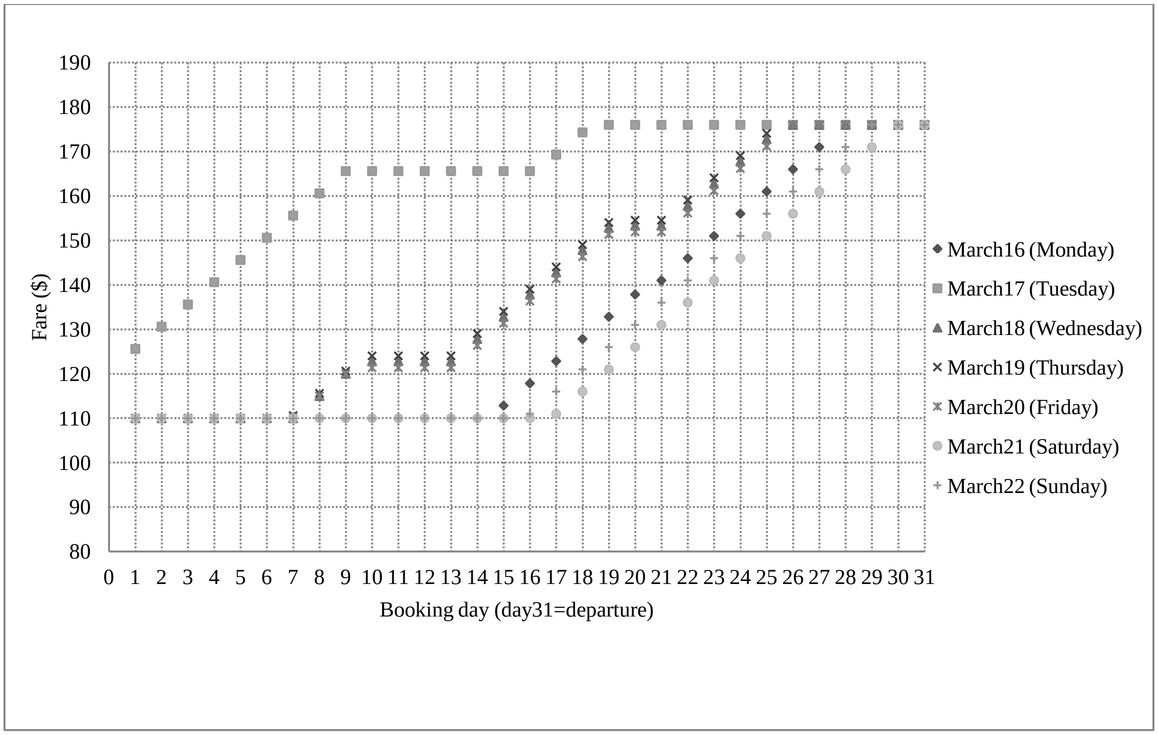


Figure 7.7 Fare strategy (origin 16-destination 10)

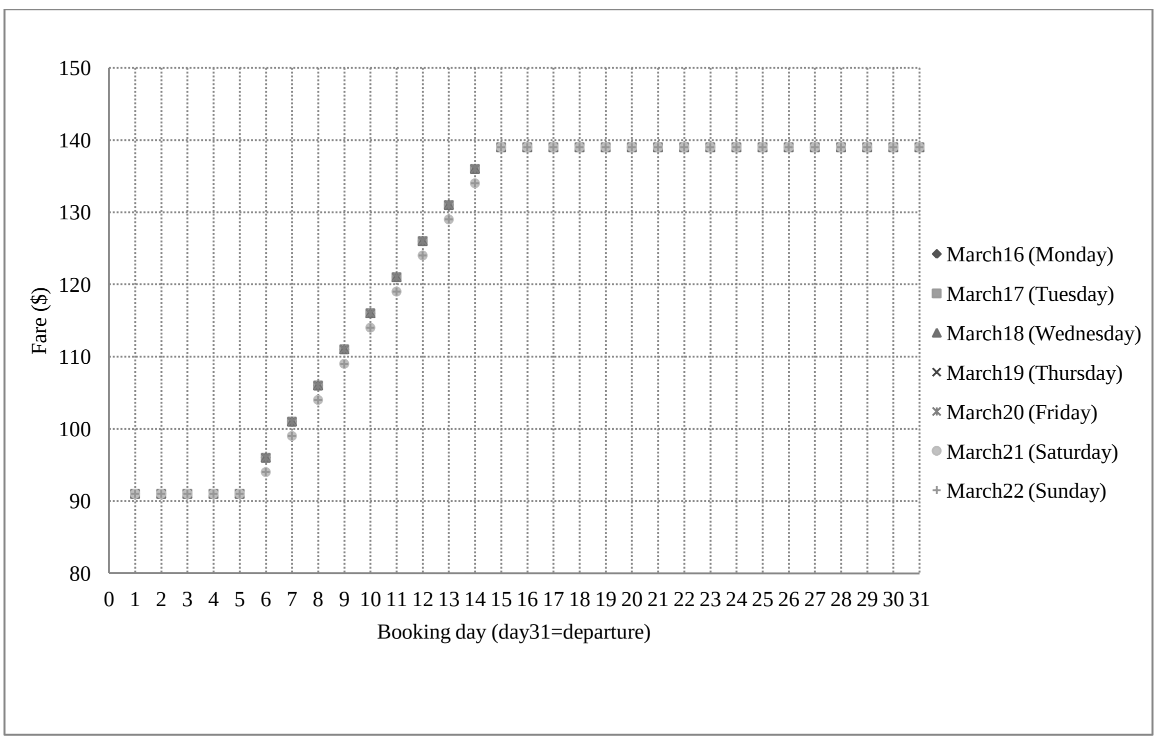


Figure 7.8 Fare strategy (origin 16-destination 12)

For Station 1, the fare strategy conforms to its demand function estimated. Its demand function indicates the highest demand on Wednesday and the lowest on Saturday particularly in the booking period 1 (Day 1-11), 2 (Day 1-20) and 4 (Day 26-29). The fare strategy suggests that the fares should be charged high on high demand day (Wednesday) and low on low demand day (Saturday). Fare strategies obtained from other stations show similar pattern to Station 1.

For Station 2, its demand function indicates high demand on Monday, Wednesday, and Friday. In particular, Monday has high demand in booking period 2 to 5 (Day 12-31), Wednesday has high demand in the booking period 1 (Day 1-11), and Friday has 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 imposes the highest fare to Monday, Wednesday, and Friday and the lowest fare to Saturday.

For Station 6, its demand function indicates high demand on Friday especially in booking periods 3 to 5 (Day 21-31). 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 imposes the highest fare on Friday and the lowest fare on Saturday.

For Station 7, its demand function indicates high demand on Wednesday especially in booking periods 2 (Day 12-20) and 4 (Day 26-29). The lowest demand is on Saturday in booking periods 1 (Day 1-11), 3 (Day 21-25), 4 (Day 26-29), and 5 (Day 30-31). Thus, the fare strategy imposes the highest fare to Friday and the lowest fare to Saturday.

For Station 8, its demand function indicates high demand on Tuesday especially in booking periods 2 (Day 12-20) and 5 (Day 30-31). Thus, the fare strategy imposes the highest fare to Tuesday and the lowest fare on Saturday.

For Station 9, its demand function indicates high demand on Thursday especially in booking periods 1 (Day 1-11), 3 (Day 21-25), 4 (Day 26-29), and 5 (Day 30-31). The lowest demand is on Saturday from booking period 2 to 5 (Day 12-31). Thus, the fare strategy imposes the highest fare on Thursday and the lowest fare on Saturday.

For Station 10, its demand function indicates high demand on Tuesday especially in booking period 2 (Day 12-20) and 5 (Day 30-31). The lowest demand is on Saturday in all booking periods. Thus, the fare strategy imposes the highest fare on Tuesday with the lowest fare on Saturday.

For Station 12, its demand function indicates high demand on Thursday in booking period 3 to 5 (Day 21-31). The lowest demand is on Saturday for all booking periods. However, due to the relatively low fares for this station, the feasible fare range imposed by the constraint is smaller compared to other stations. In this case, the day of week effect does not sufficiently influence fare strategy.

These findings support the intuition that high demand days have a larger impact on revenue maximization than low demand days. The revenue improvements for each departure day of the representative weeks in March (March 16 to March 22) and April (April 20 to April 26) are shown in Table 7.4 and Table 7.5. 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%

per day and from 0.65% to 10.60% per day for March and April representative weeks respectively. Results indicate that the application of the proposed fare strategy results into significant revenue improvement on Monday, Tuesday, and Saturday.

Table 7.4 Revenue change 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. Station6	51.63	16.74	20.23	9.8	6.71	23.57	4.17
4. Station7	18.03	2.8	8.71	2.83	2.62	15.35	13.98
5. Station8	10.36	11.46	5.52	1.32	5.5	13.01	-0.71
6. Station9	20.68	13.85	4.99	2.31	3.7	-8.77	0.98
7. Station10	16.35	17.23	-0.5	1.9	-2.23	18.98	12.23
8. Station12	20.41	11.77	6.81	2.95	5.41	45.44	25.5
9. Other 7 Stations	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total change (%)	13.76	11.52	5.26	1.92	4.43	13.74	2.45

Table 7.5 Revenue change 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. Station6	39.70	27.78	9.07	17.17	4.31	9.56	10.42
4. Station7	20.98	9.64	19.54	-5.03	-0.35	4.31	2.56
5. Station8	5.40	8.59	4.49	1.25	3.76	12.36	-4.03
6. Station9	14.75	9.51	2.20	-0.38	1.37	-15.33	4.35
7. Station10	10.93	9.40	-2.47	-3.16	-3.97	23.08	22.38
8. Station12	11.41	13.53	9.34	8.02	3.50	27.46	33.21
9. Other 7 Stations	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total change (%)	8.40	9.33	5.07	2.10	2.70	10.60	0.65

7.2 Network Revenue Optimization with Discrete Market Segmentation

Pricing and seat allocation have often been considered as two independent problems. Nevertheless, the two problems are interrelated and complementary to one another. In this section, the simultaneous RM optimization model of pricing and seat allocation is proposed. The framework incorporates choice model which realistically represent railway market by allowing for heterogeneity across categories of passengers based on latent class approach (Hetrakul and Cirillo, forthcoming). This study primarily adopts the framework of Ongprasert (2006), and further allows for pricing and seat allocation to be simultaneously optimized based on heterogeneous categories of passengers under capacity constraints determined on the basis of the railway network characteristics.

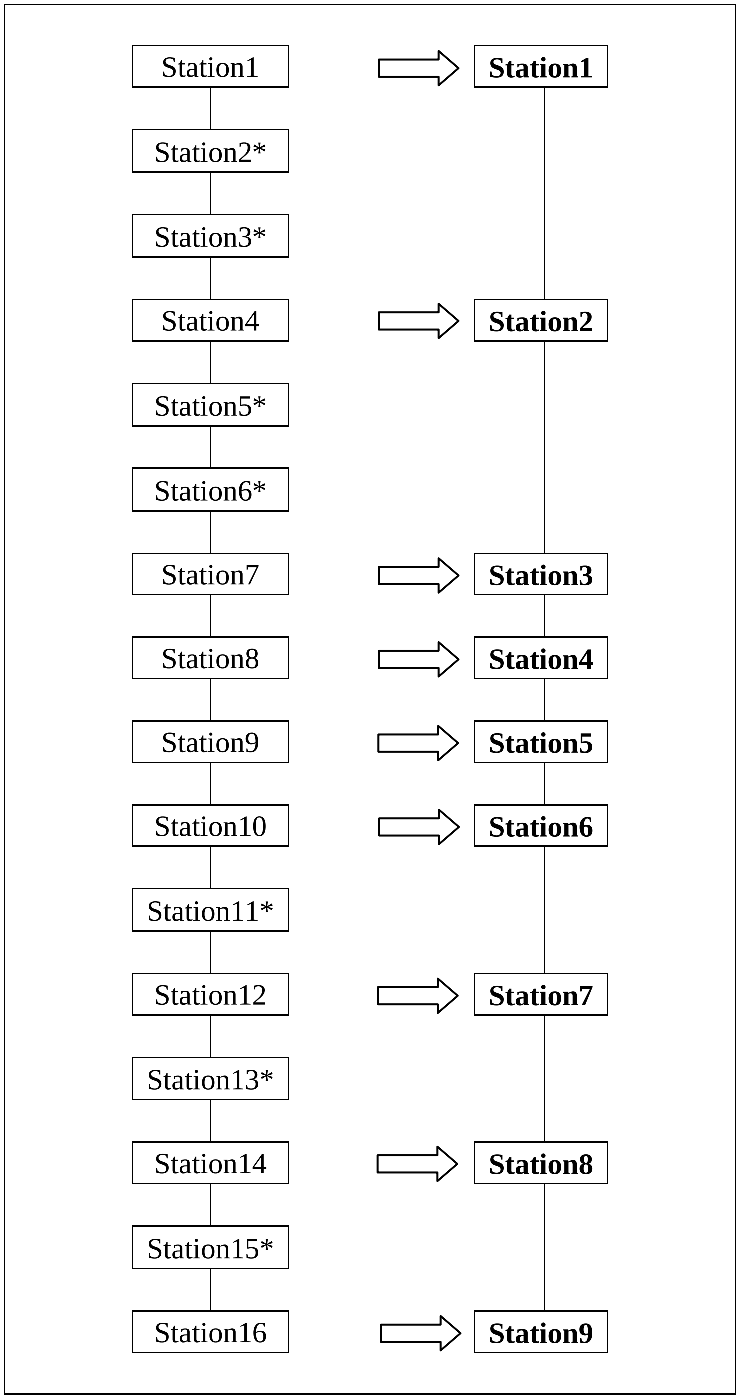
7.2.1 Optimization Framework and Procedure

The proposed RM optimization model accounts for both passengers' response to RM policy and for demand volumes. The passenger choice models, estimated with discrete choice methods, predict the timing in which passengers purchase the ticket as a function of fare and other trip attributes (Chapter 5). The demand functions account for passengers deciding not to travel with this service or for induced demand due to advantageous fare policy. Passenger volumes are estimated using log-linear regression, where independent attributes are fare and trip attributes (Section 6.2). The passenger choice models and the demand functions are then incorporated into a RM revenue optimization system in a network setting.

In this setting, it is assumed that the railway operator maximizes the ticket revenue of coach class ticket for each train trip from the south end to the north end station. The railway service in consideration departs hourly from 5:00 AM to 7:00 PM. In this analysis, we focus on the trains which depart from the south end station at four different departure times which are: 5:00 AM., 9:00 AM., 1:00 PM., and 4:00 PM. on Friday, March13, 2009. These four departure times (named Train#1 to Train#4) are selected to represent railway traffic at different time periods across the day.

7.2.2 Selected Stations

This railway network has total of 16 stations; however in this optimization problem, seven stations are excluded from the analysis because passenger demands in these markets are low and insufficient for estimating demand function. The remaining nine stations are shown in Figure 7.9 and are renumbered into station 1 to station 9. The optimization focuses on the selected stations; seats currently occupied by the excluded stations and seats currently empty (Y) are not allowed to contribute to the revenue. These seats are extracted from the total seat capacity (Z) in the constraint. The decision variables are: fare for each origin-destination pair on each booking day over the sale horizon $(fare_{od}^{j})$ and the acceptance ratio (α_{od}) which is the ratio of accepted demand over the total demand for each origin-destination pair.



* Indicates excluded stations.

Figure 7.9 Station renumbering

7.2.3 Demand Conversion

The demand functions in Section 6.2 are used in this optimization. The function provides daily passenger demand, however the proposed revenue optimization solves for each train trip which requires demand for each departure time as an input. We use conversion factors to convert daily demand to demand by departure time. The conversion factors are obtained from the historical data based on the distribution of daily passenger demand across different departure times of day. Our optimization focuses on the trains which depart from south end station (station 9) at four different departure times. Within our selected network, the corresponded departure time (t) for each station which loads passengers into these trains are shown in Table 7.6.

Table 7.6 Departure time for each origin

	Departure time (t)			
Origin	**Trian#1**	**Trian#2**	**Trian#3**	**Trian#4**
9	5:00AM	9:00AM	1:00PM	4:00PM
8	5:30AM	9:30AM	1:30PM	4:30PM
7	6:35AM	10:35AM	2:35PM	5:35PM
6	7:19AM	11:19AM	3:19PM	6:19PM
5	7:30AM	11:30AM	3:30PM	6:30PM
4	8:00AM	12:00PM	4:00PM	7:00PM
3	8:44AM	12:44PM	4:44PM	7:44PM
2	10:55AM	2:55PM	6:55PM	9:55PM

The conversion factor is denoted as f_{od}^{t} where *t* represents departure time, *o* represents origin station, and *d* represents destination station. The passenger demand by departure time can be computed from estimated demand function as follows:

$$D_{od}^{t} = f_{od}^{t} \times D_{od} \tag{7.10}$$

where

D_{od}^{t} = Number of passenger demand from origin *o* to destination *d* at departure time *t*.

f_{od}^{t} = Conversion factor from daily demand to demand by departure time.

D_{od} = Estimated passenger daily demand from origin *o* to destination *d* obtained from the demand function.

7.2.4 Problem Formulation

n = Number of stations (nine stations within the selected network).

o = Boarding station index.

d = Alighting station index.

D_{od}^{t} = Passenger demand from origin *o* to destination *d* at departure time *t*.

α_{od} = Acceptance ratio; a fraction of demand (D_{od}^{t}) to be accepted.

$fare_{od}^{j}$ = Fare for origin o destination d on booking day j.

Z = Total coach class seat capacity; equal to 260 (Railway Technology, 2011).

Y = Number of seats currently occupied by the excluded stations and seats currently empty.

$revenue$ = Revenue per train trip ($) from the selected stations in the network.

$\Pr(j|od)$ = Probability that passengers purchase the ticket on booking day j for origin o, destination d.

7.2.5 Revenue Optimization with MNL Choice Model

The problem formulation is shown as follows:

$$\max_{fare_{od}^{j},\alpha_{od}} revenue = \sum_{o=1}^{n-1} \sum_{d=o+1}^{n} \left[\alpha_{od} D_{od}^{t} \sum_{\mathrm{j}=1}^{31} \left\{ \Pr(j|od) fare_{od}^{j} \right\} \right] \tag{7.11}$$

Subject to

- Capacity constraint

$$\sum_{o=1}^{l} \sum_{d=l+1}^{n} \alpha_{od} D_{od}^{t} \leq Z - Y \tag{7.12}$$

$$0 \leq \alpha_{od} \leq 1 \tag{7.13}$$

for all $l = \{1, \ldots, n-1\}$

- Fare policy constraint

$$fare_{od}^{-} \leq fare_{od}^{j} \leq fare_{od}^{+} \tag{7.14}$$

The probability $\Pr(j|od)$ represents the share of passengers who purchase ticket on booking day j. The MNL choice model parameters estimated in Chapter 5 are used in this optimization. The probability is calculated based on the logit choice probability as:

$$\Pr(j|od) = \frac{\exp(V_j)}{\sum_{\mathrm{k}=1}^{31} \exp(V_k)} \tag{7.15}$$

where V_j is a deterministic utility of booking day j of origin o destination d. V_k is a deterministic utility of booking day k of origin o destination d.

Objective Function

The first two summations sum the passenger demand departing from origin $o \in \{1, \ldots, n-1\}$ to destination $d \in \{2, \ldots, n\}$ within the same train trip. Note that this index has different value from the station number. The value $o = 1$ represents south end station (station 9) and increases up to $o = 9$ for north end station (station 1). The index t is used to represent the corresponded departure time for each station which loads passengers into this train. The third summation represents weighted average of the ticket fare which is the product of fare ($fare_{od}^{j}$) and probability that the ticket be purchased on booking day j ($\Pr(j|od)$) over the sale horizon. The ratio of total demand which can be accepted for each origin-destination pair is denoted as α_{od}.

Constraint

The capacity constraint restricts the number of accepted passengers in each segment to be within the allowable seat capacity which is equal the total seat capacity (Z) subtracted by the seats currently occupied by the excluded stations and seats currently empty(Y). The decision variable α_{od} is an acceptance ratio used to control the number of accepted passengers to be within the allowable seat capacity. Figure 7.10 represents the capacity constraint where the line connecting each origin-destination pair represents the passenger demand. Fare policy constraint restricts fares to be within the allowable fare bounds. The fare bounds are obtained from the data based on the maximum and minimum of the average fare of a particular time of day, and day of week of departure within the entire month. These fare bounds are also adjusted to ensure that fare of shorter distance does not exceed fare of longer distance.

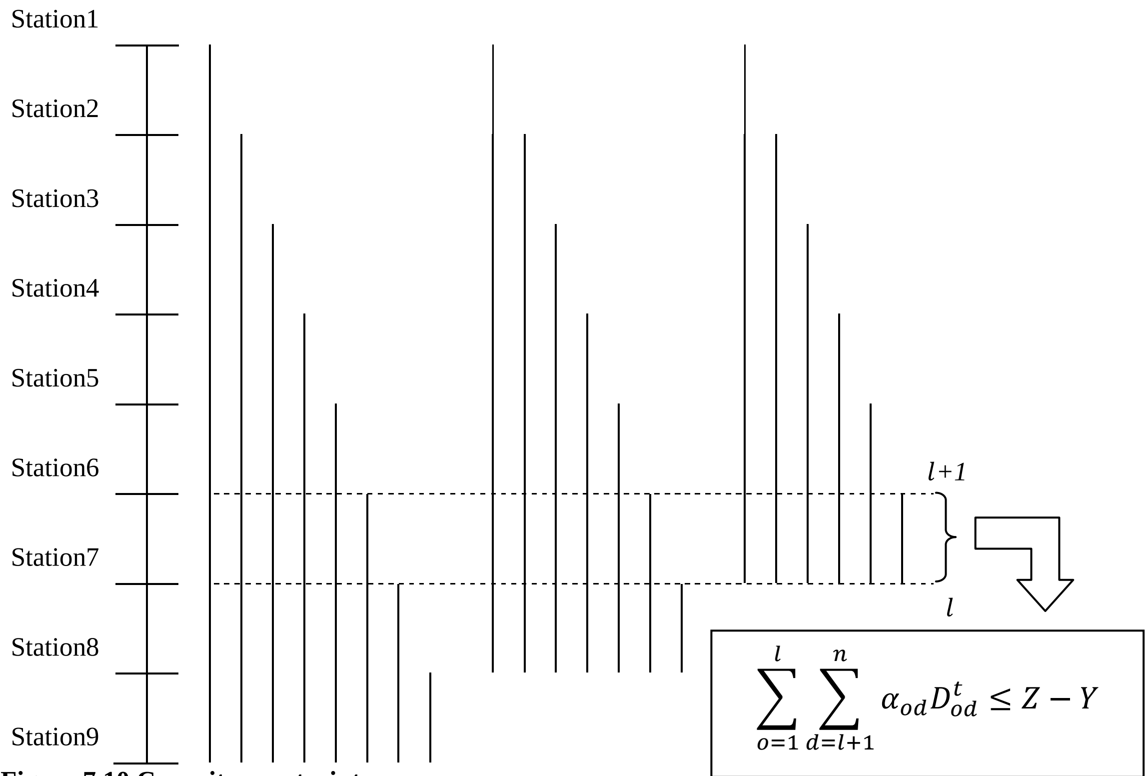


Figure 7.10 Capacity constraint

7.2.6 Revenue Optimization with LC Choice Model

The application of latent class (LC) choice model in the fare optimization allows for passenger taste heterogeneity in a discrete approach. Given the choice probability of LC model as:

$$\Pr(j|X_M, X_C) = \sum_{s=1}^{S} \Pr(s|X_M)\Pr(j|X_C, s) \quad ; \forall j \in C \tag{7.16}$$

where

s is class index; $\{1, \ldots, S\}$

X_M is class membership explanatory variable

X_C is class specific choice models explanatory variable

The corresponded optimization problem can be expressed as:

$$\max_{fare_{od}^{j},\alpha_{od}} \quad revenue$$

$$= \sum_{o=1}^{n-1} \sum_{d=o+1}^{n} \left[\alpha_{od} D_{od}^{t} \sum_{j=1}^{31} \left\{ \sum_{s=1}^{S} \Pr(s|X_M) \Pr(j|X_C, s, od)\, fare_{od}^{j} \right\} \right] \tag{7.17}$$

The LC choice model parameters estimated in Chapter 5 are used in this optimization. The latent class optimization has the same constraints as the MNL optimization.

7.2.7 Optimization Result

The optimization problem is solved as a nonlinear programming problem with LINGO 12.0, the optimization software by Lindo System Inc. (Lindo System Inc, 2010). The nonlinearity property of this problem is influenced by the exponential term in the logit choice probability function (MNL and LC). For brevity, we do not show results of all the decision variables obtained from the optimization. We chose certain categories of decision variables to show the overview of the result reported in Table 7.7 to Table 7.9.

Table 7.7 Acceptance ratio

(O,D)	5AM (Train#1)		9AM (Train#2)		1PM (Train#3)		4PM (Train#4)	
	MNL	LC	MNL	LC	MNL	LC	MNL	LC
(9,8)	1	1	1	1	0	0	0.275	0.285
(9,7)	1	1	1	1	0.861	0.861	1	1
(9,6)	0	0	1	1	0	0	1	1
(9,5)	0	0.010	0	0	0	0	1	1
(9,4)	0	0	0.054	0.048	0	0	0.195	0.186
(9,3)	0	0	0	0	0	0	0	0
(9,2)	0	0	0	0	0	0	0	0
(9,1)	0	0	0	0	0	0	0	0
(8,7)	0.839	0.336	0	1	1	1	1	1
(8,6)	0	0	0	1	1	1	1	1
(8,5)	0	1	1	1	1	1	1	1
(8,4)	0	0	0.958	0.827	0	0	0	0
(8,3)	0	0	0	0.636	0	0	0	0
(8,2)	0	0	0	0	0	0	0	0
(8,1)	0	0	0	0	0	0	0	0
(7,6)	0	0	0	0	1	1	1	1
(7,5)	1	1	1	1	1	1	1	1
(7,4)	0.331	0.322	0.866	0.891	0.714	0.714	0.587	0.591
(7,3)	0	0	0	0	0	0	0	0
(7,2)	0	0	0	0	0	0	0	0
(7,1)	1	1	0	0	0	0	0	0
(6,5)	1	1	1	1	0.464	0.468	1	1
(6,4)	1	1	0.184	0.174	0	0	0	0
(6,3)	1	1	1	1	0	0	0	0
(6,2)	1	1	0	0	0	0	0	0
(6,1)	1	1	0	0	0	0	0	0
(5,4)	0.954	1	0.366	0.201	0	0	0	0
(5,3)	1	1	1	1	0	0	0	0
(5,2)	0	0	0	0	0	0	0	0
(5,1)	0	0	0	0	0	0	0	0
(4,3)	1	1	1	1	0	0	1	1
(4,2)	0.479	0.511	0.411	0.462	0.348	0.352	0.388	0.497
(4,1)	0.295	0.302	0.347	0.370	0.227	0.226	0.316	0.340
(3,2)	0	0	0	0	0	0	0	0
(3,1)	0.374	0.432	0.600	0.718	0	0	0	0
(2,1)	1	1	1	1	1	1	1	1

Table 7.7 compares the acceptance ratio (α_{OD}) between the MNL and the LC; the acceptance policy obtained with the two models are similar.

Table 7.8 Accepted demand

(O,D)	5AM (Train#1)			9AM (Train#2)			1PM (Train#3)			4PM (Train#4)		
	Exist	MNL	LC	Exist	MNL	LC	Exist	MNL	LC	Exist	MNL	LC
(9,8)	0	1	1	9	4	4	0	0	0	3	1	2
(9,7)	0	9	9	7	63	63	11	49	49	53	89	89
(9,6)	0	0	0	0	13	12	5	0	0	39	22	22
(9,5)	0	0	0	10	0	0	10	0	0	22	9	9
(9,4)	9	0	0	65	21	22	21	0	0	72	81	80
(9,3)	1	0	0	1	0	0	0	0	0	10	0	0
(9,2)	0	0	0	4	0	0	2	0	0	1	0	0
(9,1)	1	0	0	5	0	0	0	0	0	2	0	0
(8,7)	0	3	1	3	0	2	6	8	8	1	11	12
(8,6)	0	0	0	0	0	0	1	3	3	2	4	4
(8,5)	0	0	2	0	1	1	1	4	4	1	5	5
(8,4)	1	0	0	13	14	11	11	0	0	13	0	0
(8,3)	0	0	0	2	0	1	0	0	0	4	0	0
(8,2)	0	0	0	0	0	0	0	0	0	0	0	0
(8,1)	0	0	0	1	0	0	1	0	0	1	0	0
(7,6)	0	0	0	0	0	0	5	5	5	3	5	3
(7,5)	1	4	4	1	4	3	5	5	5	2	4	4
(7,4)	25	39	37	17	79	81	38	84	84	22	61	62
(7,3)	3	0	0	1	0	0	0	0	0	5	0	0
(7,2)	0	0	0	1	0	0	0	0	0	2	0	0
(7,1)	2	0	0	11	0	0	4	0	0	2	0	0
(6,5)	0	0	0	0	3	3	0	3	3	0	0	0
(6,4)	0	0	0	0	7	7	0	0	0	0	0	0
(6,3)	0	0	0	0	3	3	0	0	0	0	0	0
(6,2)	0	0	0	0	0	0	1	0	0	1	0	0
(6,1)	0	0	0	0	0	0	2	0	0	2	0	0
(5,4)	1	8	9	0	2	1	0	0	0	0	0	0
(5,3)	0	6	7	0	4	5	0	0	0	0	0	0
(5,2)	1	0	0	1	0	0	0	0	0	1	0	0
(5,1)	9	0	0	9	0	0	4	0	0	4	0	0
(4,3)	1	4	4	0	4	4	0	0	0	1	3	3
(4,2)	24	34	34	21	34	35	44	50	51	27	32	32
(4,1)	45	43	42	49	59	58	81	89	88	59	71	71
(3,2)	4	0	0	3	0	0	0	0	0	0	0	0
(3,1)	1	10	10	3	13	14	0	0	0	1	0	0
(2,1)	2	5	5	1	6	7	2	5	6	2	2	2
Total	**131**	**166**	**166**	**238**	**334**	**336**	**255**	**306**	**306**	**358**	**401**	**399**

In Table 7.8 the number of accepted passengers $(\alpha_{od} \times D_{od}^{t})$ by the modeling system is compared to the actual demand. Results show that the proposed strategy increases the total number of passengers in all the four trains. Based on our assumption that limits the number of seats to those occupied in the actual situation, we can conclude that the optimal solution suggests accepting more short-haul passengers , thus occupying capacity in a shorter duration and allow for the same number of seats to serve more passengers.

Table 7.9 Revenue per train trip ($)

(O,D)	5AM (Train#1)			9AM (Train#2)			1PM (Train#3)			4PM (Train#4)		
	Exist	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**
(9,8)	0	28	9	396	190	64	0	0	0	171	86	31
(9,7)	0	1,013	994	764	6,907	6,888	1,433	7,326	7,326	7,521	13,310	13,310
(9,6)	0	0	0	0	1,882	1,686	894	0	0	7,244	4,317	4,317
(9,5)	0	0	1	1,704	0	0	1,808	0	0	4,433	1,863	1,849
(9,4)	1,203	0	0	9,589	3,709	3,606	3,575	0	0	14,529	19,035	18,810
(9,3)	137	0	0	183	0	0	0	0	0	1,959	0	0
(9,2)	0	0	0	332	0	0	396	0	0	222	0	0
(9,1)	149	0	0	694	0	0	0	0	0	422	0	0
(8,7)	0	287	37	244	0	48	690	1,123	1,085	132	1,492	1,145
(8,6)	0	0	0	0	0	43	205	655	655	390	871	770
(8,5)	0	0	194	0	99	72	212	733	731	191	975	943
(8,4)	95	0	0	1,717	2,186	1,698	2,038	0	0	2,693	0	0
(8,3)	0	0	0	212	0	62	0	0	0	827	0	0
(8,2)	0	0	0	0	0	0	0	0	0		0	0
(8,1)	0	0	0	124	0	0	223	0	0	223	0	0
(7,6)	0	0	0	0	0	0	566	658	658	354	584	250
(7,5)	86	456	395	114	357	298	715	688	688	272	611	491
(7,4)	2,064	4,225	4,086	1,860	8,663	8,879	5,025	12,461	12,461	2,790	9,088	9,153
(7,3)	273	0	0	152	0	0	0	0	0	760	0	0
(7,2)	0	0	0	107	0	0	0	0	0	386	0	0
(7,1)	274	0	0	1,449	0	0	847	0	0	389	0	0
(6,5)	0	0	0	0	125	84	0	199	200	0	0	0
(6,4)	0	0	0	0	334	278	0	0	0	0	0	0
(6,3)	0	0	0	0	290	272	0	0	0	0	0	0
(6,2)	0	0	0	0	0	0	140	0	0	124	0	0
(6,1)	0	0	0	0	0	0	281	0	0	298	0	0
(5,4)	32	336	342	0	83	35	0	0	0	0	0	0
(5,3)	0	488	479	0	313	224	0	0	0	0	0	0
(5,2)	93	0	0	124	0	0	0	0	0	140	0	0
(5,1)	795	0	0	978	0	0	520	0	0	523	0	0
(4,3)	71	220	162	0	259	190	0	0	0	71	246	202
(4,2)	2,366	3,396	3,361	2,263	3,429	3,362	5,293	6,747	6,830	3,061	4,320	4,251
(4,1)	4,781	4,416	4,331	5,430	6,103	5,896	10,629	12,364	12,278	7,117	9,865	9,794
(3,2)	396	0	0	370	0	0	0	0	0	0	0	0
(3,1)	104	978	956	343	1,299	1,350	0	0	0	119	0	0
(2,1)	70	181	174	25	227	135	90	243	278	75	100	57
Total	**12,989**	**16,023**	**15,520**	**29,174**	**36,456**	**35,169**	**35,580**	**43,197**	**43,191**	**57,436**	**66,761**	**65,372**
% Improve		**23.36**	**15.80**		**24.96**	**20.55**		**21.41**	**21.39**		**16.24**	**13.82**

Finally, Table 7.9 shows the revenue comparison per each train trip across four train departures. Results indicate that the RM strategy can improve revenues from low traffic trains (Train#1, 2, and 3 from 15.80% to 24.96% per train trip) significantly and from high traffic train but in a less remarkable entity (Train#4 from 13.82% to 16.24% per train trip). In comparison between the LC and the MNL, it was found that the optimization problems which incorporate the LC choice model generally provide less revenue. A possible explanation for this behavior can be derived from the fact that the LC models enable passengers in different classes to respond to fare differently. Thus, the behavior of the price sensitive passengers can be revealed more realistically with the LC than with the MNL which assumes homogeneity across populations. Based on the results obtained, two markets are selected and detailed analysis in terms of pricing and seat allocation are provided in the next section.

7.2.8 Pricing Result

In this section, we select two major markets which are (9,7)[7] and (4,1) to elaborate the result in detail. For these two markets, we consider the trip which loads passengers into Train#2 and Train#4 that depart from south end station at 9AM and 4PM respectively. These correspond to the departure times of 9AM and 4PM for market (9,7) and departure times of 12PM and 7PM for market (4,1) based on Table 7.6.

Fare of market (9,7) is shown in Figure 7.11 and Figure 7.13 for 9AM and 4PM departure time respectively. The results from the MNL and the LC models are compared to the existing fare, which is representative fare pattern of a particular departure time of day and day of week obtained from the data. In general, the optimization with the LC model provides stepwise fare pattern which realistically reflects passenger price

[7] The term denotes (origin, destination) pair eg. (9,7) represents trip from Station 9 to Station7.

sensitivity across the sale horizon. The use of the LC choice model also enables to distinguish between leisure and business travelers, depending on departure time and day of week. When compared across different departure times, the fares of market (9,7) are higher in the 4PM departure compared to 9AM departure. This is primarily due to the difference in fare bounds across different departure times imposed in the constraints as well as differences in passenger price sensitivity across different departure times. It is expected that passengers departing 4PM which has higher traffic be less price sensitive than the 9AM departure.

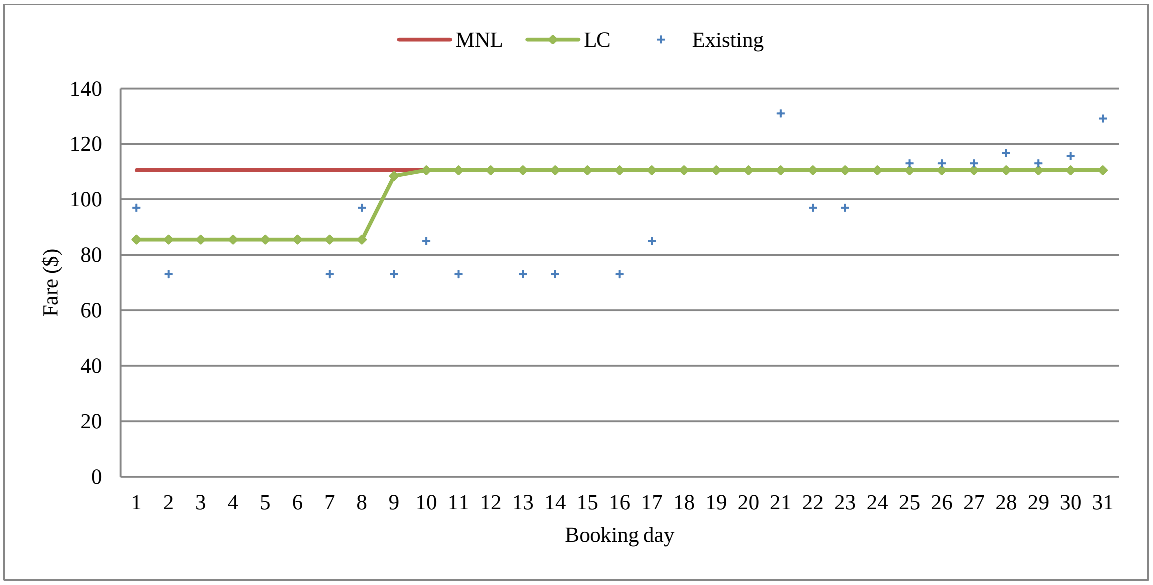


Figure 7.11 Station (9,7) 9AM departure fare

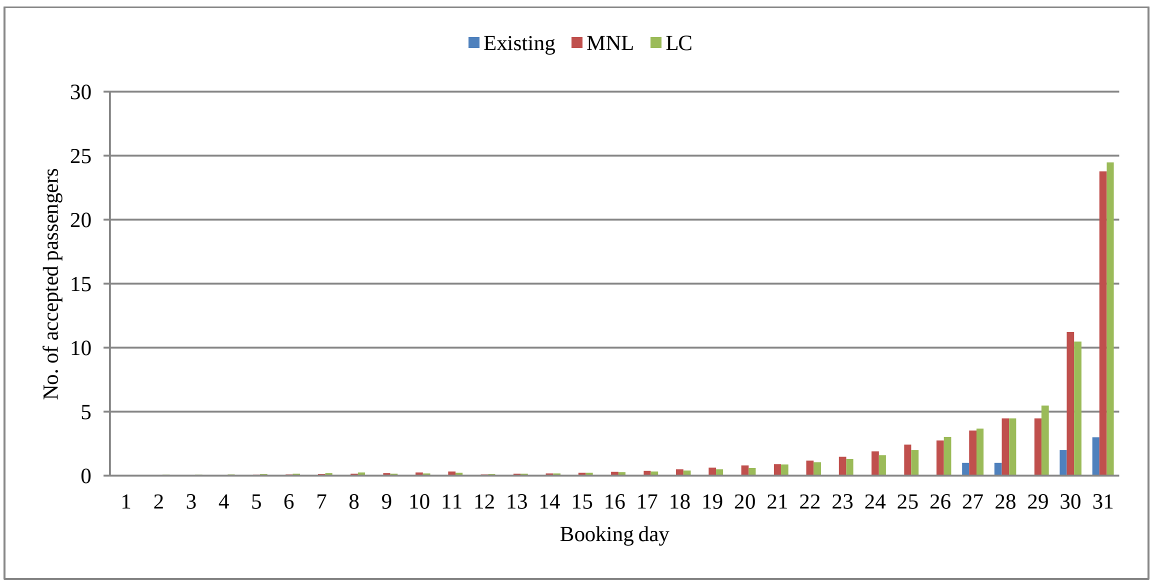


Figure 7.12 Station (9,7) 9AM departure demand

Figure 7.12 shows the corresponded number of accepted passengers in market (9,7) of 9AM departure. The response of passenger demand in both the MNL and the LC are realistic; when the new fare is lower than the existing (from booking day 25 onward), the passenger demand increase from the existing significantly. Figure 7.14 also shows the same pattern for 4PM departure, with the number of passengers increasing significantly on booking day 30 and 31 when fares are lower than the existing.

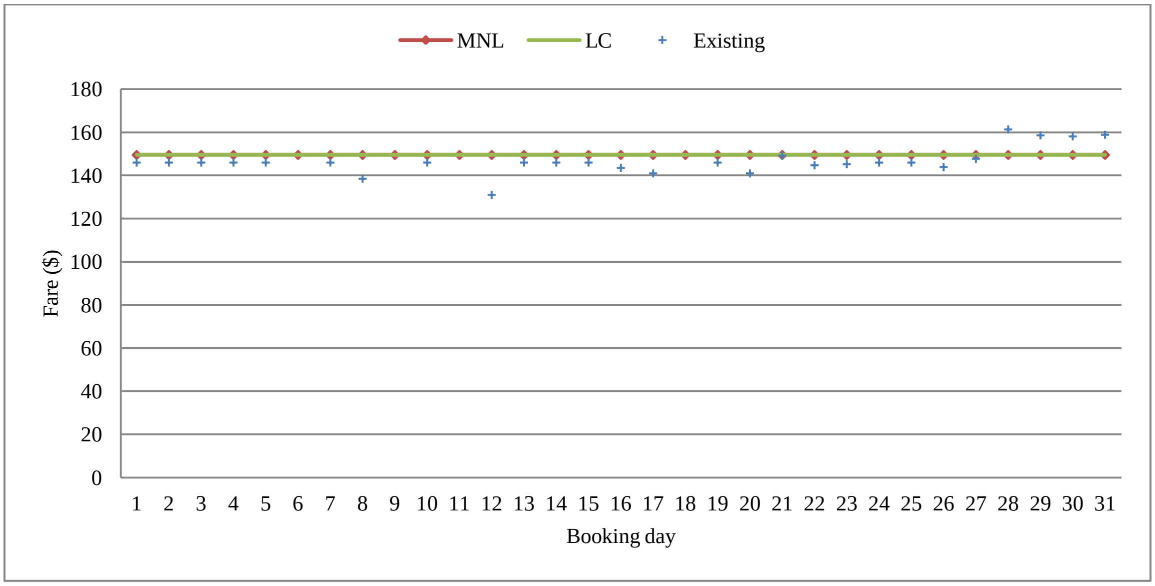


Figure 7.13 Station (9,7) 4PM departure fare

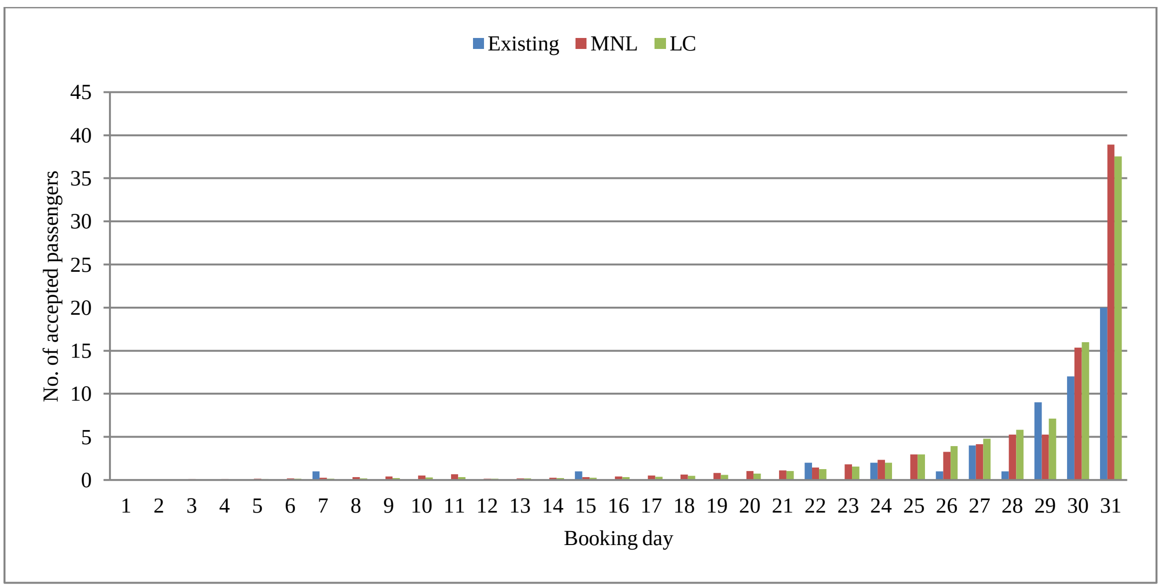


Figure 7.14 Station (9,7) 4PM departure demand

The results of market (4,1) are shown in Figure 7.15 to Figure 7.18. Results show similar pattern to market (9,7) with the LC model providing stepwise fare pattern. In market (4,1), the 7PM departure has a higher traffic than 12PM departure and passengers are expected to be less price sensitive. The passenger demand responses to fare are realistic following the pattern observed in market (9,7).

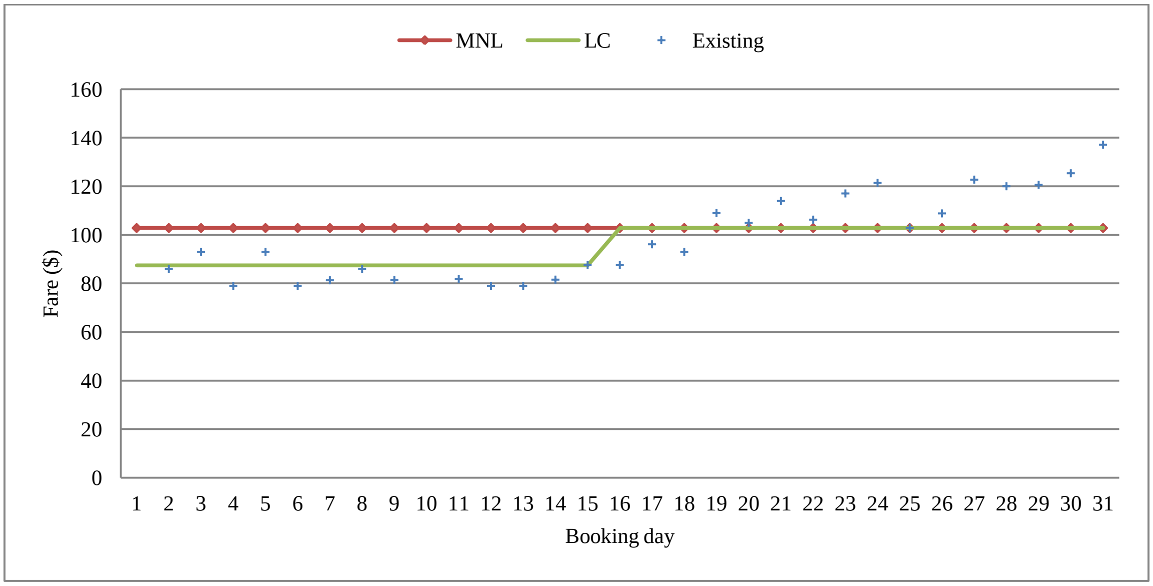


Figure 7.15 Station (4,1) 12PM departure fare

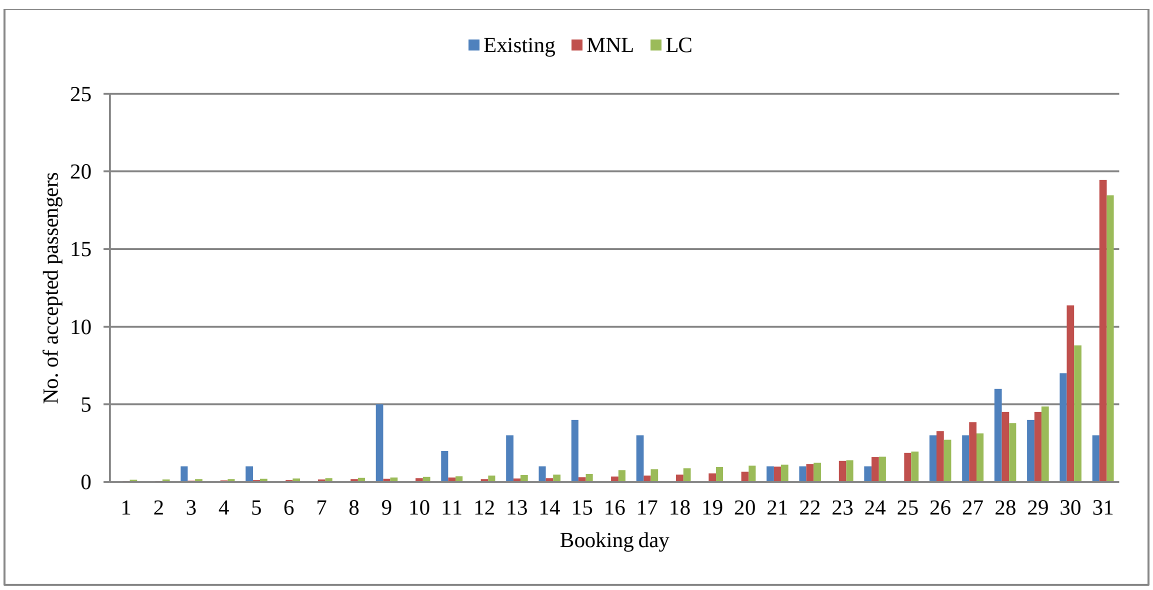


Figure 7.16 Station (4,1) 12PM departure demand

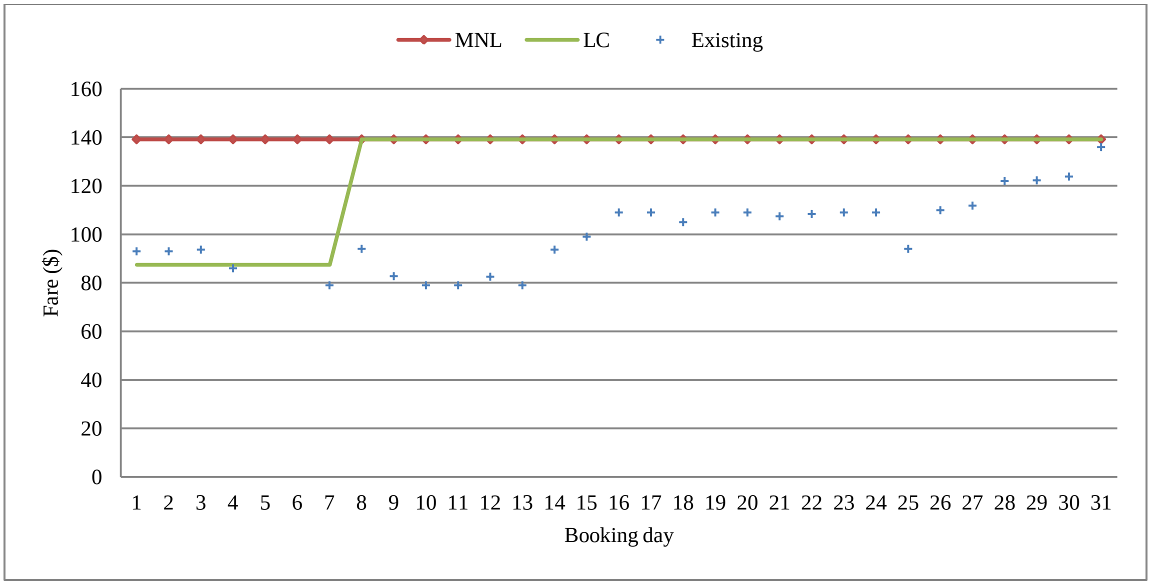


Figure 7.17 Station (4,1) 7PM departure fare

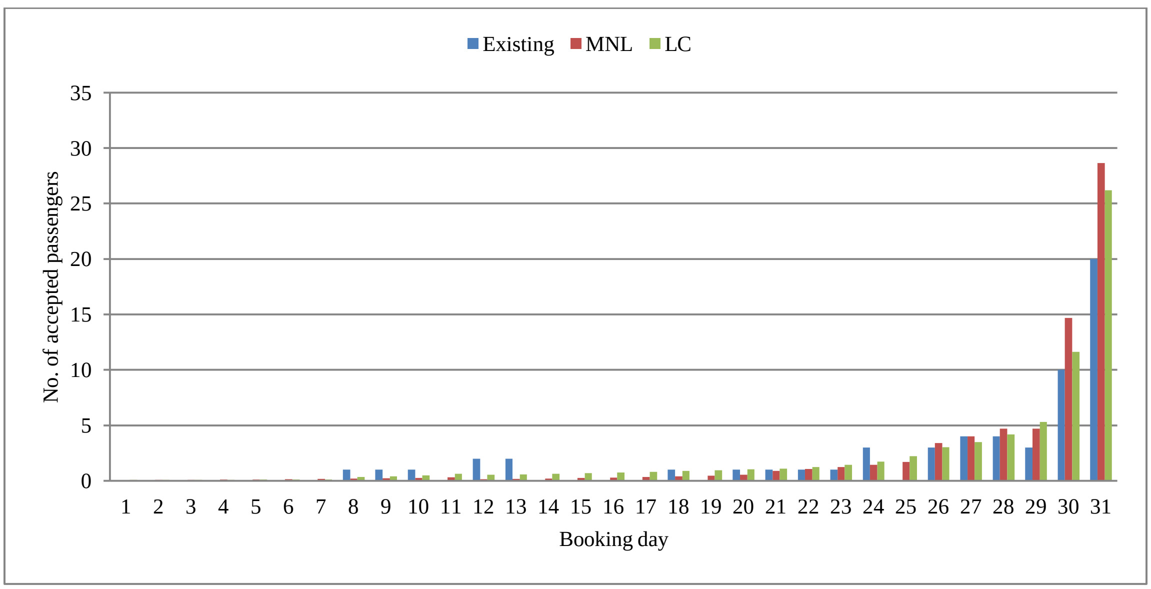


Figure 7.18 Station (4,1) 7PM departure demand

Based on these two markets, the total accepted passengers and corresponded revenue across different departure times of focused are shown in Figure 7.19 and Figure 7.20 respectively.

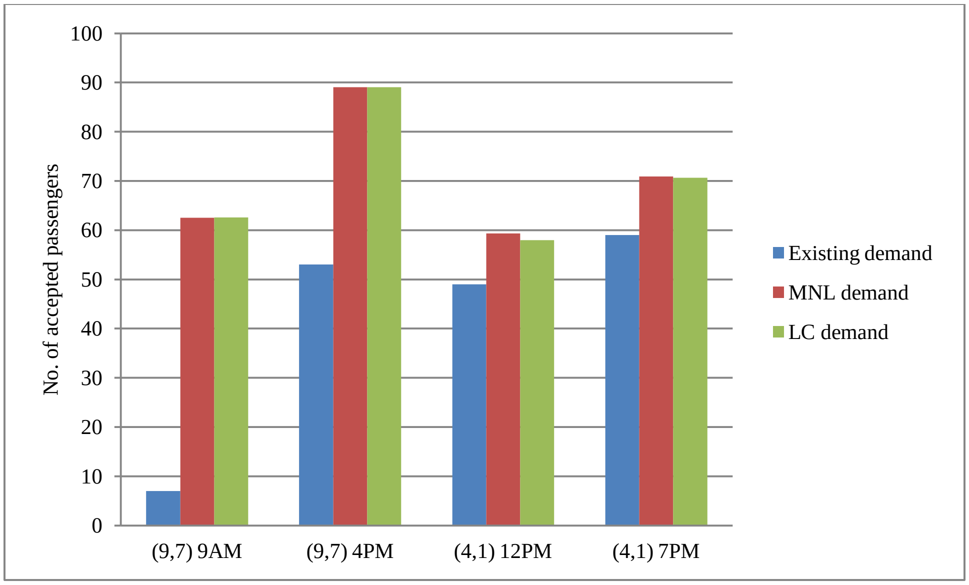


Figure 7.19 Total accepted passengers of major markets

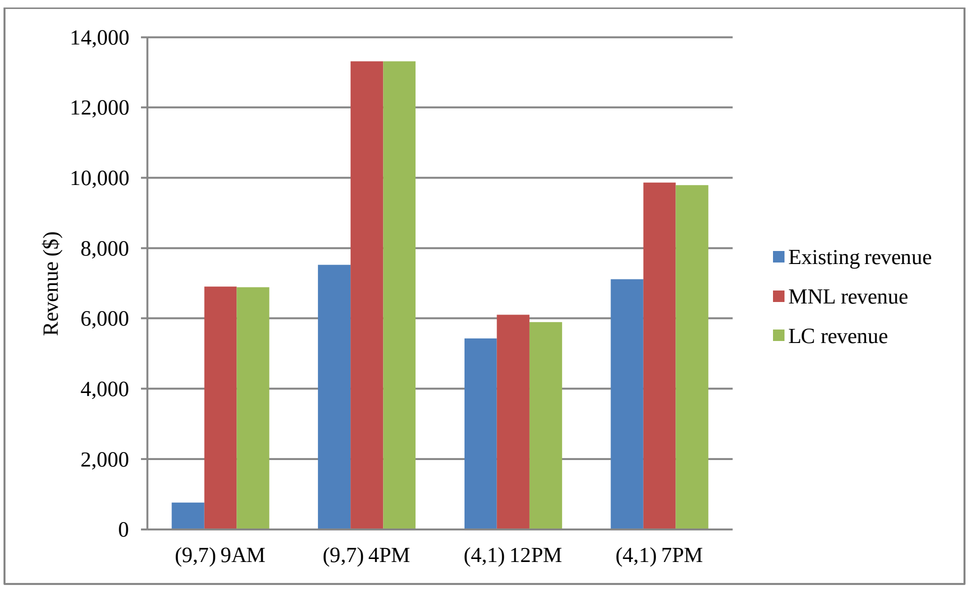


Figure 7.20 Revenue ($) per train trip of major markets

7.2.9 Capacity Redistribution

This section illustrates how the proposed seat allocation strategy influences capacity distribution of each segment in the network. For brevity, we shall focus the analysis on Train#4 which has the maximum traffic across the four departure times. Table 7.10

represents the share (percentage) of each leg capacity utilized by its related markets (highlighted in grey). The allowable capacity of each leg is shown in the last row of Table 7.10. Result shows that the proposed shares of capacity increase slightly for markets (9,4), (4,2), and (4,1) and significantly for markets (9,7) and (7,4). On the other hand, the proposed shares of capacity for markets (9,6) and (9,5) decrease. This occurs across all the legs these markets utilize.

Table 7.10 Capacity redistribution of Train#4

	Leg 9-8 (%)			Leg 8-7 (%)			Leg 7-6 (%)			Leg 6-5 (%)			Leg 5-4 (%)			Leg 4-3 (%)			Leg 3-2 (%)			Leg 2-1 (%)		
OD	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**
(9,8)	1	1	1																					
(9,7)	26	44	44	24	40	40																		
(9,6)	19	11	11	18	10	10	19	12	12															
(9,5)	11	4	4	10	4	4	11	5	5	14	5	5												
(9,4)	36	40	40	33	37	36	35	42	43	44	50	50	51	57	57									
(9,3)	5	0	0	5	0	0	5	0	0	6	0	0	7	0	0	8	0	0						
(9,2)	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0			
(9,1)	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	2	0	0	2	0	0	3	0	0
(8,7)				0	5	5																		
(8,6)				1	2	2	1	2	2															
(8,5)				0	2	2	0	2	3	1	3	3												
(8,4)				6	0	0	6	0	0	8	0	0	9	0	0									
(8,3)				2	0	0	2	0	0	2	0	0	3	0	0	3	0	0						
(8,2)				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
(8,1)				0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
(7,6)							1	2	1															
(7,5)							1	2	2	1	3	2												
(7,4)							11	32	33	14	38	39	15	43	43									
(7,3)							2	0	0	3	0	0	4	0	0	4	0	0						
(7,2)							1	0	0	1	0	0	1	0	0	2	0	0	2	0	0			
(7,1)							1	0	0	1	0	0	1	0	0	2	0	0	2	0	0	3	0	0

	Leg 9-8 (%)			Leg 8-7 (%)			Leg 7-6 (%)			Leg 6-5 (%)			Leg 5-4 (%)			Leg 4-3 (%)			Leg 3-2 (%)			Leg 2-1 (%)		
OD	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**	**Exist**	**MNL**	**LC**
(6,5)										0	0	0												
(6,4)										0	0	0	0	0	0									
(6,3)										0	0	0	0	0	0	0	0	0						
(6,2)										1	0	0	1	0	0	1	0	0	1	0	0			
(6,1)										1	0	0	1	0	0	2	0	0	2	0	0	3	0	0
(5,4)													0	0	0									
(5,3)													0	0	0	0	0	0						
(5,2)													1	0	0	1	0	0	1	0	0			
(5,1)													3	0	0	3	0	0	4	0	0	5	0	0
(4,3)																1	3	3						
(4,2)																22	30	31	26	31	31			
(4,1)																48	67	67	57	69	69	81	97	97
(3,2)																			0	0	0			
(3,1)																			1	0	0	1	0	0
(2,1)																						3	3	3
%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Total	202	202	202	221	221	221	203	191	188	162	160	159	142	142	142	122	106	106	103	103	103	73	73	73

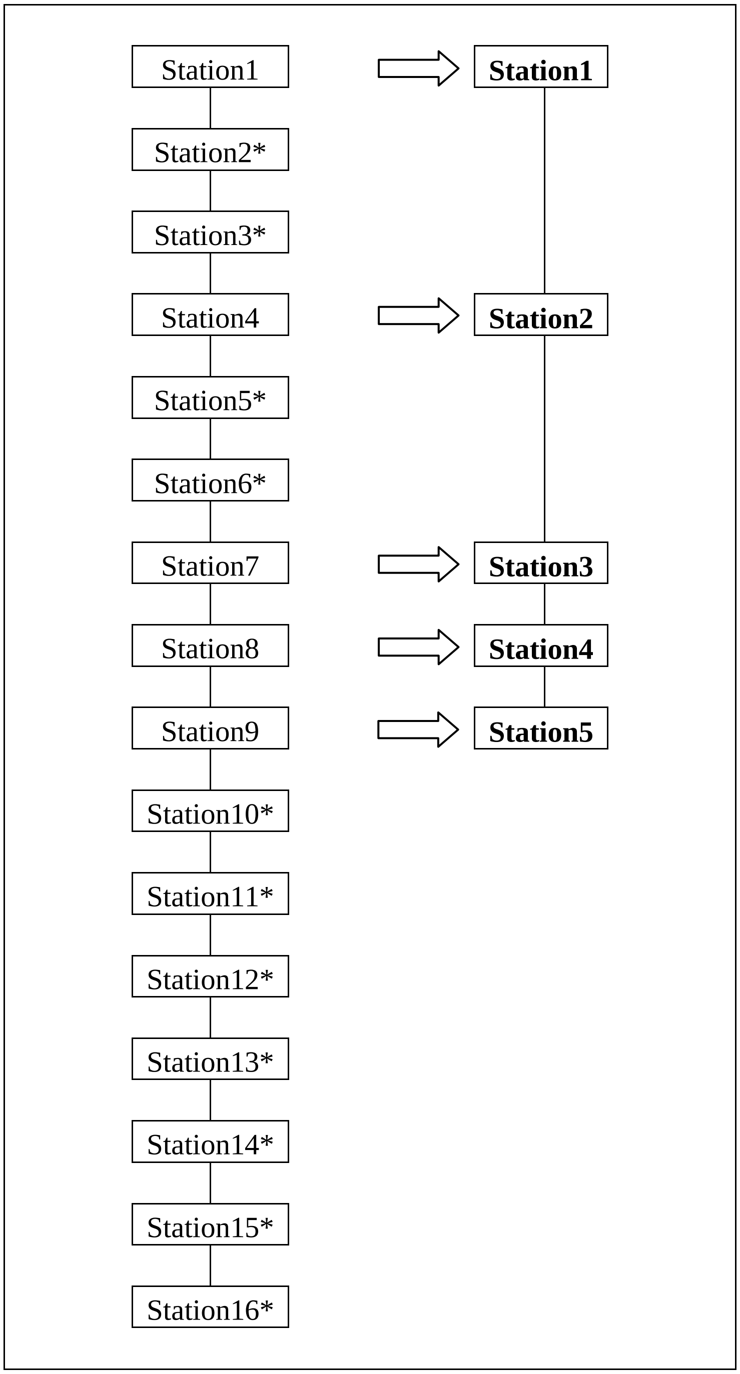
In addition, we compare the cumulated leg-based capacity consumption of Train#4 over the sale horizon in Appendix B between the existing condition and the strategy from optimizations with the MNL and the LC models.

7.3 Network Revenue Optimization with Continuous Market Segmentation

The primary purpose of this section is to incorporate mixed logit passenger choice models which account for taste heterogeneity based on continuous segmentation. Given that mixed logit model with parametric distribution is consistent and applicable only in the case that the distribution of the unknown random coefficient can be reasonably assumed, this section introduces mixed logit with non-parametric distribution (Section 7.3.2) to resolve difficulties associated with the identification of underlying unknown random distribution. With this approach, B-spline curve is proposed to capture randomness and heterogeneity presented in the population. The method adopts the model estimation framework of Bastin et al. (2010) where the method can explicitly estimate the shape of the unknown distributions.

7.3.1 Selected Stations

In this section, the problem setting and optimization procedure has the same structure as Section 7.2 where the choice models in Chapter 5 and demand functions in Section 6.2 are incorporated. However, the network considered in this section consists of five stations depicted in Figure 7.21 to allow for detailed analysis. With this selected network, the optimization objective is to maximize ticket revenue obtained from these five stations per each train trip.



* Indicates excluded stations.

Figure 7.21 Station renumbering

7.3.2 Non-parametric Mixed Logit Model

In this section, a mixed logit model with non-parametric distribution is proposed. We assume that the distribution of each random coefficient in the mixed logit function is random, and if we assume the independence between these values, each value can be considered separately. This allows us to draw from the univariate random variables. More specifically, if a random variable X has a bounded support, the elegant way to construct the non-parametric distribution is the use of B-spline functions. The bounded support is considered consistent with the underlying behavior assumption (Bastin et al., 2010). We propose the B-spline approximation of degree three given by:

$$F_X^{-1}(u) \approx C(u) = \sum_{i=0}^{n} \pi_i \, N_{i,3}(u), \tag{7.18}$$

where the piecewise cubic polynomial functions $N_{i,3}(u)$ form an easily computable basis for B-spline function in [0,1]. The coefficients π_i is called the control point, and the functions $N_{i,3}(u)$ depend on the choice of special points, called knots in [0,1]. With these basis and knots choice, $C(u)$ is monotonically increasing if the control points have the same property, that is $\pi_0 \leq \pi_1 \leq \pi_n$.

With non-parametric distribution, the mixed logit model specifies price sensitivity $(\tilde{\beta}_{fare})$ with a uniform B-spline curve. The utility function corresponded to mixed logit choice model in Chapter 5 has the form:

$$U_i(j,k) = \left(\beta_{advbk} \times advbk_j\right) + \left(\tilde{\beta}_{fare} \times fare_j\right) + \left(\beta_{wknd}^{k} \times wknd_i\right) + \varepsilon_{ij} \tag{7.19}$$

where the price coefficient is estimated with a B-spline approximation of degree three given by:

$$\tilde{\beta}_{fare} \approx C(u) = \sum_{i=0}^{n} \pi_i \, N_{i,3}(u) \tag{7.20}$$

The constrained optimization in Bastin et al. (2010) is adopted to estimate the knots point (π_i) of the non-parametric distribution and to ensure its monotonicity. For price coefficient, seven control points $(\pi_1, \pi_2, \dots, \pi_7)$ have been estimated for each B-spline, where π_1 and π_7 give the bounds of the distribution, and the knot vector is defined on the percentile 0, 0.25, 0.5, and 1. An example of the estimation result for one of the models incorporated in the revenue optimization (model 5 based on Chapter 5) is shown in Table 7.11.The estimation results of all 10 models are shown in Appendix C.

Table 7.11 Passenger choice model with B-spline

MNL				ML			B-spline				LC						
											Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat			Est	T-Stat		Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.220	17.952	*	-0.450	89.143	*	advbk	-0.568	72.854	*	advbk	-0.199	-11.688	*	-0.539	-12.988	*
price.period1	-0.008	1.942		-6.225	54.812	*	price1	-0.406	47.219	*	price.period1	-0.083	-10.519	*	0.023	1.497	
price.period2	-0.017	6.735	*	(3.630)	56.304	*	price2	-0.406	37.111	*	price.period2	-0.084	-11.562	*	-0.010	-0.746	
price.period3	-0.015	8.598	*				price3	-0.406	27.721	*	price.period3	-0.084	-12.478	*	-0.012	-0.950	
price.period4	-0.013	9.154	*				price4	-0.406	11.498	*	price.period4	-0.076	-11.299	*	-0.024	-2.042	*
price.period5	-0.004	3.109	*				price5	4.569	7.813	*	price.period5	-0.106	-6.418	*	-0.015	-1.235	
price.period6	-0.002	1.533					price6	6.698	15.579	*	price.period6	-0.059	-8.864	*	-0.035	-2.974	*
							price7	6.698	15.578	*							
											Class Model	**Class1**			**Class2**		
											Class Size	0.575			0.426		
											Variable	Est	T-Stat		Est	T-Stat	
											Intercept	0.785	8.454	*	-0.785	-8.454	*
											Monday	0.186	2.215	*	-0.186	-2.215	*
wknd.period1	1.473	8.601	*	0.468	12.337	*		0.176	1.361		Tuesday	-0.408	-5.255	*	0.408	5.255	*
wknd.period2	0.833	5.862	*	0.491	42.742	*		0.153	1.206		Wednesday	-0.413	-5.272	*	0.413	5.272	*
wknd.period3	0.396	3.485	*	0.518	36.889	*		0.245	1.396		Thursday	-0.456	-5.764	*	0.456	5.764	*
wknd.period4	0.352	4.280	*	0.355	6.400	*		-0.495	4.276	*	Friday	-0.317	-3.977	*	0.317	3.977	*
wknd.period5	-0.551	5.574	*	-0.163	3.125	*		-0.603	5.292	*	Saturday	0.180	1.163		-0.180	-1.163	
wknd.period6	0.498	7.342	*	1.332	23.902	*		0.022	0.177								
											Early morning						
											AM peak	-0.917	-11.348	*	0.917	11.348	*
											AM off peak	-0.305	-5.495	*	0.305	5.495	*
											PM off peak	-0.009	-0.182		0.009	0.182	
											PM peak	-0.190	-3.484	*	0.190	3.484	*
No. of observations		11,536			11,536				11,536		No. of observations					11,536	
Rho-squared:		0.4327			0.4525				0.6736		Rho-squared:					0.4444	
Adjusted rho-squared:		**0.4323**			**0.4523**				**0.6733**		**Adjusted rho-squared:**					**0.4437**	
Log-likelihood at optimal		-22,474			-21,688				-12,929		Log-likelihood at optimal					-22,011	
Log-likelihood at zero		-39,614			-39,614				-39,614		Log-likelihood at zero					-39,614	
Log-likelihood at constant		-22,530			-22,530				-22,530		Log-likelihood at constant					-22,530	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

7.3.3 Revenue Optimization with ML Choice Model

Mixed logit (ML) probabilities are the integral of standard logit probabilities over a density of parameters (β). Choice probabilities of a mixed logit model can be expressed in the form:

$$P_{ni}(\beta) = \int L_{ni}\,(\beta) f(\beta) d\beta \tag{7.21}$$

where $L_{ni}(\beta)$ is the logit probability evaluated at parameter β:

$$L_{ni}(\beta) \; = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}} \tag{7.22}$$

and $f(\beta)$ is a density function. $V_{ni}(\beta)$ is deterministic term observed by the analyst, which depends on the parameters β. Usually, the utility is linear in β, thus $V_{ni}(\beta) = \beta' x_{ni}$. The mixed logit probability then takes its usual form:

$$P_{ni} = \int \left(\frac{\exp(\beta' x_{ni})}{\sum_{J} \exp(\beta' x_{nj})} \right) f(\beta) d(\beta) \tag{7.23}$$

The application of mixed logit choice model in the revenue optimization allows for passenger taste heterogeneity in a continuous approach. Given the choice probability of mixed logit specified in the choice model as:

$$\Pr(j) = \int \left(\frac{\exp\left(\beta' x_{nj}\right)}{\sum_{\mathrm{K}} \exp(\beta' x_{nk})} \right) f(\beta) d(\beta) \tag{7.24}$$

The corresponded optimization problem could be expressed as:

$$\max_{fare_{od}^{j},\alpha_{od}} \quad revenue$$

$$= \sum_{o=1}^{n-1} \sum_{d=o+1}^{n} \left[\alpha_{od} D_{od}^{t} \sum_{\text{j}=1}^{31} \int \left(\frac{\exp(\beta' x_{nj})}{\sum_{\text{K}} \exp(\beta' x_{nk})} \right) f(\beta) d(\beta) fare_{od}^{j} \right] \tag{7.25}$$

The optimization is subjected to the same set of constraints described in Section 7.2 except that only five stations are considered instead of nine stations.

7.3.4 Incorporating Random Coefficient Parameters in the Optimization

This section describes how the random coefficient parameters estimated from mixed logit models in Chapter 5 and Section 7.3.2 are incorporated into the optimization framework.

A. Random coefficient with parametric distribution

The mixed logit choice model with parametric distribution estimated in Chapter 5 assumes that price coefficient is log-normally distributed. Based on the estimation method for mixed logit model in Chapter 5 using AMLET (Bastin, 2011), the lognormal distribution is specified as a transformation of the normal distribution given by $\exp(\mu + \sigma X)$ where X is assumed to be normally distributed based on standard normal distribution. To incorporate the random coefficient parameters into revenue optimization, we discretize the price sensitivity (β_{price}) into six segments based on the distribution of X. Given X is distributed as a standard normal, Figure 7.22 shows the probability of X lying in each of the six regions where the cut points have increment of σ. The averaged price sensitivity of each segment (β_{fare}^{i}) can then be written as $\exp(\mu + \sigma X_i)$ where X_i is the mean value of X within region i calculated from the probability density function of the standard normal distribution. The averaged price sensitivity obtained is further truncated at [-1.0, 0.5] to exclude extreme behaviors. This approach approximates the random coefficient parameters from mixed logit by segmenting price sensitivity (β_{fare}) into six segments and assigning each value to the population based on the probability distribution in Figure 7.22.

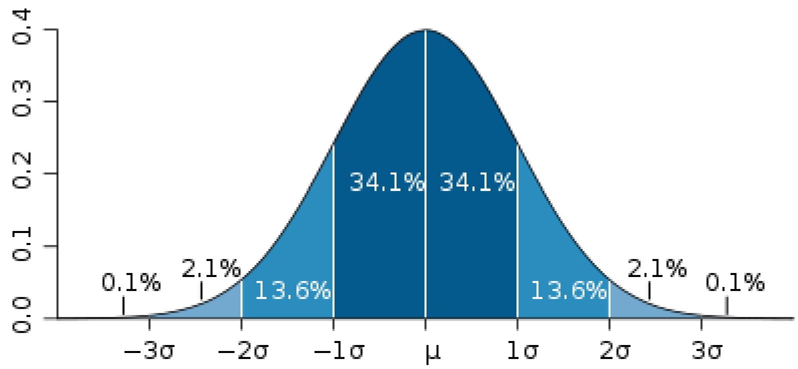


Figure 7.22 Standard normal distribution

B. Random coefficient with non-parametric distribution

Based on the estimation of mixed logit model with non-parametric B-spline mentioned in Section 7.3.2, the control points estimated are truncated at [-1.0, 0.5] to exclude extreme behaviors which is not usually welcome because they are difficult to interpret. These control points are used to construct spline curves using De Boor's algorithm (Lee, 1982) based on polynomial of degree three. Examples of the spline curves estimated are shown in Figure 7.23.

To incorporate random coefficient into revenue optimization, we discretize the spline curve into seven regions equally based on the x axis. For each region i, the averaged price sensitivity (β_{fare}^{i}) is obtained by drawing randomly from the distribution based on the shape of the spline curve within the region. The probability mass of each averaged price sensitivities are equally 1/7 since the spline region is equally divided on x axis. This approach approximates the random coefficient parameters from mixed logit by segmenting price sensitivity (β_{fare}) into seven segments and assigning each value to the population with the probability of 1/7 each.

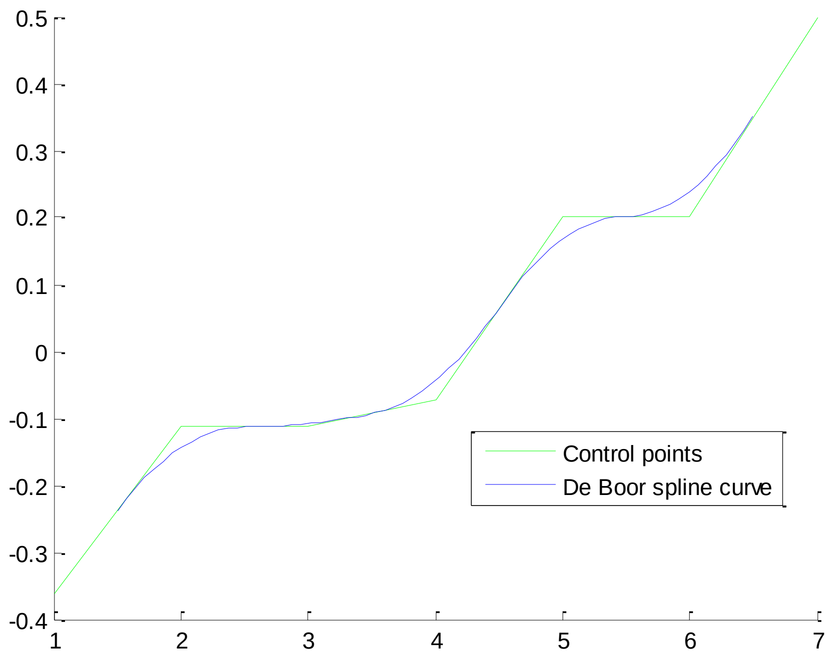


Figure 7.23 Example of B-spline curve estimated with De Boor's algorithm

7.3.5 Optimization Result

The optimization problem is solved as a nonlinear programming problem with LINGO 12.0, the optimization software by Lindo System Inc. (Lindo System Inc, 2010). The nonlinearity property of this problem is influenced by the exponential function of the logit choice probability. For brevity, we do not show results of all the decision variables obtained from the optimization. We chose certain categories of decision variables to show the overview of the result reported in Table 7.12 to Table 7.14.

Table 7.12 Acceptance ratio

(O,D)	5AM (Train#1)				9AM (Train#2)				1PM (Train#3)				4PM (Train#4)			
	MNL	LC	ML	Spline	MNL	LC	ML	Spline	MNL	LC	ML	Spline	MNL	LC	ML	Spline
(5,4)	0.544	0.559	1	0.560	1	1	1	0.958	1	1	1	1	0.926	0.999	0.865	1
(5,3)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0.587
(5,2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(5,1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(4,3)	1	1	0.745	1	1	1	0.271	1	1	0	1	1	1	1	1	0
(4,2)	0.381	0.452	0.470	0.455	0.324	0.328	0.375	0.334	0.319	0.332	0.323	0.339	0.321	0.343	0.324	0.343
(4,1)	0.295	0.261	0.390	0.262	0.197	0.197	0.347	0.200	0.197	0.208	0.196	0.196	0.268	0.262	0.267	0.294
(3,2)	1	0	1	0	0	0	1	0	1	0	1	0	1	0	1	1
(3,1)	0.338	0.521	1	0.517	0.608	0.596	1	0.702	0.656	0	0.666	0.900	0.403	0.613	0.426	0.056
(2,1)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 7.13 Accepted demand

(O,D)	5AM (Train#1)					9AM (Train#2)					1PM (Train#3)					4PM (Train#4)				
	Exist	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**
(5,4)	1	5	5	3	5	0	6	6	3	5	0	1	1	1	1	0	3	3	3	3
(5,3)	0	6	6	3	6	0	4	4	2	5	0	1	0	1	1	0	2	2	2	2
(5,2)	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
(5,1)	9	0	0	0	0	9	0	0	0	0	4	0	0	0	0	4	0	0	0	0
(4,3)	1	4	4	2	4	0	4	4	1	4	0	5	0	5	5	1	3	3	3	0
(4,2)	24	27	32	28	32	21	30	31	25	31	44	46	48	46	48	27	27	28	27	27
(4,1)	45	43	38	47	38	49	41	41	52	41	81	77	81	77	75	59	60	59	60	63
(3,2)	4	5	0	1	0	3	0	0	1	0	0	1	0	1	0	0	2	0	2	1
(3,1)	1	9	14	7	14	3	14	14	8	15	0	5	0	5	6	1	4	5	4	0
(2,1)	2	5	5	2	5	1	6	7	2	7	2	5	6	6	6	2	2	2	2	2
Total	**88**	**104**	**104**	**94**	**104**	**87**	**106**	**106**	**95**	**107**	**131**	**141**	**136**	**142**	**142**	**95**	**102**	**102**	**102**	**99**

Table 7.12 shows acceptance ratio (α_{od}) across four departure times. This acceptance ratio results in the number of accepted passengers ($\alpha_{od} \times D_{od}^{t}$) shown in Table 7.13. Result shows that the proposed strategy suggests increasing total accepted passengers in all the four trains. Based on our assumption that the solution cannot utilize more seats than currently occupied in the existing condition, this means that the solution suggests accepting more short-haul passengers which occupy capacity in a shorter duration thus allowing for the same number of seats to serve more passengers.

Table 7.14 Revenue per train trip ($)

(O,D)	5AM (Train#1)					9AM (Train#2)					1PM (Train#3)					4PM (Train#4)				
	Exist	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**	**Exist**	**MNL**	**LC**	**ML**	**Spline**
(5,4)	32	191	197	138	202	0	255	254	153	204	0	60	59	62	47	0	156	170	145	103
(5,3)	0	488	480	193	474	0	343	325	211	282	0	79	0	85	45	0	228	205	249	87
(5,2)	93	0	0	0	0	124	0	0	0	0	0	0	0	0	0	140	0	0	0	0
(5,1)	795	0	0	0	0	978	0	0	0	0	520	0	0	0	0	523	0	0	0	0
(4,3)	71	220	221	136	221	0	289	288	56	201	0	428	0	432	308	71	246	244	248	0
(4,2)	2,366	2,703	3,206	2,771	3,214	2,263	3,373	3,416	2,758	3,411	5,293	6,182	6,426	6,228	6,409	3,061	3,575	3,816	3,590	3,645
(4,1)	4,781	4,413	3,905	4,848	3,910	5,430	4,757	4,743	5,934	4,668	10,629	10,739	11,304	10,614	10,479	7,117	8,388	8,175	8,296	8,805
(3,2)	396	481	0	132	0	370	0	0	154	0	0	161	0	161	0	0	203	0	202	96
(3,1)	104	882	1,360	719	1,347	343	1,573	1,543	859	1,416	0	620	0	627	776	119	480	641	506	28
(2,1)	70	181	185	83	190	25	248	268	93	266	90	245	278	274	268	75	100	51	112	47
Total	**8,708**	**9,560**	**9,553**	**9,022**	**9,558**	**9,533**	**10,837**	**10,837**	**10,219**	**10,449**	**16,532**	**18,514**	**18,068**	**18,482**	**18,333**	**11,106**	**13,376**	**13,303**	**13,348**	**12,812**
% Improve		**9.78**	**8.85**	**3.29**	**9.43**		**13.68**	**13.68**	**7.19**	**9.61**		**11.99**	**9.29**	**11.79**	**10.89**		**20.44**	**19.78**	**20.19**	**15.36**

Table 7.14 shows the revenue comparison per each train trip of the selected five stations across four train trips. Result shows that the revenue improvement ranges from 3.29% to 20.44% per train trip depending on the departure time and the choice model. Train#4 has the most revenue improvement compared to other trains. The B-spline choice model results in the revenue improvement from 9.43% to 15.36% per train trip, its improvement is comparable to other models for Train#1 and Train#3, while for Train#2 and Tran#4 the B-spline revenue improvements are relatively lower than other models. This could be influenced by variability of the price sensitivity imposed by the B-spline model which better represents heterogeneity of the passengers and reveals the behavior of price sensitive passengers more explicitly. In turn, this results in lower fare strategy for the two train trips (Train#2 and Train#4) and consequently contributes to less revenue improvement. Detailed fare strategy analysis for the selected high demand market is shown in the next section.

7.3.6 Pricing Result

In this section, we select market (4,1) which has the most traffic in our selected network to elaborate the result in detail. For this market, we consider four different departure times which load passengers into Train#1 to Train#4. These correspond to the departure times from station 4 at 8AM, 12PM, 4PM, and 7PM respectively based on Table 7.6.

Fares of market (4,1) are shown in Figure 7.24, 7.26, 7.28, and 7.30 for the four train trips departing from Station4 at 8AM, 12PM, 4PM, and 7PM respectively. The results from all the models are compared to the existing fare, which is representative fare pattern of a particular departure time of day and day of week obtained from the data. In general, while the MNL, the LC, and the ML with parametric distribution provide similar results across 4 departure times, the optimizations with B-spline choice model provide distinct fare pattern which is consistent with our expectation that price sensitivity should decrease as time approaches departure date.

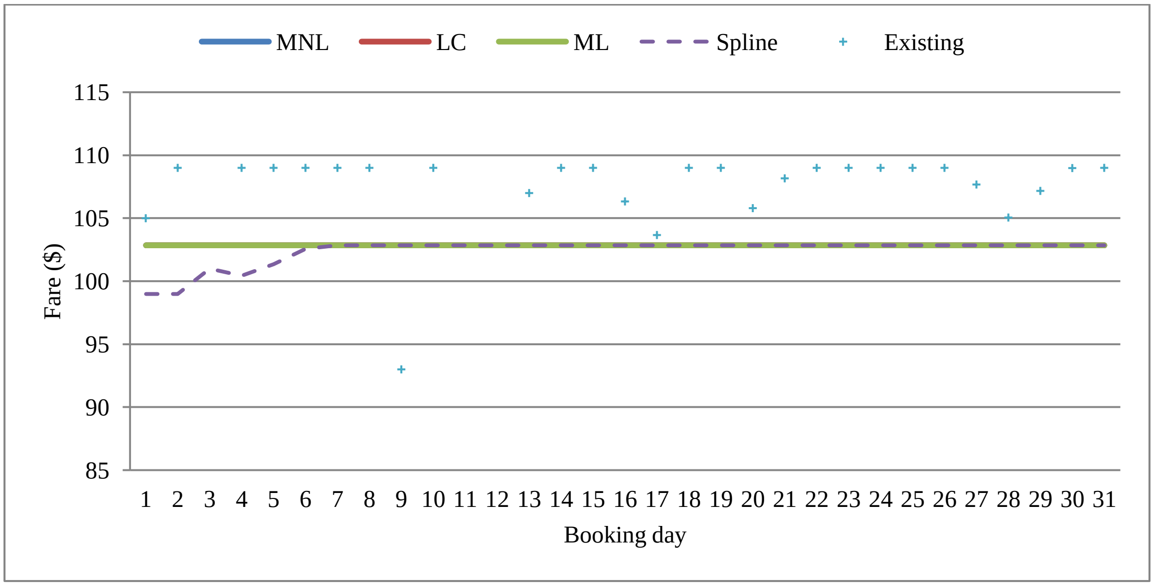


Figure 7.24 Market (4,1) 8AM departure fare

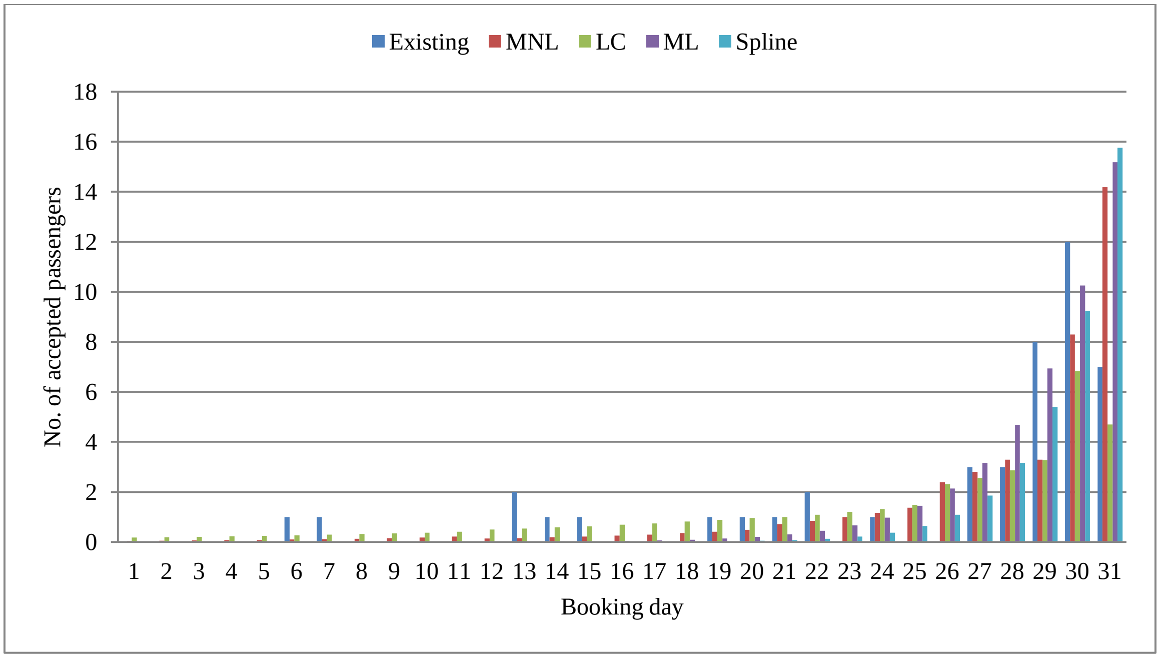


Figure 7.25 Market (4,1) 8AM departure demand

Figure 7.25, 7.27, 7.29, and 7.31 show the corresponded number of accepted passengers in market (4,1) on each day over the sale horizon across four departure times. The responses of passenger demands to the choice models are consistent with the expectation of price sensitivity; with greater number of passengers when the new fares are lower than existing, and lower number of passengers vice versa. This behavioral pattern is observed across all the four train departure times.

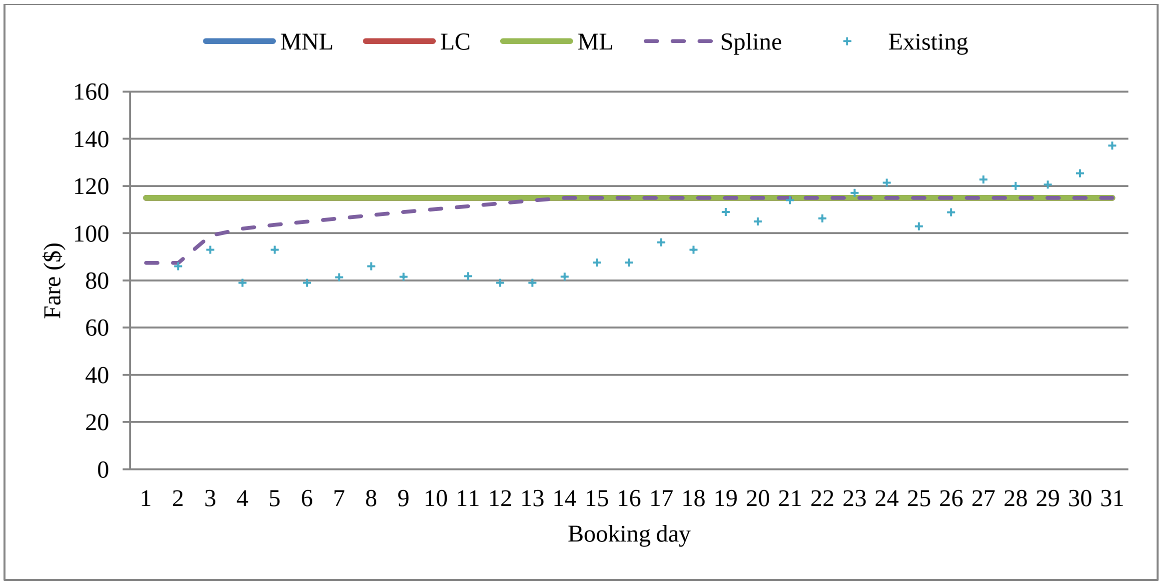


Figure 7.26 Market (4,1) 12PM departure fare

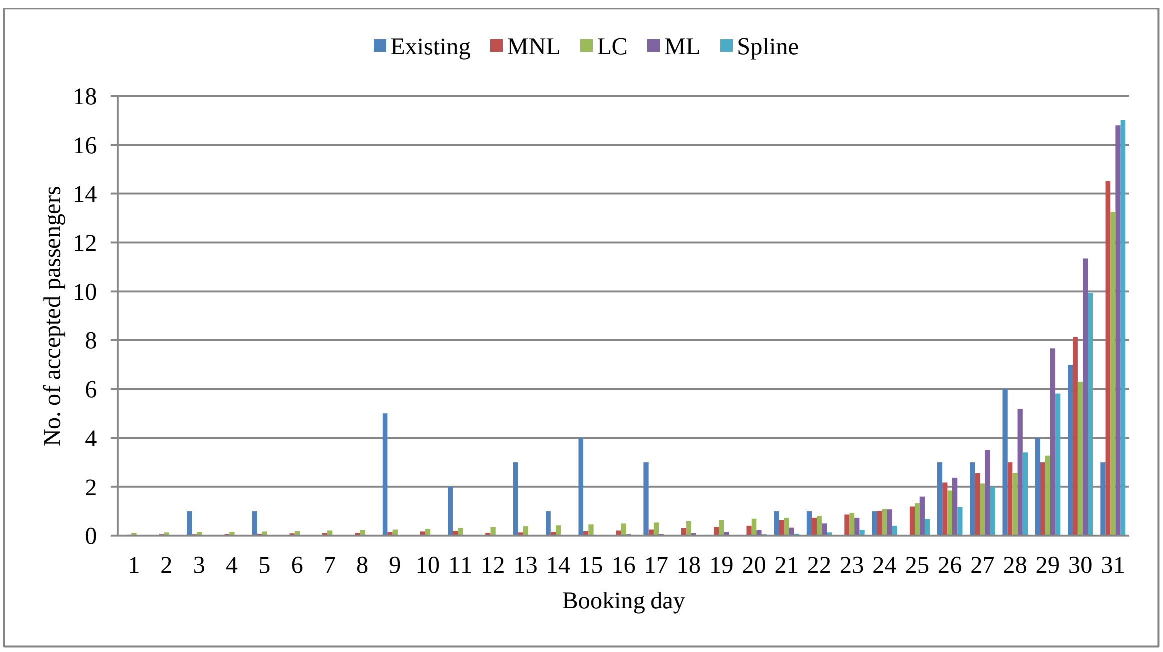


Figure 7.27 Market (4,1) 12PM departure demand

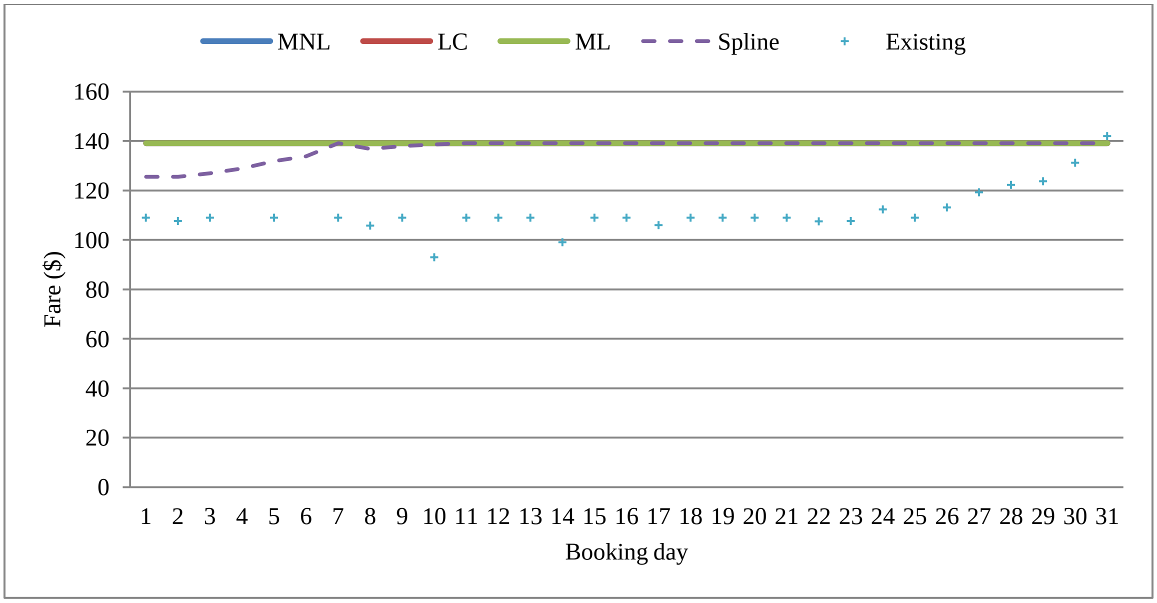


Figure 7.28 Market (4,1) 4PM departure fare

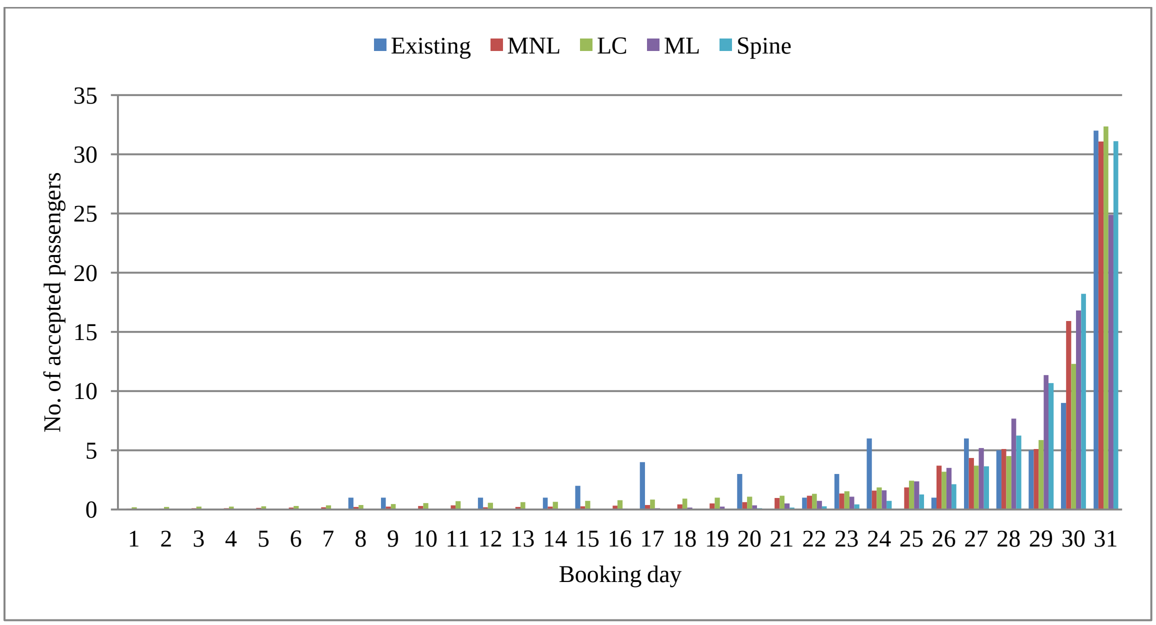


Figure 7.29 Market (4,1) 4PM departure demand

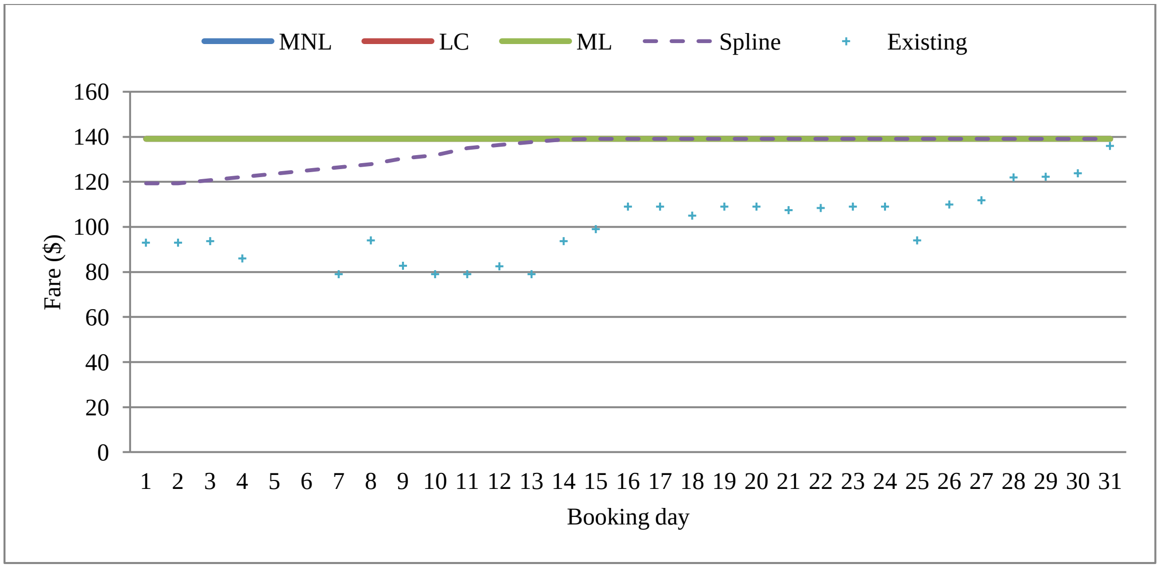


Figure 7.30 Market (4,1) 7PM departure fare

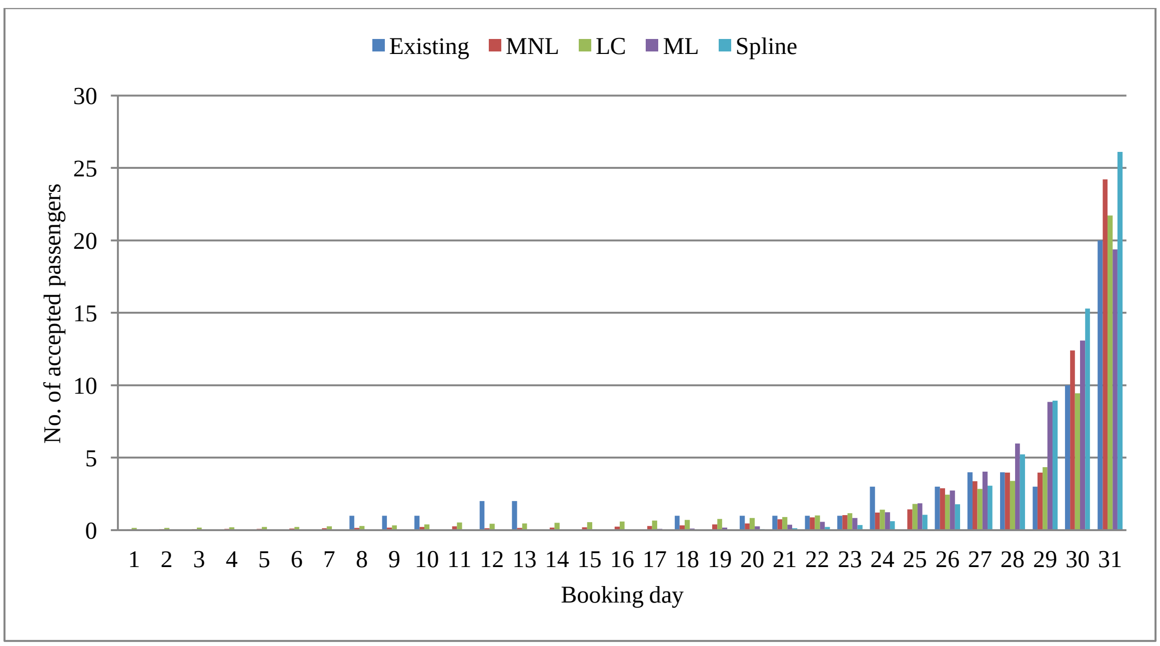


Figure 7.31 Market (4,1) 7PM departure demand

7.4 Conclusions

In this chapter, the impacts of incorporating passenger choice models in RM strategy are assessed. RM optimization models are developed where the passenger choice models and demand functions are incorporated under assumptions and capacity constraints determined on the basis of the single-leg and network characteristics.

In single-leg approach, the deriving pricing strategy leads to a potential increase in revenues ranging from 1.92% to 13.76% per day and from 0.65% to 10.60% per day for the representative weeks of March and April respectively. This case study is based on the assumption that each market independently optimizes its ticket revenue without allowing for capacity redistribution across the network.

In network approach, the framework accounts for shared capacity which enables for capacity redistribution across the network. We first propose a methodological framework to incorporate a latent class model of ticket purchase timing decision in the railway pricing and seat allocation problem. The approach allows RM strategy to explicitly account for passenger taste heterogeneity by distinguishing between leisure and business travelers, depending on departure time and day of week. Results obtained have illustrated the impacts that the strategy derived from the optimization procedure has on the existing conditions in terms of fare, capacity distribution, and revenue. Seat allocation policy results into more short-haul trips acceptance, which contributes to greater revenue than long-haul trip with the same seat capacity. The solution from the proposed framework results in significant revenue improvement from 16.24% to 24.96% per train trip and from 13.82% to 21.39% per train trip for the MNL and the LC choice model, respectively, depending on the train departure time. Finally, the optimization system also provides indications on how to redistribute capacity efficiently across the markets considered.

Then, the mixed logit models with parametric and non-parametric distribution are proposed. The non-parametric approach adopts B-spline curves as a polynomial approximation of arbitrary distribution without assuming the shape of the distribution. Results show that mixed logit with non-parametric B-spline improves revenues from 9.43% to 15.36% per train trip, which is comparable to other choice models (from 3.29% to 20.44% per train trip). However, the pricing strategy provided by B-spline is more in line with the expectation based on passenger price sensitivity behavior.

This chapter has illustrated how passenger choice models developed can complement RM strategy in term of pricing and seat allocation. In particular, it shows that accounting for passenger taste heterogeneity based on discrete and continuous segmentation approach results in more realistic representation of the passenger behavior when incorporated in RM strategy.

Chapter 8 : Dynamic Discrete Choice Model

The increasing use of internet as a major ticket distribution channel has resulted in passengers becoming more strategic to fare policy. This potentially induces passengers to book the ticket well in advance in order to obtain a lower fare ticket, and later adjust their ticket when they are sure about trip scheduling. This is especially true in flexible refund markets where ticket cancellation and exchange behavior has been recognized as having major impacts on revenues. Therefore, when modeling this behavior, it is important to account for the characteristic of the passenger that optimally makes decision over time based on trip schedule and fare uncertainty.

In this chapter, we propose an inter-temporal choice model of ticket cancellation and exchange for railway passengers where customers are assumed to be forward-looking agents. A dynamic discrete choice model (DDCM) is applied to predict the timing of ticket cancellations and exchanges in response to fare and trip schedule uncertainty. Passengers' decisions involve a two step process. First, the passenger decides whether to keep or adjust the ticket. Once the decision to adjust the ticket has been made, the passenger has the choice to cancel the ticket or to change departure time. The problem is formulated as an optimal stopping problem, and a two step look-ahead policy is adopted to approximate the dynamic programming problem.

To this scope, the main setting of our problem is outlined in Section 8.1. Based on the analysis of ticket cancellation and exchange data in Section 4.2, the selected sample for model development is described in Section 8.2. From the comprehensive review in Section 2.3, we formulate a dynamic discrete choice model and formalize the algorithm used for the dynamic programming problem in this study in Section 8.3 and Section 8.4 respectively. The approach is applied to simulated and real ticket reservation data for intercity railway trips in Section 8.5 and Section 8.6 respectively. In these sections, the superiority of the method proposed for modeling exchange and cancel decisions is demonstrated. Finally, conclusions drawn from the empirical analysis and future research directions are outlined in Section 8.7.

8.1 Problem Setting

In this chapter, we propose a dynamic framework based on discrete choice models developed in the context of railway revenue management. To the authors' knowledge, this is the first attempt to incorporate dynamics in individual choices in revenue management modeling and in particular to formalize tickets' exchange and cancel decisions for railway intercity trips. The railway operator in consideration offers fully refundable fare and provides flexibility in ticket exchange which makes ticket cancellation and exchange decision to be very crucial to the RM system. Passengers are inclined to purchase ticket early and adjust their ticket later when they are more certain about trip schedules. The model accounts for passengers' trip adjustment choice and explicitly specifies the probability of exchanging ticket as a function of the set of available exchange tickets. The choice set is constituted by all departure times offered by the railway operator between a specific origin-destination pair.

8.2 Sample Selection

Data mentioned in Section 4.2 are used for the model development in this chapter. The problem is further simplified by considering only passengers who made weekday trips from south end terminal station to 3 major destinations (named STA1, STA2, and STA3), and purchased the ticket 15 days before departure which results into a time horizon of 16 days for each decision maker (from 15 days before departure until departure day). This results in 696 valid individual passenger records for model estimation.

8.3 Problem Formulation

8.3.1 Passenger Stopping Problem

We consider a passenger set $\mathfrak{I} = \{1, \dots, M\}$ where each passenger $i \in \mathfrak{I}$ can be in one of the two possible states $s_{it} = \{0,1\}$ in time period $t \in \{0,1, \dots, T\}$. Passenger is considered to be in the decision process when $s_{it} = 0$ and out of the decision process when $s_{it} = 1$. In each time period t, passenger i in state $s_{it} = 0$ has two options:

1. To make change to the ticket (either exchange or cancel). Once decided to adjust the ticket, the passenger makes the choice of $j \in \mathfrak{J}_t$ which is composed of exchange (departure time specific exchange decision at time period t) and cancel alternatives and obtain a terminal period payoff u_{ijt}. The utility of exchange is primarily a function of fare difference between the original and the exchange ticket at time period t. The utility of ticket cancellation is primarily a function of trip characteristics, and the refund amount.
2. To keep the original ticket and obtain a one-period payoff U_{ikt}, which is normalized to have a mean of zero before departure day and equal to c on departure day.

The two-step decision process assumes that, at each time period, the passenger decides whether to keep or change the ticket. The optimal time period in which passenger decides to change the ticket is denoted by τ, where the passenger chooses the ticket change alternative j_t^* that maximizes the utility from $\mathfrak{I}$. The passenger decision is the optimal stopping problem at time t:

$$D\left(u_{i1t}, \dots, u_{iJt}, U_{ikt}, t\right) = \max_{\tau}\left\{\sum_{k=t}^{\tau-1} U_{ikt} + E\left[\max_{j\in\mathfrak{I}} u_{ij\tau}\right]\right\} \tag{8.1}$$

Let $v_{it} = \max_{j\in\mathfrak{I}} u_{ijt}$. We assume that v_{it} is Gumbel distributed with a scale factor equals to 1. Based on dynamic programming theory, the passenger's decision can be transformed into:

$$D(v_{it}, U_{ikt}) = \max\left\{v_{it}, U_{ikt} + E\left[D\left(v_{i,t+1}\right)\right]\right\} \tag{8.2}$$

The reservation utility is defined by function:

$$W_{it} = U_{ikt} + E\left[D\left(v_{i,t+1}\right)\right] \tag{8.3}$$

And consider the optimal policy:

$$\begin{cases} v_{it} & \text{if } v_{it} \geq W_{it} \\ W_{it} & \text{otherwise} \end{cases} \tag{8.4}$$

The problem can be simplified as:

$$D(v_{it}) = \max(v_{it}, W_{it}) \tag{8.5}$$

A. Keep Ticket Probability

The passenger i will keep the ticket at time t when $W_{it} \geq v_{it}$. Let π_{i0t} denotes the probability of keeping ticket until the next period, which can be written as:

$$\pi_{i0t} = P[\, \upsilon_{it} \leq W_{it}] = P[\text{keep}|s_{it} = 0] \quad (8.6)$$

$$= F_{\upsilon}[W_{it},\, \upsilon_{it}] = e^{-e^{-(W_{it}-r_{it})}} \quad (8.7)$$

where r_{it} is the mode of the distribution of υ_{it} that is:

$$r_{it} = \ln G(e^{V_{i1t}}, \dots, e^{V_{ijt}}) \quad (8.8)$$

B. Change Ticket Probability

The probability of ticket change is $P[\text{change}|s_{it} = 0] = 1 - \pi_{i0t}$ and the choice specific ticket change probability is:

$$\pi_{ijt} \quad = P\big[U_{ijt} \geq U_{ilt}, \forall l \neq j, u_{it} \geq W_{it}\big] \quad (8.9)$$

$$= P\big[U_{ijt} \geq W_{it}\big|U_{ijt} \geq U_{ilt}, \forall l \neq j\big]P\big[U_{ijt} \geq U_{ilt}, l \neq j\big] \quad (8.10)$$

$$= (1 - \pi_{i0t})P\big[U_{ijt} \geq U_{ilt}, l \neq j\big] \quad (8.11)$$

8.3.2 Objective Function and Parameters to Estimate

The parameter estimation is performed by maximizing the likelihood function:

$$\mathcal{L}(\beta) = \prod_{i=1}^{N}\prod_{t=0}^{T} P_{it}\,[\text{decision}] \quad (8.12)$$

The decision probability is presented as:

$$P_{it}[\text{decision}] = P_{it}[\text{decision}, s_{it} = 0\,] + P_{it}[\text{decision}, s_{it} = 1] \quad (8.13)$$

$$= P_{it}[\text{decision}|\, s_{it} = 0\,]P[s_{it} = 0\,] + P_{it}[\text{decision}|\, s_{it} = 1]P[s_{it} = 1\,] \quad (8.14)$$

The state s_{it} is observed in the data set, if the passenger has not changed the ticket, $P[s_{it} = 0\,] = 1$ and $P[s_{it} = 1\,] = 0$. Once the passenger change the ticket, the passenger is considered to be out of the decision process, therefore $P[s_{it} = 0\,] = 0$ and $P[s_{it} = 1\,] = 1$. When the passenger is out of the decision process, the probability does not affect the result of the likelihood function. As a result, the completed likelihood function in this problem is:

$$\mathcal{L}(\beta) = \prod_{(i,t)\in V} P_{it}[\text{decision}, s_{it} = 0\,] \tag{8.15}$$

where $V = \{(i,t)|i \in \{1,\dots,M\}, t \in \{0,\dots,T\}\}$. The decisions include keeping the ticket and ticket change specific choice. Thus $P_{it}[\text{decision}, s_{it} = 0\,] = \{\pi_{i0t}, \pi_{ijt}\}$.

8.4 Dynamic Estimation Process

The estimation process is done with maximum likelihood estimation method. First π_{i0t} must be obtained in order to calculate π_{ijt}. The probability π_{i0t}, depends on W_{it} which can be calculated from : $W_{it} = U_{ikt} + E\left[D\left(\upsilon_{i,t+1}\right)\right]$, assuming that r_{it} is the mode of the distribution of υ_t.

W_{it} is composed of two parts: the utility of the current ticket attributes (U_{ikt}) and the expected utility in the next time period $\left(E\left[D\left(\upsilon_{i,t+1}\right)\right]\right)$. At each time period, the passenger is assumed to have a perception about the future scenarios, which are characterized by the alternative attributes changing over time. The expectation utility accounts for the possible market conditions in the passenger's perceived scenario; in our specification, the fare of each departure time specific exchange decision has been selected as independent variable in the utility specification. A passenger is assumed to have a perception of future attributes for a limited number of time periods, denoted by T. At time period t, the passenger faces two alternatives, keeping the ticket or changing the ticket. The passenger will continue the decision process into the period $t+1$ only if he had decided to keep the ticket in time period t. Therefore, the decision process can be characterized by a scenario tree with a unique pattern (shown in Figure 8.1). This scenario tree constitutes the base for the expected utility calculation. The following steps describe the procedure to calculate $\pi_{i0,0}$ and $E[D(\upsilon_{i1})]$ which will be indicated by $E[D_1]$ because all the expectations in the example are for individual i.

The procedure for calculating the expected utility will be described in detail as follows:

- First, we assume that the passenger has the expectation over a limited number of future time periods, which is limited to two in order to reduce the number of leaves in the scenario tree. At time period $t = 0$, the passenger can anticipate the future ticket characteristics (i.e. fare) from time period $t = 1$ and $t = 2$. The terminal time period expected utility $E[D_3] = 0$ because the passenger knows nothing for time period 3 when being at time period 0.
- Calculate $E[D_1]$. In order to obtain $\pi_{i0,0}$ from equation (8.6), the reservation utility (W_{i0}) is required. The reservation utility (W_{i0}) can be obtained from equation (8.3) $W_{i0} = U_{ik0} + E[D_1]$ which requires the calculation of $E[D_1]$. At time 0, the passenger has two alternatives for successive time 1, keep the ticket or change the ticket. The second term at the right hand side of the function $E[D_1] = \mathrm{E}\{\max[\upsilon_1, U_{ik1} + E[D_2]]\}$ represents the utility of keeping alternative; therefore when calculating $E[D_2]$, it is necessary that the term corresponded to the left leave of the tree be obtained (indicated by dash line in Figure 8.1). The calculation $E[D_2] = \mathrm{E}\{\max[\upsilon_2, U_{ik2} + E[D_3]]\}$ demands the same function to be calculated for time period 3 ($E[D_3]$) which is assumed to be zero according to the above assumption. The process of calculating $E[D_1]$ is recursive with known

utility at the end of the perspective horizon (assumed to two periods in this formulation). After $E[D_1]$ is calculated, reservation utility at time 0 (W_{i0}) can be obtained.

- This calculation procedure can be repeated to calculate $\pi_{i0,1}$ with the assumption that respondent can anticipate characteristics for time period 3 and $E[D_4] = 0$.

The reason that a terminal value for the expected utility has to be fixed at zero is because it is difficult to predict a particular value for the individual's perspective when future time period is far beyond his knowledge of information. This means that in the long term, the individual has not enough information to predict the future; passengers cannot anticipate the utility of keeping or cancelling the ticket. With this approach, after a limited number of time periods, information on future ticket fare attribute is just ignored.

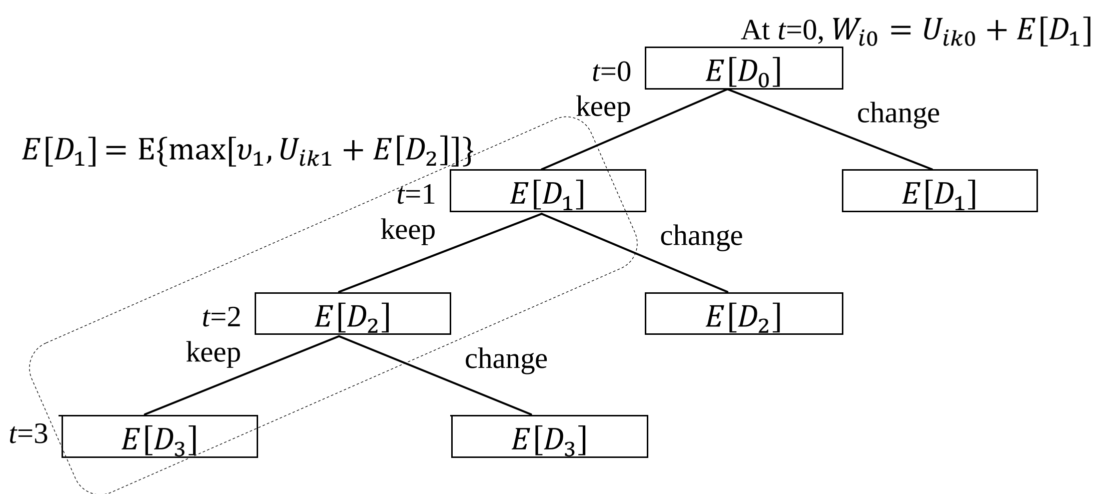


Figure 8.1 Scenario tree

8.5 Experiment with Simulated Data

Synthetic ticket reservation data over time periods are simulated to validate the proposed dynamic discrete choice formulation. The data are created assuming that the characteristics of choice behavior are known; by adopting this procedure it is possible to test the ability of the dynamic discrete choice model to recover the true value of the parameters used to generate the data and to reproduce observed choice of individuals over time. Comparisons with static models, in the form of multinomial logit, are also presented.

8.5.1 Data Construction

The simulated data are partially simulated from the real individuals' record whose characteristics are described in Section 8.2. Synthetic data assume that passengers have the same origins and destinations as the real data, while individual characteristics, departure day of week, and departure time, vary from the real data. Concerning individual characteristics, the group size variable is generated from a uniform distribution and varies between 1 and 3 persons. Departure day of week is assumed to be uniformly distributed across the weekdays, while departure time is assumed to be uniformly distributed on discrete hour clock time between 5:00 AM and 7:00 PM. Ticket fare of the original departure time and other departure times within the same departure day are constructed for each day over the decision horizon based on historical data; the constructed fares vary by departure day of week and time of day.

Each individual is supposed to provide responses over a 16 day time period starting from 15 days before departure until the departure day. A total of (16×696)

observations are then generated. There are 17 alternatives in the choice set, the first 15 alternatives refer to departure time specific exchange decisions (5:00 AM to 7:00 PM), the 16th alternative is cancel, and the 17th alternative is keeping the ticket. An important assumption in the data construction process is that if at one period the passenger decides to make a change to his ticket, then this passenger will no longer be part of the decision process in the next time period (he is out of the market). This results in a total of 10,199 observed decisions that are valid for model estimation. True value parameters have been used to determine individual choices. Synthetic observations are then used to estimate both the static multinomial logit (MNL) and the dynamic discrete choice model (DDCM).

8.5.2 Model Specification

The model specification considers 16 discrete time horizons defined by $t \in \{0,1,\ldots,15\}$ where t also represents the number of day from original ticket purchase. The first time period is the day when original ticket is purchased ($t = 0$), (day1). The last time period is departure day ($t = 15$), (day16). The utility specification is defined as follows:

$$
\left.
\begin{array}{l}
U_{i5t=}\beta_{cost}(f_{5t} - f_{b0}) + \varepsilon_{i5t} + \varepsilon_i \\
\quad \dots\dots\dots\dots\dots\dots\dots\dots. \\
U_{ijt=}\beta_{cost}\left(f_{jt} - f_{b0}\right) + \varepsilon_{ijt} + \varepsilon_i \\
\quad \dots\dots\dots\dots\dots\dots\dots\dots\dots. \\
U_{i19t=}\ \beta_{cost}(f_{19t} - f_{b0}) + \varepsilon_{i19t} + \varepsilon_i
\end{array}
\right\} \text{15 Exchange alternatives (5:00 – 19:00).} \tag{8.16}
$$

$$
U_{ict=}\ ASC_{cnl} + \beta_{ev}ev + \beta_{Fri}Fri + \beta_{STA3}STA3 + \beta_{ref}f_{b0} + \varepsilon_{ict} + \varepsilon_i
$$

$$
U_{ikt=}\begin{cases} c + \varepsilon_{ikt} + \varepsilon_i & \text{if } t = 15 \\ \varepsilon_{ikt} + \varepsilon_i & \text{if } t < 15 \end{cases}
$$

The utility of individual i on alternative j is denoted by U_{ijt}. For ticket exchange decision, the index j indicates 15 exchange departure times (5:00 AM to 7:00 PM). The utility of exchange (U_{ijt}) includes exchange cost defined as the difference between the original fare (f_{b0}) and the new fare (f_{jt}) at time t. The model allows a passenger to exchange the ticket for the same departure time as in the original ticket; transactions of this type are observed in the real data. This decision will result in the passenger paying the difference between the original cost ($t = 0$) and the cost at time t. The utility of cancel (U_{ict}) includes alternative specific constant (ASC), refund, dummy of original departure in the evening (3:00-7:00 PM.), dummy of original departure on Friday, and dummy of STA3 destination. The utility of keep (U_{ikt}) has two different specifications. In the last time period ($t = 15$) passengers deciding to keep the ticket obtain the utility which includes the constant term referring to utility of traveling with the original ticket. In other time periods ($t < 15$) the systematic term of the keep utility is normalized to zero. ε_{ijt} is the random error term for each individual i, alternative j at a given time period t. ε_i is the individual error term which is assumed to be constant across all observations produced by the same respondent.

To evaluate the ability to recover the true value of the model, the root mean square deviation (RMSD) is adopted as measure of differences between the true values and the estimated coefficient values. The RMSD is defined as:

$$RMSD(\hat{\theta}) = \sqrt{E\left[(\hat{\theta} - \theta)^2\right]} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{n}} \tag{8.17}$$

where n is the number of parameters. Using the simulated data with the utility specification defined above, two models are estimated: dynamic discrete choice model (DDCM) and static multinomial logit (MNL) model. In the static model, the attributes of the future ticket characteristic (fare) are not considered when making decisions in each time period. The model is simply formulated as a traditional MNL model with 17 alternatives (15 exchange decisions, 1 cancel, and 1 keep). The dynamic model with the algorithm defined in the formulation is coded in the C language and makes use of optimization tools available in AMLET (Another Mixed Logit Estimation Tool), (Bastin, 2011).The static model is estimated using ALOGIT (ALOGIT, 2007). The utilities specifications are assumed to be the same for the static and the dynamic model; the derived estimation results are compared in Table 8.1.

8.5.3 Estimation Result

All parameters in both the MNL and the DDCM models are statistically significant at the 5% significance level. The RMSD value obtained with the dynamic model is lower than the MNL model (0.82 compared to 3.93); this indicates that the dynamic model outperforms the MNL model in recovering the true value of the parameters. The rho-squared value obtained from the dynamic model exceeds that obtained from the MNL model (0.7070 compared to 0.3814); this indicates that the dynamic model has a better goodness of fit than the MNL model.

Table 8.1 Estimation result: simulated data

	Exchange	Cancel	Keep	True Value	MNL Est	MNL T-stat		Dynamic (2-SL) Est	Dynamic (2-SL) T-stat	
ASC cancel		x		-5.00	-12.730	-14.2	*	-5.434	6.7	*
Orig Deptt 3-7 pm		x		2.50	1.392	5.2	*	3.927	8.1	*
Depart Friday		x		-2.00	-1.456	-3.8	*	-2.018	3.3	*
STA3 destination		x		4.00	2.517	10.1	*	5.252	10.8	*
Exchange cost	x			-0.02	-0.040	-77.8	*	-0.020	12.1	*
Refund		x		0.03	0.032	8.4	*	0.030	5.5	*
Keep (day 16)			x	-7.00	-6.477	-8.9	*	-5.868	27.0	*
Cancel day1		x		3.00	6.609	9.1	*	2.275	13.2	*
Exchange day16	x			1.50	-6.144	-8.4	*	2.343	8.5	*
Early exchange	x			-2.00	-6.366	-41.4	*	-2.598	13.6	*
Log-likelihood (0)						-28,896			-2,908	
Log-likelihood (final)						-17,875			-852	
Likelihood ratio index (rho-squared)						0.3814			0.7070	
RMSD						3.93			0.82	
No. of individuals						696			696	
No. of observations						10,199			10,199	

* Statistically significant at the 5% significance level.

8.5.4 Model Validation

The results obtained from the estimation are used to validate how well the models reproduce the observed simulated decisions (Table 8.2).

Table 8.2 Validation result: simulated data

Choice	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
Exchange	MNL	3	3	3	3	2	2	2	3	3	2	3	2	3	2	3	3
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
2	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
Exchange	MNL	4	4	4	4	4	6	7	4	4	5	5	5	5	3	3	2
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
3	Observed	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
Exchange	MNL	4	5	7	6	6	8	6	6	6	7	4	3	3	2	2	2
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
4	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
Exchange	MNL	15	10	11	11	12	11	11	12	11	13	13	13	13	12	10	7
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
5	Observed	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	27
Exchange	MNL	31	31	30	22	23	20	21	21	20	17	14	15	23	19	19	17
	DDCM	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	20
6	Observed	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	18
Exchange	MNL	44	36	30	34	30	27	27	31	29	33	29	27	26	25	21	17
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18
7	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
Exchange	MNL	50	53	35	33	36	32	28	39	38	44	43	46	47	33	30	33
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25
8	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26
Exchange	MNL	48	51	35	34	37	38	42	34	40	24	27	26	35	38	33	33
	DDCM	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	22
9	Observed	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	27
Exchange	MNL	44	46	37	37	38	32	45	37	38	43	40	39	37	35	29	22
	DDCM	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	29
10	Observed	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	28
Exchange	MNL	17	16	17	15	14	19	17	20	17	21	21	23	19	16	17	14
	DDCM	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	30
11	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
Exchange	MNL	16	17	17	21	19	17	26	21	21	19	17	19	19	23	17	13
	DDCM	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	19
12	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28
Exchange	MNL	12	18	22	19	20	17	20	19	18	20	16	19	14	15	14	15
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28

Choice	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
13	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32
Exchange	MNL	23	27	38	43	37	32	36	29	31	39	32	29	28	28	25	22
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37
14	Observed	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43
Exchange	MNL	59	77	69	79	67	109	64	61	66	56	58	73	47	50	51	41
	DDCM	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	42
15	Observed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76
Exchange	MNL	183	151	132	122	135	113	127	137	137	110	132	104	108	107	113	113
	DDCM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	77
16	Observed	54	0	1	0	0	0	0	1	0	0	0	0	0	0	0	94
Cancel	MNL	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	67
	DDCM	42	0	2	0	2	0	1	3	0	0	0	0	1	3	0	88
17	Observed	638	637	636	635	634	634	633	632	632	632	632	632	632	632	632	148
Keep	MNL	90	93	150	153	154	152	155	157	153	179	180	188	205	224	246	210
	DDCM	653	637	634	636	632	633	631	629	632	631	632	630	631	629	632	172
Total	Observed	4	1	0	1	1	0	1	0	0	0	0	0	0	0	0	390
Exchange	MNL	552	545	487	483	481	482	479	476	479	453	452	444	427	408	386	355
(1-15)	DDCM	1	1	1	0	1	1	2	1	0	1	0	2	0	0	0	372
	Total	**696**	**638**	**637**	**636**	**635**	**634**	**634**	**633**	**632**	**632**	**632**	**632**	**632**	**632**	**632**	**632**

The validation in Table 8.2 indicates that the major drawback of the MNL model is the over-prediction of exchange decisions, especially exchange decision choice 15 (exchange to 7 PM); which is characterized by low fare. The MNL model predicts the cancel decisions in the first time period (day1) precisely; however, its prediction on cancel decisions in the last time period (day16) is less good than the one produced by DDCM. More importantly, in the last time period (day16) with a high number of exchange decision, the DDCM model clearly outperforms the MNL. We selected the first (day1) and last time period (day16) of cancel decision and exchange decisions to show in detail the prediction capability of the DDCM over the MNL (Figure 8.2).

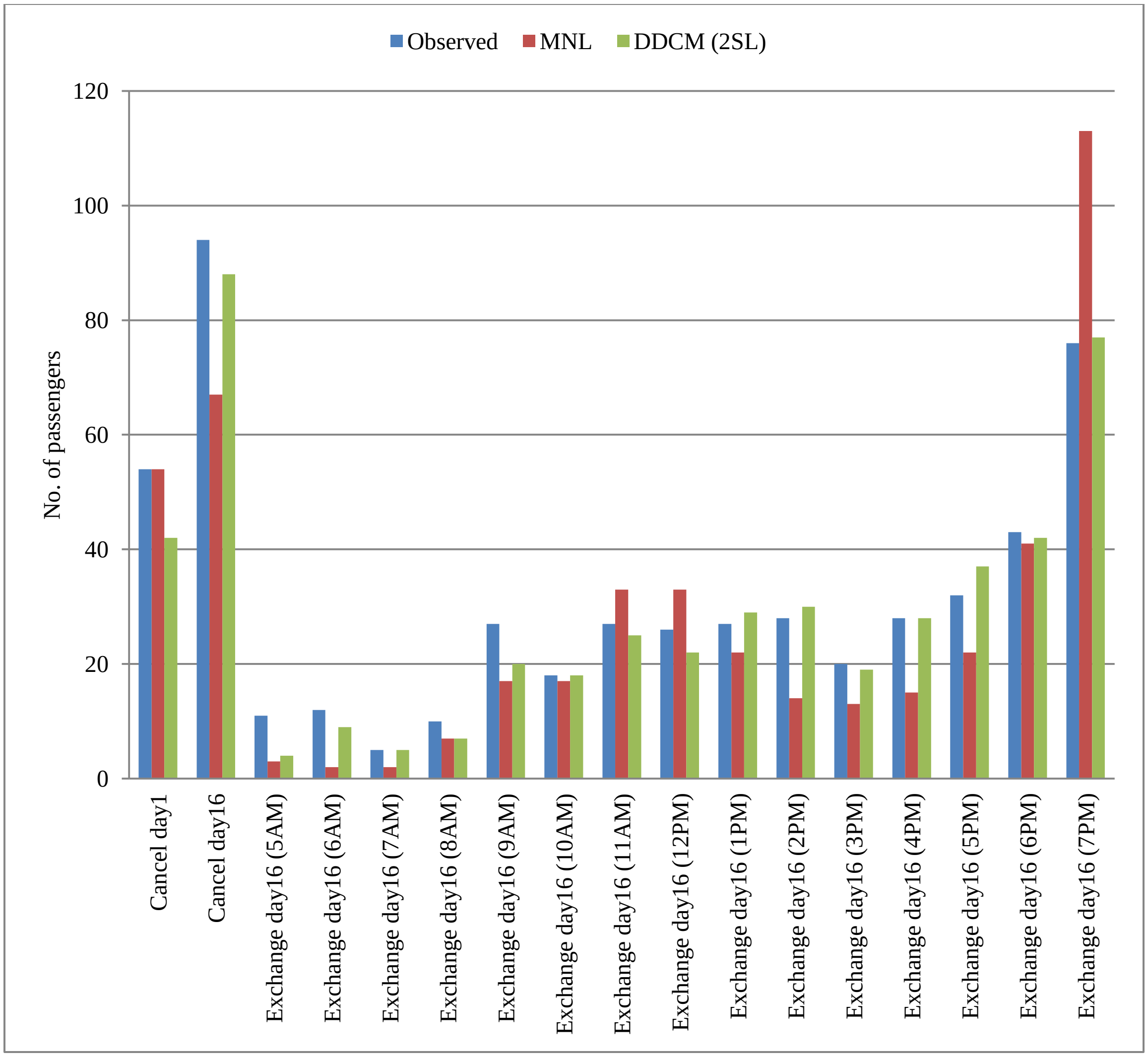


Figure 8.2 Simulated data validation: departure time specific exchange and cancel decision

Figure 8.3 to Figure 8.5 briefly summarize the predictions over different time period (day) where exchange decision is aggregated over all departure times.

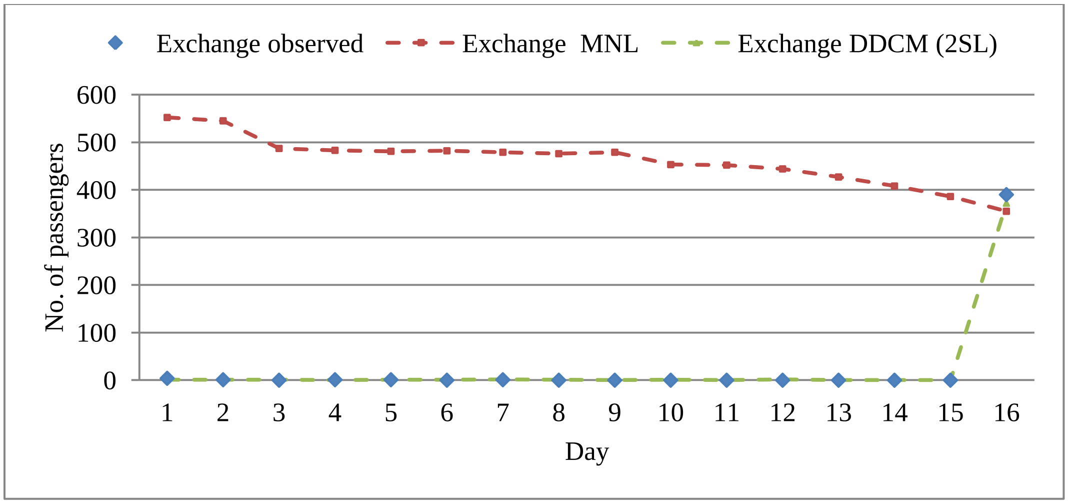


Figure 8.3 Validation of exchange decision: simulated data

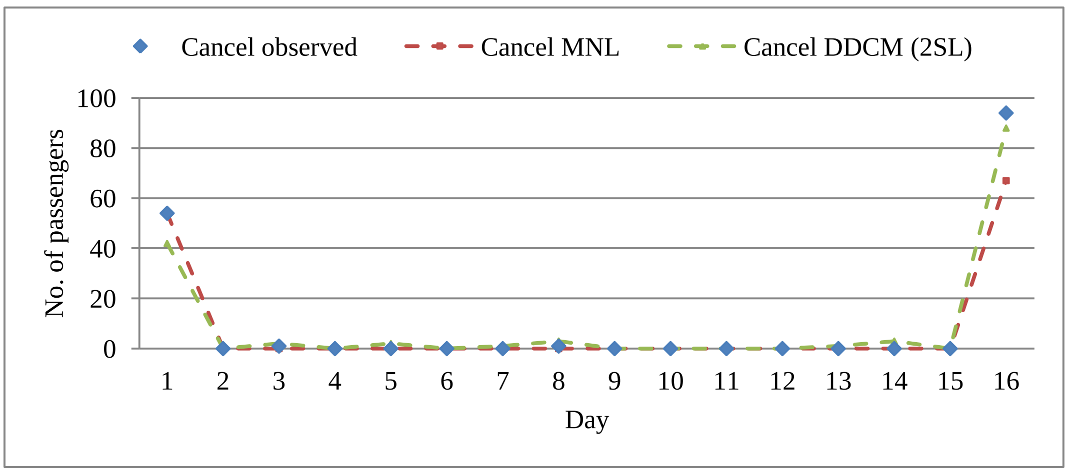


Figure 8.4 Validation of cancel decision: simulated data

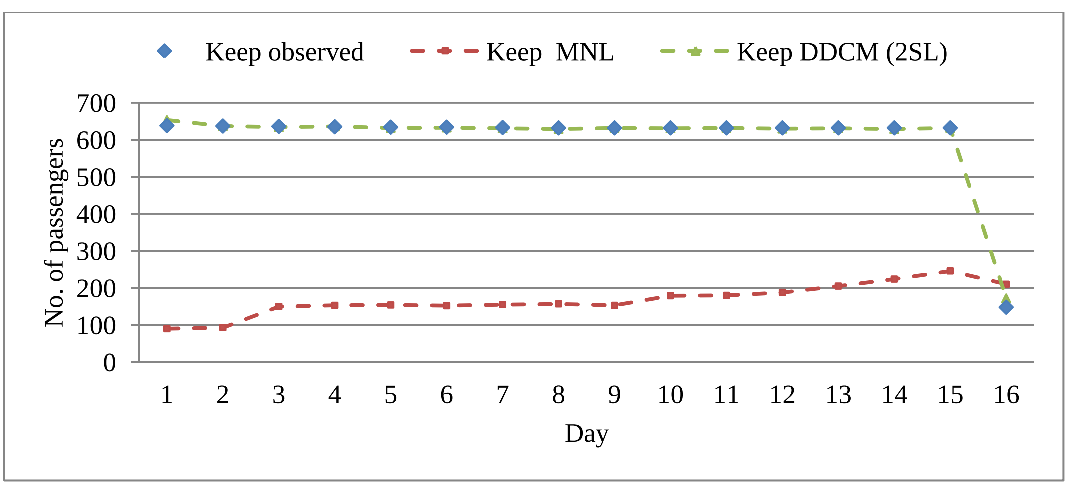


Figure 8.5 Validation of keep decision: simulated data

The choice probability for each alternative observed and predicted together with measure of errors for the simulated data experiment is reported in Table 8.3. The absolute error D is used to represent measure of error which is defined as:

$$D = |M_{pred} - M_{obs}| \tag{8.18}$$

where D is error norm; M_{pred} is a vector of predicted choice probability, and M_{obs} is a vector of observed choice probability. The D value obtained from the dynamic model is significantly smaller than the corresponding value in the MNL model (0.176 compared to 22.214) indicating a better prediction capability of the dynamic model over the MNL model.

Table 8.3 Model validation: choice probability of simulated data experiment

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Exchange day1	0.0057	0.7931	0.0014
Exchange day2	0.0016	0.8542	0.0016
Exchange day3	0.0000	0.7645	0.0016
Exchange day4	0.0016	0.7594	0.0000
Exchange day5	0.0016	0.7575	0.0016
Exchange day6	0.0000	0.7603	0.0016
Exchange day7	0.0016	0.7555	0.0032
Exchange day8	0.0000	0.7520	0.0016
Exchange day9	0.0000	0.7579	0.0000
Exchange day10	0.0000	0.7168	0.0016
Exchange day11	0.0000	0.7152	0.0000
Exchange day12	0.0000	0.7025	0.0032
Exchange day13	0.0000	0.6756	0.0000
Exchange day14	0.0000	0.6456	0.0000
Exchange day15	0.0000	0.6108	0.0000
Exchange day16	0.6171	0.5617	0.5886
Cancel day1	0.0776	0.0776	0.0603
Cancel day2	0.0000	0.0000	0.0000
Cancel day3	0.0016	0.0000	0.0031
Cancel day4	0.0000	0.0000	0.0000
Cancel day5	0.0000	0.0000	0.0031
Cancel day6	0.0000	0.0000	0.0000
Cancel day7	0.0000	0.0000	0.0016
Cancel day8	0.0016	0.0000	0.0047
Cancel day9	0.0000	0.0000	0.0000
Cancel day10	0.0000	0.0000	0.0000
Cancel day11	0.0000	0.0000	0.0000
Cancel day12	0.0000	0.0000	0.0000
Cancel day13	0.0000	0.0000	0.0016
Cancel day14	0.0000	0.0000	0.0047
Cancel day15	0.0000	0.0000	0.0000
Cancel day16	0.1487	0.1060	0.1392

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Keep day1	0.9167	0.1293	0.9382
Keep day2	0.9984	0.1458	0.9984
Keep day3	0.9984	0.2355	0.9953
Keep day4	0.9984	0.2406	1.0000
Keep day5	0.9984	0.2425	0.9953
Keep day6	1.0000	0.2397	0.9984
Keep day7	0.9984	0.2445	0.9953
Keep day8	0.9984	0.2480	0.9937
Keep day9	1.0000	0.2421	1.0000
Keep day10	1.0000	0.2832	0.9984
Keep day11	1.0000	0.2848	1.0000
Keep day12	1.0000	0.2975	0.9968
Keep day13	1.0000	0.3244	0.9984
Keep day14	1.0000	0.3544	0.9953
Keep day15	1.0000	0.3892	1.0000
Keep day16	0.2342	0.3323	0.2722
D		**22.2140**	**0.1760**

8.6 Experiment with Real Ticket Reservation Data

8.6.1 Data Construction

Observations relative to ticket exchange and cancellation from 696 individuals were obtained from internet purchases of railway intercity tickets mentioned in Section 8.2. Based on the trip schedule of real data, ticket fare of the original departure time and other departure times within the same departure day are constructed for each day over the decision horizon based on historical data. In each time period, if the passenger decides to change or cancel the current ticket, then the same passenger will no longer be in the decision process in the next time period; all observations occurred in the period after ticket exchange are excluded from the dataset. This results in a total of 7,268 observed decisions that are valid for model estimation.

8.6.2 Model Specification

The specification follows the same structure proposed in the simulated data experiment. The model specification considers 16 discrete time periods defined by $t \in \{0,1,\dots,15\}$ where t also represents the number of day from original ticket purchase. The first time period is the day when original ticket is purchased $(t = 0)$, (day1). The last time period is departure day $(t = 15)$, (day16) The utility specification is defined as follows:

$$
\left.
\begin{aligned}
&U_{i5t=}\beta_{cost}(f_{5t} - f_{b0}) + \beta_{dfi\_exc}t + \varepsilon_{i5t} + \varepsilon_i \\
&\dots\dots\dots\dots\dots\dots\dots\dots \\
&U_{ijt=}\beta_{cost}\left(f_{jt} - f_{b0}\right) + \beta_{dfi\_exc}t + \varepsilon_{ijt} + \varepsilon_i \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&U_{i19t=}\ \beta_{cost}(f_{19t} - f_{b0}) + \beta_{dfi\_exc}t + \varepsilon_{i19t} + \varepsilon_i
\end{aligned}
\right\} \text{15 Exchange alternatives.}
\tag{8.19}
$$

$$
\begin{aligned}
U_{ict=}\ ASC_{cnl} &+ \beta_{gp}gp + \beta_{mr}mr + \beta_{ev}ev + \beta_{Mon}Mon + \beta_{Fri}Fri \\
&+ \beta_{STA1}STA1 + \beta_{STA3}STA3 + \beta_{ref}f_{b0} + \beta_{dfi\_cnl}t + \varepsilon_{ict} + \varepsilon_i
\end{aligned}
$$

$$
U_{ikt=}\begin{cases} c + \varepsilon_{ikt} + \varepsilon_i & \text{if } t = 15 \\ \varepsilon_{ikt} + \varepsilon_i & \text{if } t < 15 \end{cases}
$$

The utility of individual i on alternative j is denoted by U_{ijt}. For ticket exchange decision, the index j indicates 15 exchange departure time (5:00 AM to 7:00 PM). The utility of exchange (U_{ijt}) includes exchange cost which is defined as the difference between the original fare (f_{b0}) and new fare (f_{jt}) at time t, and day from issue (dfi) which is the number of day from original ticket purchase equal to t where $t = 0$ on the day of original purchase and $t = 15$ on departure day. The utility of cancel (U_{ict})

includes alternative specific constant (ASC), refund, dummy of group traveler, dummies of original departure in the morning (5:00-9:00 AM.) and evening (3:00-7:00 PM.), dummies of original departure on Monday and Friday, dummies of STA1 and STA3 destination. The utility of keep (U_{ikt}) is defined in two cases. In the last time period $(t = 15)$ passenger deciding to keep the ticket obtain an utility that includes the constant term relative to the utility of traveling with the original ticket. In other time periods $(t < 15)$ the systematic term of the keep utility is normalized to zero. ε_{ijt} is the random error term for each individual i, alternative j at a given time period t. ε_i is the individual error term which is assumed to be constant across all observations produced by the same respondent.

8.6.3 Estimation Result

The results obtained from model estimation are shown in Table 8.4. Most of the variables are statistically significant at the 5% significance level. The results obtained from the dynamic model shows negative sign in a number of variables associated with cancel decision which are: group traveler (party size includes more than one passenger), evening departure (original departure time from 3:00-7:00 PM.), original departure on Friday, and STA1 destination. This indicates low tendency of passenger with these characteristics to cancel their ticket. On the other hand, passengers with morning departure (original departure time from 5:00-9:00 AM.), original departure on Monday, and STA3 destination have a positive sign for the corresponding structural coefficients, indicating that passengers with these characteristics have higher likelihood to cancel the ticket. In particular, passengers traveling early in the week and traveling alone (typically associated with business travelers) are more likely to cancel their ticket which is in line with the results of Iliescu (2008).

The exchange cost and refund have the expected sign indicating disutility associated with paying additional cost to exchange ticket and the utility of receiving refund when ticket is canceled respectively. The variable of keeping the ticket on departure day (day16) shows negative sign which could be explained by the fact that the fare of the original ticket possessed by the passenger is higher compared to a ticket hypothetically exchanged to other departure times. Another reason could be that passengers intentionally want to exchange/cancel the ticket but could not find an alternative departure time which economically matches their schedule.

The day from issues (number of days since the original ticket is purchased) has positive sign for the variable associated with exchange and cancel decision; this indicates that it is preferable for passengers to adjust their ticket later. This is line with expectations and consistent with results obtained by Iliescu (2008) who found that the odds of ticket change increase as the departure date approaches due to a strong effect of “last minute” change of plan. More specifically, the day from issue coefficient for the cancel decision has larger magnitude compared to the day from issue coefficient for the exchange decisions. This is intuitive based on this operator’s refund policy; passengers are fully refunded if the ticket is canceled at least one hour before departure, while late tickets exchange are possible but limited by the uncertainty about seats availability.

The dummy variables of cancel on the original purchase date (day1) and exchange on the departure day (day16) show large magnitude indicating that a high number of cancellation and exchange occurs on the day they purchase ticket and on the departure day respectively. These results are in line with Iliescu (2008) and Graham et al. (2010) which found that ticket changes are more likely to happen in recently purchased ticket (especially within the first week) and are more likely to occur as the departure date approaches. Finally, the variable associated with early exchange (exchanging to departure time earlier than original ticket) shows negative sign which indicates that passengers gain less utility when making early exchange compared to later exchange (which is the base case). In term of goodness of fit, although the rho-squared value obtained from the dynamic model is lower than the MNL model (0.2791 compared to 0.6295); but the analysis in the next section shows that the dynamic model is superior to MNL model in term of prediction capability.

Table 8.4 Estimation result: real data

	Exchange	Cancel	Keep	MNL Est	MNL T-stat		Dynamic (2-SL) Est	Dynamic (2-SL) T-stat	
ASC cancel		x		-6.297	12.9	*	-3.652	57.1	*
>1 psg		x		-0.869	2.1	*	-1.090	1.5	
Orig Deptt 5-9 am		x		0.143	0.8		0.639	1.2	
Orig Deptt 3-7 pm		x		-0.327	1.9		-0.760	1.4	
Depart Monday		x		0.556	1.8		2.740	3.0	*
Depart Friday		x		-0.286	1.8		-0.451	1.0	
STA1 destination		x		-0.435	2.3	*	-0.306	0.6	
STA3 destination		x		0.557	2.5	*	1.648	2.6	*
Exchange cost	x			-0.011	19.3	*	-0.026	3.7	*
Refund		x		0.014	6.0	*	0.042	9.6	*
Keep (day 16)			x	1.885	11.0	*	-3.547	12.8	*
Day from issue	x			-1.217	35.3	*	0.189	5.8	*
Day from issue		x		0.163	5.9	*	0.266	35.4	*
Cancel (day 1)		x		5.629	18.3	*	3.169	42.7	*
Exchange (day 16)	x			17.050	30.5		1.578	10.2	*
Early exchange	x			-3.299	24.5	*	-1.751	12.1	*
Log-likelihood (0)					-20,592			-4,324	
Log-likelihood (final)					-7,629			-3,117	
Likelihood ratio index (rho-squared)					0.6295			0.2791	
No. of individuals					696			696	
No. of observations					7,268			7,268	

* Statistically significant at the 5% significance level.

8.6.4 Model Validation

To test the prediction capabilities of the model proposed, the resulting coefficients of the model have been used to replicate the choice observed in the sample. Results are reported in Table 8.5.

Table 8.5 Validation result: real data

Choice	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Observed	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Exchange	MNL	5	3	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	DDCM	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	Observed	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Exchange	MNL	6	3	1	1	0	0	0	0	0	0	0	0	0	0	0	1
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Observed	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
Exchange	MNL	6	4	2	1	0	0	0	0	0	0	0	0	0	0	0	1
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Observed	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2
Exchange	MNL	13	7	3	1	0	0	0	0	0	0	0	0	0	0	0	2
	DDCM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
5	Observed	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Exchange	MNL	20	12	5	2	1	0	0	0	0	0	0	0	0	0	0	3
	DDCM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Observed	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Exchange	MNL	24	14	6	2	1	0	0	0	0	0	0	0	0	0	0	3
	DDCM	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
7	Observed	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1
Exchange	MNL	29	19	7	2	1	0	0	0	0	0	0	0	0	0	0	4
	DDCM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8	Observed	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	9
Exchange	MNL	29	19	7	3	1	0	0	0	0	0	0	0	0	0	0	4
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Observed	3	1	0	0	0	0	0	2	0	0	0	0	0	3	0	9
Exchange	MNL	29	18	8	3	1	0	0	0	0	0	0	0	0	0	0	4
	DDCM	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	Observed	1	0	0	0	1	0	0	0	3	0	0	0	0	0	0	6
Exchange	MNL	22	14	7	2	1	0	0	0	0	0	0	0	0	0	0	4
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Observed	2	2	1	1	0	0	0	1	0	0	0	0	1	0	0	12
Exchange	MNL	26	16	7	3	1	0	0	0	0	0	0	0	0	0	0	0
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	Observed	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	8
Exchange	MNL	32	21	9	3	1	0	0	0	0	0	0	0	0	0	0	5
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Choice	Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
13	Observed	4	1	0	0	0	0	0	1	1	0	1	0	1	0	0	5
Exchange	MNL	40	23	10	4	1	0	0	0	0	0	0	0	0	0	0	6
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Observed	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	3
Exchange	MNL	62	34	13	4	1	0	0	0	0	0	0	0	0	0	0	8
	DDCM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	Observed	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Exchange	MNL	95	154	194	213	221	218	219	214	213	200	203	190	183	175	167	36
	DDCM	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	Observed	181	6	8	1	6	2	8	4	6	0	5	4	5	6	20	25
Cancel	MNL	181	1	2	3	4	4	5	6	7	8	9	11	12	14	16	4
	DDCM	85	0	0	0	0	0	0	0	0	0	0	0	0	2	0	20
17	Observed	495	483	469	467	460	457	447	438	428	427	420	414	405	393	369	285
Keep	MNL	77	135	201	224	233	235	233	227	219	221	215	220	219	216	210	283
	DDCM	606	494	483	468	467	460	457	447	438	428	427	419	414	403	393	344
Total	Observed	20	6	6	1	1	1	2	5	4	1	2	2	4	6	4	59
Exchange	MNL	438	359	280	243	230	221	219	215	213	200	203	190	183	175	167	80
(11-15)	DDCM	5	1	0	1	0	0	0	0	0	0	0	1	0	0	0	5
	Total	**696**	**495**	**483**	**469**	**467**	**460**	**457**	**447**	**438**	**428**	**427**	**420**	**414**	**405**	**393**	**369**

Figure 8.6 to Figure 8.8 briefly summarize the predictions over different time periods (days) where exchange decisions are aggregated for all exchange departure times. The validation results show that the DDCM slightly under-predicts cancellation and although it is not able to predict the cancellation on the first time period (day1) as well as the MNL, it is capable of predicting cancellation on the last time period (day16) reasonably well. In term of exchange, DDCM slightly under-predicts the total number of exchange except for the first (day1) and the last time period (day16) which are characterized by a relatively high exchange rate; however, the MNL drastically over predicts exchange decisions throughout all time periods. The prediction of keep obtained

from DDCM is reasonably close to the observed value while the MNL significantly under predicts the keep decision as a consequence of over prediction in exchange.

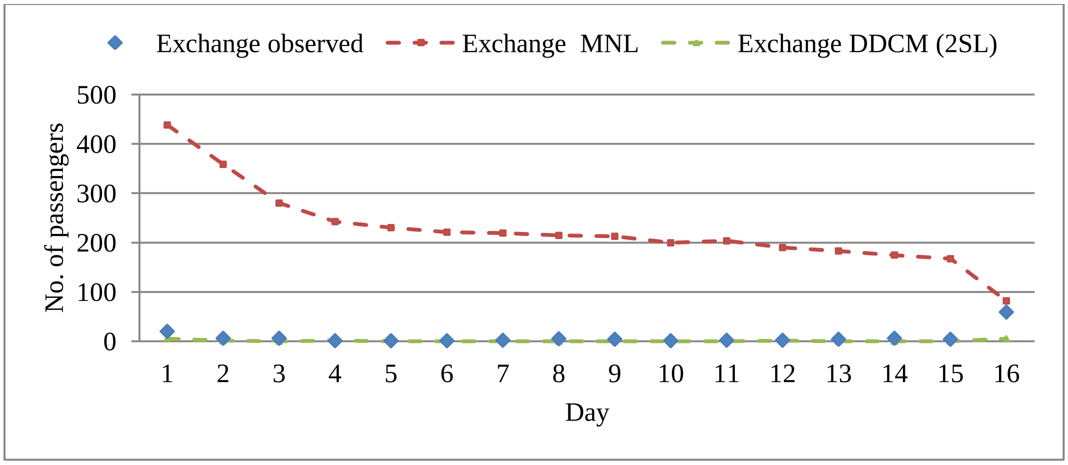


Figure 8.6 Validation of exchange decision: real data

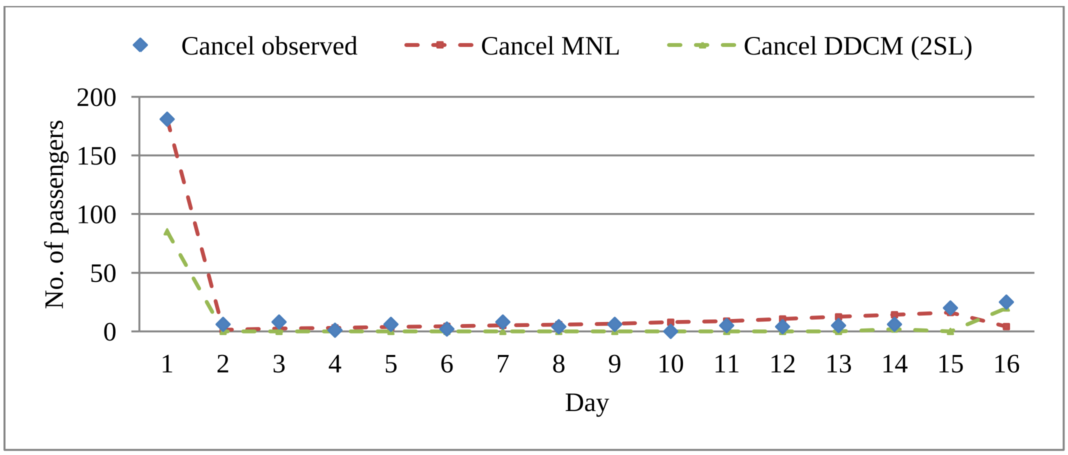


Figure 8.7 Validation of cancel decision: real data

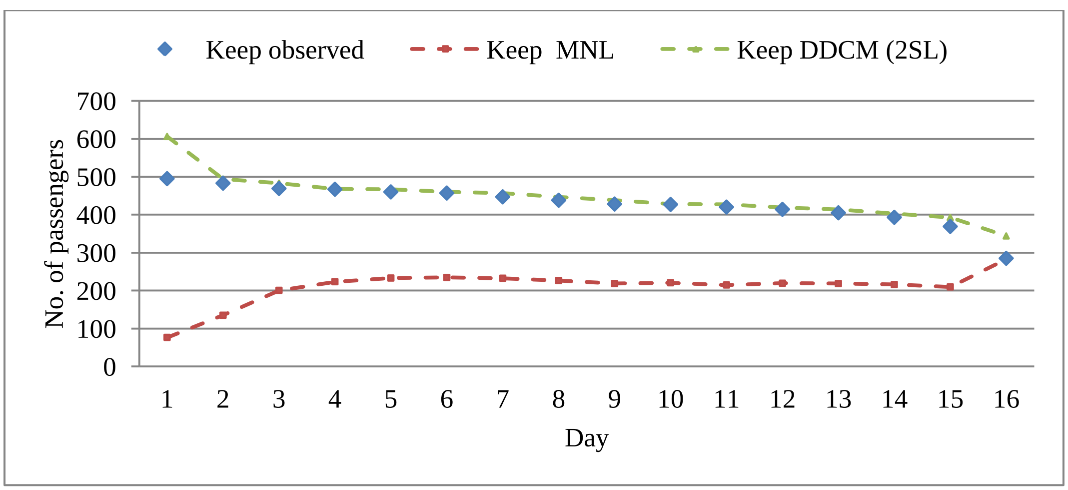


Figure 8.8 Validation of keep decision: real data

The choice probability for each alternative observed and predicted together with measure of errors for the real data experiment is reported in Table 8.6. It shows that the D value of the dynamic model is significantly smaller than the corresponding value obtained from the MNL model (1.194 compared to 15.179) indicating a much better prediction capability of the dynamic model over the MNL model.

Table 8.6 Model validation: choice probability of real data experiment

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Exchange day1	0.0287	0.6297	0.0072
Exchange day2	0.0121	0.7244	0.0020
Exchange day3	0.0124	0.5797	0.0000
Exchange day4	0.0021	0.5171	0.0021
Exchange day5	0.0021	0.4929	0.0000
Exchange day6	0.0022	0.4804	0.0000
Exchange day7	0.0044	0.4799	0.0000
Exchange day8	0.0112	0.4799	0.0000
Exchange day9	0.0091	0.4856	0.0000
Exchange day10	0.0024	0.4662	0.0001
Exchange day11	0.0047	0.4763	0.0000
Exchange day12	0.0048	0.4521	0.0024
Exchange day13	0.0097	0.4420	0.0000
Exchange day14	0.0148	0.4314	0.0000
Exchange day15	0.0102	0.4254	0.0000
Exchange day16	0.1599	0.2220	0.0136
Cancel day1	0.2601	0.2601	0.1221
Cancel day2	0.0121	0.0026	0.0000
Cancel day3	0.0166	0.0048	0.0000
Cancel day4	0.0021	0.0064	0.0000
Cancel day5	0.0128	0.0079	0.0000
Cancel day6	0.0043	0.0093	0.0000
Cancel day7	0.0175	0.0112	0.0000
Cancel day8	0.0089	0.0130	0.0000
Cancel day9	0.0137	0.0148	0.0000
Cancel day10	0.0000	0.0182	0.0000
Cancel day11	0.0117	0.0206	0.0000
Cancel day12	0.0095	0.0252	0.0000
Cancel day13	0.0121	0.0300	0.0000
Cancel day14	0.0148	0.0351	0.0049
Cancel day15	0.0509	0.0410	0.0000
Cancel day16	0.0677	0.0111	0.0542

Alternative	Observed	Predicted (Static)	Predicted (Dynamic)
Keep day1	0.7112	0.1102	0.8707
Keep day2	0.9758	0.2729	0.9980
Keep day3	0.9710	0.4155	1.0000
Keep day4	0.9957	0.4765	0.9979
Keep day5	0.9850	0.4991	1.0000
Keep day6	0.9935	0.5102	1.0000
Keep day7	0.9781	0.5090	1.0000
Keep day8	0.9799	0.5072	1.0000
Keep day9	0.9772	0.4995	1.0000
Keep day10	0.9976	0.5156	0.9999
Keep day11	0.9836	0.5030	1.0000
Keep day12	0.9857	0.5226	0.9976
Keep day13	0.9783	0.5280	1.0000
Keep day14	0.9704	0.5336	0.9951
Keep day15	0.9389	0.5336	1.0000
Keep day16	0.7723	0.7669	0.9322
D		**15.1790**	**1.1940**

8.7 Conclusions

This chapter has proposed a dynamic discrete choice model for ticket cancellation and exchange in the context of railway ticket purchase for intercity trips. The methodological framework proposed considers forward-looking agents that maximize their inter-temporal payoffs when deciding about exchanging or cancelling their ticket. The classical formulation based on the optimal stopping problem derived from dynamic programming is preserved here, while an innovative and elegant scenario tree formulation is proposed to solve the issue of calculating the passengers' expected utility over time. The model is estimated using maximum likelihood estimation, which seems particularly appropriated in this finite horizon problem. The analysis makes an important contribution in the context of discrete choice models for revenue management as it allows us to account for temporal effects on individual decisions that are usually treated in a static context. The model has been successfully estimated using both simulated and real data; results show that the DDCM outperforms the MNL in reproducing the initial values assumed in the simulated dataset and in reproducing the actual choices in both synthetic and real data.

Chapter 9 : Conclusions and Future Research Directions

This chapter concludes the dissertation by summarizing the important findings and the contributions from this research. We then discuss future avenues for this research and the possible extension to other problems in RM.

9.1 Conclusions

This dissertation has mainly investigated the improvement of prediction accuracy in choice models of ticket purchase timing accounting for taste heterogeneity (Chapter 5). The approach exploits elements available in the ticket reservation data to segment passengers primarily by departure time characteristics. The application of the proposed choice models to RM strategy is demonstrated by incorporating them into RM optimization models (Chapter 7) which optimize ticket revenue. The RM optimization models account for demand side in term of purchase timing (Chapter 5) and demand volume (Chapter 6) and simultaneously solve for pricing and seat allocation strategy under the capacity constraints determined on the basis of the railway network characteristics.

More importantly, this dissertation has investigated the impact of incorporating inter-temporal effects in ticket cancellation and exchange decision using a dynamic discrete choice model (DDCM) framework (Chapter 8). A dynamic econometric model accounting for realistic ticket exchange, refund policy, and the evolving characteristics of fare is developed. The timing of the ticket cancellations and exchanges is formulated as an optimal stopping problem where passengers decide when to adjust tickets. The model structure is further enriched by allowing for departure time specific exchange decisions based on alternatives offered by the railway service.

9.2 Contributions

The primary contributions from this research are:

1. The choice models accounting for taste heterogeneity are developed focusing on segmenting passengers between business and leisure travel. Instead of segmenting passengers by trip purpose, which is not available in the ticket reservation data, departure time characteristics are used to classify passenger into categories (classes). With this approach, passenger segments can be identified, and the differences in their behavioral preferences among classes can be captured.
2. To assess the impact of the proposed choice models in RM strategy, an optimization model system is developed to maximize revenues from ticket sale on single-leg and network problems. The framework incorporates parameters estimated from the choice models and demand functions as key inputs to represent passenger response to RM policy.
3. In the network problem, the optimizations allow for pricing and seat allocation to be optimized simultaneously under capacity constraints determined on the basis of the railway network characteristics. The framework allows for capacity redistribution; thus capacity resources are efficiently utilized across the selected network. The incorporated choice models accounting for heterogeneity allow the RM optimization to distinguish between leisure and business travelers, depending on departure time and day of week. Results show that seat allocation policy which

accepts more short-haul trips contributes to greater revenue than long-haul trips with the same seat capacity.

4. For the network revenue optimization problem, the mixed logit choice model with non-parametric distribution is proposed as an alternative approach to resolve difficulties associated with the identification of underlying unknown random distribution. Results show that the pricing strategy provided by mixed logit with non-parametric B-spline is more realistic and intuitive compared to other models based on passenger price sensitivity behavior.
5. This dissertation is the first study that develops a ticket cancellation and exchange model using dynamic discrete choice models (DDCM). Inter-temporal effects in ticket cancellation and exchange decisions are usually neglected or treated using static models incorporating lagged effects.
6. The proposed ticket cancellation and exchange model realistically represents exchange, refund, and fare policy of the railway service. Complementary to the existing studies, the proposed model enables us to predict not only the timing of cancellations and exchanges, but also the new departure times. Prediction of new departure times is expected to further support railway operator's RM capacity planning. The analysis based on simulated and real ticket reservation data indicates that the DDCM provides more intuitive results and better prediction accuracy than the multinomial logit (MNL) models.

9.3 Future Research Directions

This research has illustrated how a railway operator can exploit its existing data sources to better understand the choice behavior of railway passengers. In this dissertation, the passenger choice models of ticket purchase timing capable of supporting railway RM decision are developed and incorporated in the RM strategy. The dynamic discrete choice model (DDCM) of ticket cancellation and exchange is developed to investigate the impact of inter-temporal effects in ticket cancellation and exchange decision. It is expected that these frameworks will be extended to other RM applications. The following areas indicate possible avenues for future research:

For the dynamic model of ticket cancellation and exchange, in term of model specification, it would be desirable to allow for the fare to have dynamic attribute in the utility specification by specifying the process as a random walk. It is interesting to account for heterogeneity of passengers in the model by incorporating the unobservable (or latent class) segments within the population using latent class (LC) approach. It will also be interesting for the proposed DDCM framework to be applied to ticket purchase timing behavior and compare its performance with the static choice models of ticket purchase timing proposed in this dissertation.

From the dynamic model estimation perspective, it will be desirable to examine the impact of number of steps used in the look-ahead policy toward the model performance. Currently, the framework allows for the passengers to look ahead for two time periods. Increasing the number of steps in the look-ahead policy is expected to improve model performance.

It will be desirable for the proposed dynamic model to be incorporated in the revenue optimization where decision variables include capacity planning parameters such as demand acceptance to test the impact of the dynamic model toward revenue and explore the possibility of adopting overbooking strategy used in the airline.

Future research on the choice model which considers other modes of transportation is desirable. This involves data collection on other transportation modes' service attributes. This will be able to represent impact from attributes of other transportation modes toward passenger demand.

Future research which investigates other choice dimensions of the railway RM will be desirable, for instance, choice of departure time or departure day when making ticket purchase. However, for these choice dimensions, additional data collection to construct plausible choice set for individual might be necessary.

From the RM revenue optimization perspective, it will be desirable to consider the network with hub and spoke characteristic which involves station transfer and more complex capacity constraints. It will also be interesting to optimize ticket revenue over multiple departures simultaneously by accounting for demand shifts across departures with choice models of departure time or departure day, thereby efficiently balancing passenger demand and improve total revenue.

More specifically the dynamic discrete choice framework developed in this dissertation can be adapted and transferred to other case studies: departure time and route choice modeling under dynamic tolling, activity scheduling for activity based travel demand analysis.

Appendices

Appendix A. Choice models accounting for taste heterogeneity

Table A-1 Model 1 Result

	MNL			ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.129	7.851	*	-0.198	17.328	*
price.period1	-0.001	0.244		-9.218	6.708	*
price.period2	-0.005	1.618		(6.957)	(6.116)	*
price.period3	-0.007	2.305	*			
price.period4	-0.006	2.247	*			
price.period5	-0.003	1.283				
price.period6	0.000	0.002				
wknd.period1	0.822	4.768	*	2.058	10.601	*
wknd.period2	0.909	5.527	*	1.004	7.519	*
wknd.period3	0.605	2.983	*	0.153	1.289	
wknd.period4	0.592	4.482	*	-0.058	0.686	
wknd.period5	0.287	1.570		-0.100	0.929	
wknd.period6	-0.215	1.445		-0.058	0.388	
No. of observations		1,361			1,361	
Rho-squared:		0.1485			0.1940	
Adjusted rho-squared:		**0.1457**			**0.1920**	
Log-likelihood at optimal		-3,980			-3,767	
Log-likelihood at zero		-4,674			-4,674	
LL at constant		-3,988			-3,988	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.091	-4.553	*	-0.332	-1.934	
price.period1	-0.003	-0.702		0.009	0.275	
price.period2	-0.005	-1.000		-0.001	-0.073	
price.period3	-0.006	-1.222		-0.003	-0.250	
price.period4	-0.005	-1.045		-0.004	-0.606	
price.period5	-0.004	-0.818		0.000	-0.062	
price.period6	-0.005	-1.116		0.005	0.727	
Class Model	**Class1**			**Class2**		
Class Size	0.723			0.277		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	1.006	2.477	*	-1.006	-2.477	*
Monday	-1.145	-2.908	*	1.145	2.908	*
Tuesday	-1.310	-2.895	*	1.310	2.895	*
Wednesday	-0.856	-2.480	*	0.856	2.480	*
Thursday	-0.517	-1.490		0.517	1.490	
Friday	-0.703	-2.013	*	0.703	2.013	*
Saturday	1.049	0.285		-1.049	-0.285	
Early morning	0.491	1.238		-0.491	-1.238	
AM peak	1.195	1.509		-1.195	-1.509	
AM off peak	0.731	2.147	*	-0.731	-2.147	*
PM off peak	-0.048	-0.216		0.048	0.216	
PM peak	-0.335	0.000		0.335	0.000	
No. of observations					1,361	
Rho-squared:					0.1563	
Adjusted rho-squared:					**0.1507**	
Log-likelihood at optimal					-3,943	
Log-likelihood at zero					-4,674	
LL at constant					-3,988	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.191	16.589	*
price.adult	-0.032	9.420	*
price.child (2-15)	-	-	-
price.senior (62+)	-0.024	4.442	*
price.unacc child (8-11)	-	-	-
price.student advantage	-0.026	7.526	*
price.adultAAA member	-0.021	4.750	*
price.childAAA member	-	-	-
price.military adult	-	-	-
price.disabled adult	0.006	0.209	
price.others	-0.051	11.458	*
price.period6	-0.027	7.941	*
wknd.period1	1.096	4.242	*
wknd.period2	0.730	4.176	*
wknd.period3	0.294	0.915	
wknd.period4	0.385	2.269	*
wknd.period5	0.555	1.567	
wknd.period6	-0.059	0.272	
No. of observations	1,361		
Rho-squared:		0.1552	
Adjusted rho-squared:		**0.1513**	
Log-likelihood at optimal		-3,948	
Log-likelihood at zero		-4,674	
LL at constant		-3,988	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-2 Model 2 Result

MNL				ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.184	42.872	*	-0.370	141.708	*
price.period1	-0.006	3.634	*	-5.318	77.001	*
price.period2	-0.012	9.173	*	(2.830)	(60.153)	*
price.period3	-0.011	10.783	*			
price.period4	-0.010	10.380	*			
price.period5	-0.005	5.287	*			
price.period6	-0.003	3.366	*			
wknd.period1	1.240	25.231	*	1.897	20.459	*
wknd.period2	1.099	27.923	*	0.681	11.363	*
wknd.period3	0.407	13.106	*	0.347	8.765	*
wknd.period4	0.466	14.830	*	0.041	0.977	
wknd.period5	-0.472	13.477	*	-0.483	10.257	*
wknd.period6	0.260	9.369	*	0.516	11.436	*
No. of observations		37,373			37,373	
Rho-squared:		0.2970			0.3397	
Adjusted rho-squared:		**0.2969**			**0.3396**	
Log-likelihood at optimal		-90,226			-84,742	
Log-likelihood at zero		-128,338			-128,338	
LL at constant		-90,487			-90,487	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.139	-21.091	*	-0.899	-18.162	*
price.period1	0.000	0.111		0.014	1.560	
price.period2	-0.003	-1.684		-0.082	-2.069	*
price.period3	-0.005	-2.778	*	-0.049	-5.557	*
price.period4	-0.001	-0.737		-0.069	-7.931	*
price.period5	0.003	1.789		-0.077	-8.052	*
price.period6	-0.010	-4.348	*	-0.059	-6.941	*
Class Model	**Class1**			**Class2**		
Class Size	0.619			0.381		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.181	4.845	*	-0.181	-4.845	*
Monday	-0.402	-14.288	*	0.402	14.288	*
Tuesday	-0.338	-11.586	*	0.338	11.586	*
Wednesday	-0.375	-11.908	*	0.375	11.908	*
Thursday	-0.286	-9.502	*	0.286	9.502	*
Friday	-0.213	-7.690	*	0.213	7.690	*
Saturday	-0.019	-0.433		0.019	0.433	
Early morning	1.085	19.204	*	-1.085	-19.204	*
AM peak	0.985	22.737	*	-0.985	-22.737	*
AM off peak	0.474	15.353	*	-0.474	-15.353	*
PM off peak	0.096	3.491	*	-0.096	-3.491	*
PM peak	0.113	4.036	*	-0.113	-4.036	*
No. of observations					37,373	
Rho-squared:					0.3034	
Adjusted rho-squared:					**0.3032**	
Log-likelihood at optimal					-89,402	
Log-likelihood at zero					-128,338	
LL at constant					-90,487	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.248	43.103	*
price.adult	-0.024	10.139	*
price.child (2-15)	-0.004	0.279	
price.senior (62+)	-0.023	9.401	*
price.unacc child (8-11)	0.001	0.072	
price.student advantage	-0.010	2.838	*
price.adultAAA member	-0.010	3.720	*
price.childAAA member	-	-	-
price.military adult	-0.028	7.638	*
price.disabled adult	-0.024	4.788	*
price.others	-0.043	17.059	*
price.period6	-0.019	8.249	*
wknd.period1	2.012	40.426	*
wknd.period2	0.878	34.139	*
wknd.period3	-0.044	1.492	
wknd.period4	0.050	1.751	
wknd.period5	-0.133	4.703	*
wknd.period6	0.236	7.935	*
No. of observations		37,373	
Rho-squared:		0.2904	
Adjusted rho-squared:		**0.2903**	
Log-likelihood at optimal		-91,070	
Log-likelihood at zero		-128,338	
LL at constant		-90,487	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-3 Model 3 Result

	MNL			ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.242	33.967	*	-0.372	105.968	*
price.period1	0.002	1.224		-7.130	49.790	*
price.period2	-0.009	6.020	*	(4.668)	(43.775)	*
price.period3	-0.010	9.128	*			
price.period4	-0.011	11.796	*			
price.period5	-0.007	8.166	*			
price.period6	-0.003	3.244	*			
wknd.period1	1.440	16.329	*	2.087	9.865	*
wknd.period2	1.161	17.137	*	0.207	1.269	
wknd.period3	0.311	4.461	*	0.193	2.343	*
wknd.period4	0.381	7.161	*	0.091	0.968	
wknd.period5	-0.449	6.181	*	-0.465	4.209	*
wknd.period6	0.155	3.433	*	0.887	9.297	*
No. of observations		19,032			19,032	
Rho-squared:		0.3557			0.3899	
Adjusted rho-squared:		**0.3555**			**0.3898**	
Log-likelihood at optimal		-42,106			-39,874	
Log-likelihood at zero		-65,356			-65,356	
LL at constant		-41,704			-41,704	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.197	-32.279	*	-1.710	-0.002	
price.period1	-0.027	-10.192	*	0.121	4.389	*
price.period2	-0.034	-13.480	*	0.034	0.974	
price.period3	-0.032	-14.000	*	-0.008	-0.244	
price.period4	-0.030	-13.591	*	-0.104	-2.665	*
price.period5	-0.026	-11.217	*	-0.117	-2.714	*
price.period6	-0.034	-15.545	*	-0.051	-1.769	

Class Model	**Class1**			**Class2**		
Class Size	0.617			0.383		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.457	8.606	*	-0.457	-8.606	*
Monday	-0.396	-8.689	*	0.396	8.689	*
Tuesday	-0.375	-8.547	*	0.375	8.547	*
Wednesday	-0.358	-8.061	*	0.358	8.061	*
Thursday	-0.332	-7.578	*	0.332	7.578	*
Friday	-0.215	-4.851	*	0.215	4.851	*
Saturday	-0.029	-0.396		0.029	0.396	
Early morning	0.742	10.146	*	-0.742	-10.146	*
AM peak	0.715	13.463	*	-0.715	-13.463	*
AM off peak	0.155	3.992	*	-0.155	-3.992	*
PM off peak	-0.143	-4.313	*	0.143	4.313	*
PM peak	0.051	1.550		-0.051	-1.550	
No. of observations					19,032	
Rho-squared:					0.3688	
Adjusted rho-squared:					**0.3684**	
Log-likelihood at optimal					-41,255	
Log-likelihood at zero					-65,356	
LL at constant					-41,704	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.382	78.332	*
price.adult	-0.080	49.469	*
price.child (2-15)	-0.006	0.235	
price.senior (62+)	-0.075	30.008	*
price.unacc child (8-11)	-0.688	18.544	*
price.student advantage	-0.054	12.524	*
price.adultAAA member	-0.061	28.722	*
price.childAAA member	-	-	-
price.military adult	-0.082	10.460	*
price.disabled adult	-0.069	8.539	*
price.others	-0.102	46.711	*
price.period6	-0.073	45.156	*
wknd.period1	2.653	29.238	*
wknd.period2	1.034	16.240	*
wknd.period3	-0.024	0.330	
wknd.period4	-0.175	3.260	*
wknd.period5	-0.522	5.994	*
wknd.period6	0.034	0.702	
No. of observations		19,032	
Rho-squared:		0.3602	
Adjusted rho-squared:		**0.3599**	
Log-likelihood at optimal		-41,814	
Log-likelihood at zero		-65,356	
LL at constant		-41,704	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-4 Model 4 Result

MNL				ML			LC							MNL with Socioeconomics			
							Choice Model	**Class1**			**Class2**						
Variable	Est	T-Stat		Est	T-Stat		Variable	Est	T-Stat		Est	T-Stat		Variable	Est	T-Stat	
advbk	-0.141	15.520	*	-0.314	52.483	*	advbk	-0.086	-6.378	*	-0.500	-9.411	*	advbk	-0.192	20.516	*
price.period1	-0.017	4.707	*	-5.688	46.679	*	price.period1	-0.013	-2.761	*	0.032	2.121	*	price.adult	-0.023	5.198	*
price.period2	-0.021	5.909	*	(4.536)	(53.574)	*	price.period2	-0.013	-2.660	*	-0.001	-0.070		price.child (2-15)	-	-	-
price.period3	-0.019	5.892	*				price.period3	-0.012	-2.818	*	0.001	0.061		price.senior (62+)	-0.025	4.552	*
price.period4	-0.017	5.675	*				price.period4	-0.009	-2.140	*	-0.011	-0.991		price.unacc child (8-11)	-	-	-
price.period5	-0.013	4.299	*				price.period5	-0.007	-1.512		-0.011	-0.982		price.student advantage	-0.018	4.093	*
price.period6	-0.009	3.236	*				price.period6	-0.017	-2.609	*	-0.004	-0.410		price.adultAAA member	-0.018	3.756	*
														price.childAAA member	-	-	-
							Class Model	**Class1**			**Class2**			price.military adult	0.081	1.225	
							Class Size	0.574			0.426			price.disabled adult	5.460	6.934	*
							Variable	Est	T-Stat		Est	T-Stat		price.others	-0.040	8.479	*
							Intercept	0.987	3.384	*	-0.987	-3.384	*	**price.period6**	-0.017	3.925	*
							Monday	-1.418	-6.349	*	1.418	6.349	*				
wknd.period1	1.282	12.189	*	3.329	27.284	*	Tuesday	-1.224	-5.848	*	1.224	5.848	*	wknd.period1	1.614	16.890	*
wknd.period2	1.118	12.777	*	1.428	19.718	*	Wednesday	-1.321	-6.015	*	1.321	6.015	*	wknd.period2	0.935	13.166	*
wknd.period3	0.650	7.678	*	0.298	4.862	*	Thursday	-1.443	-6.325	*	1.443	6.325	*	wknd.period3	0.428	6.034	*
wknd.period4	0.622	8.762	*	-0.304	5.237	*	Friday	-1.150	-5.589	*	1.150	5.589	*	wknd.period4	0.400	5.993	*
wknd.period5	0.013	0.129		-0.735	7.376	*	Saturday	-0.716	-2.674	*	0.716	2.674	*	wknd.period5	0.307	2.963	*
wknd.period6	-0.684	6.821	*	-1.015	11.171	*								wknd.period6	-0.685	5.449	*
							Early morning	-	-		-	-					
							AM peak	0.974	4.980	*	-0.974	-4.980	*				
							AM off peak	0.667	3.779	*	-0.667	-3.779	*				
							PM off peak	0.077	0.476		-0.077	-0.476					
							PM peak	-0.113	-0.677		0.113	0.677					
No. of observations		4,472			4,472		No. of observations					4,472		Number of observations		4,472	
Rho-squared:		0.2078			0.2889		Rho-squared:					0.2122		Rho-squared:		0.2074	
Adjusted rho-squared:		**0.2069**			**0.2883**		**Adjusted rho-squared:**					**0.2105**		**Adjusted rho-squared:**		**0.2062**	
Log-likelihood at optimal		-12,166			-10,920		Log-likelihood at optimal					-12,097		Log likelihood at optimal		-12,172	
Log-likelihood at zero		-15,357			-15,357		Log-likelihood at zero					-15,357		Log-likelihood at zero		-15,357	
LL at constant		-12,330			-12,330		LL at constant					-12,330		LL at constant		-12,330	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-5 Model 5 Result

Variable	MNL Est	MNL T-Stat		ML Est	ML T-Stat	
advbk	-0.220	17.952	*	-0.450	89.143	*
price.period1	-0.008	1.942		-6.225	54.812	*
price.period2	-0.017	6.735	*	(3.630)	(56.304)	*
price.period3	-0.015	8.598	*			
price.period4	-0.013	9.154	*			
price.period5	-0.004	3.109	*			
price.period6	-0.002	1.533				
wknd.period1	1.473	8.601	*	0.468	12.337	*
wknd.period2	0.833	5.862	*	0.491	42.742	*
wknd.period3	0.396	3.485	*	0.518	36.889	*
wknd.period4	0.352	4.280	*	0.355	6.400	*
wknd.period5	-0.551	5.574	*	-0.163	3.125	*
wknd.period6	0.498	7.342	*	1.332	23.902	*
No. of observations		11,536			11,536	
Rho-squared:		0.4327			0.4525	
Adjusted rho-squared:		**0.4323**			**0.4523**	
Log-likelihood at optimal		-22,474			-21,688	
Log-likelihood at zero		-39,614			-39,614	
LL at constant		-22,530			-22,530	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.199	-11.688	*	-0.539	-12.988	*
price.period1	-0.083	-10.519	*	0.023	1.497	
price.period2	-0.084	-11.562	*	-0.010	-0.746	
price.period3	-0.084	-12.478	*	-0.012	-0.950	
price.period4	-0.076	-11.299	*	-0.024	-2.042	*
price.period5	-0.106	-6.418	*	-0.015	-1.235	
price.period6	-0.059	-8.864	*	-0.035	-2.974	*
Class Model	**Class1**			**Class2**		
Class Size	0.575			0.426		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.785	8.454	*	-0.785	-8.454	*
Monday	0.186	2.215	*	-0.186	-2.215	*
Tuesday	-0.408	-5.255	*	0.408	5.255	*
Wednesday	-0.413	-5.272	*	0.413	5.272	*
Thursday	-0.456	-5.764	*	0.456	5.764	*
Friday	-0.317	-3.977	*	0.317	3.977	*
Saturday	0.180	1.163		-0.180	-1.163	
Early morning	-	-		-	-	
AM peak	-0.917	-11.348	*	0.917	11.348	*
AM off peak	-0.305	-5.495	*	0.305	5.495	*
PM off peak	-0.009	-0.182		0.009	0.182	
PM peak	-0.190	-3.484	*	0.190	3.484	*
No. of observations					11,536	
Rho-squared:					0.4444	
Adjusted rho-squared:					**0.4437**	
Log-likelihood at optimal					-22,011	
Log-likelihood at zero					-39,614	
LL at constant					-22,530	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.517	58.129	*
price.adult	-0.123	36.196	*
price.child (2-15)	26.057	2.985	*
price.senior (62+)	-0.116	25.077	*
price.unacc child (8-11)	-	-	-
price.student advantage	-0.072	15.558	*
price.adultAAA member	-0.090	17.187	*
price.childAAA member	-	-	-
price.military adult	0.406	7.946	*
price.disabled adult	-0.075	20.376	*
price.others	-0.151	29.506	*
price.period6	-0.118	34.320	*
wknd.period1	3.522	51.264	*
wknd.period2	0.908	3.191	*
wknd.period3	-0.125	0.435	
wknd.period4	-0.502	1.939	
wknd.period5	-0.657	2.299	*
wknd.period6	0.129	0.511	
Number of observations		11,536	
Rho-squared:		0.4212	
Adjusted rho-squared:		**0.4207**	
Log likelihood at optimal		-22,930	
Log-likelihood at zero		-39,614	
LL at constant		-22,530	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-6 Model 6 Result

MNL				ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.160	30.287	*	-0.392	100.986	*
price.period1	-0.018	6.775	*	-5.421	35.523	*
price.period2	-0.024	9.831	*	(4.252)	(29.595)	*
price.period3	-0.021	9.799	*			
price.period4	-0.018	8.377	*			
price.period5	-0.012	5.593	*			
price.period6	-0.008	4.074	*			
wknd.period1	1.163	21.812	*	4.299	57.340	*
wknd.period2	1.203	29.228	*	1.747	43.787	*
wknd.period3	0.751	19.745	*	0.168	4.589	*
wknd.period4	0.537	17.042	*	-0.664	20.754	*
wknd.period5	-0.149	3.310	*	-1.187	27.385	*
wknd.period6	-0.504	13.923	*	-1.353	30.550	*
No. of observations		29,514			29,514	
Rho-squared:		0.2779			0.3500	
Adjusted rho-squared:		**0.2778**			**0.3499**	
Log-likelihood at optimal		-73,182			-65,877	
Log-likelihood at zero		-101,351			-101,351	
LL at constant		-73,807			-73,807	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.084	-10.870	*	-0.672	-16.043	*
price.period1	0.017	3.864	*	-0.014	-0.917	
price.period2	0.018	3.851	*	-0.130	-2.310	*
price.period3	0.018	3.826	*	-0.073	-4.953	*
price.period4	0.022	4.696	*	-0.087	-5.805	*
price.period5	0.029	6.196	*	-0.093	-6.019	*
price.period6	0.010	1.792		-0.077	-5.371	*
Class Model	**Class1**			**Class2**		
Class Size	0.566			0.434		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.952	8.602	*	-0.952	-8.602	*
Monday	-1.212	-13.484	*	1.212	13.484	*
Tuesday	-1.061	-12.027	*	1.061	12.027	*
Wednesday	-1.161	-12.777	*	1.161	12.777	*
Thursday	-1.196	-13.013	*	1.196	13.013	*
Friday	-0.859	-11.172	*	0.859	11.172	*
Saturday	-0.837	-9.565	*	0.837	9.565	*
Early morning	0.977	11.940	*	-0.977	-11.940	*
AM peak	0.852	14.129	*	-0.852	-14.129	*
AM off peak	0.428	8.998	*	-0.428	-8.998	*
PM off peak	0.109	2.949	*	-0.109	-2.949	*
PM peak	-0.076	-2.027	*	0.076	2.027	*
No. of observations					29,514	
Rho-squared:					0.2816	
Adjusted rho-squared:					**0.2814**	
Log-likelihood at optimal					-72,809	
Log-likelihood at zero					-101,351	
LL at constant					-73,807	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.293	69.420	*
price.adult	-0.064	27.225	*
price.child (2-15)	-0.071	5.034	*
price.senior (62+)	-0.063	24.982	*
price.unacc child (8-11)	-0.063	10.836	*
price.student advantage	-0.054	19.977	*
price.adultAAA member	-0.057	23.248	*
price.childAAA member	-	-	-
price.military adult	0.396	0.786	
price.disabled adult	-0.063	7.736	*
price.others	-0.088	34.916	*
price.period6	-0.058	24.431	*
wknd.period1	1.582	30.329	*
wknd.period2	0.855	23.879	*
wknd.period3	0.389	12.255	*
wknd.period4	0.398	12.808	*
wknd.period5	0.296	6.590	*
wknd.period6	-0.520	11.507	*
No. of observations		29,514	
Rho-squared:		0.2760	
Adjusted rho-squared:		**0.2758**	
Log-likelihood at optimal		-73,383	
Log-likelihood at zero		-101,351	
LL at constant		-73,807	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-7 Model 7 Result

MNL				ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.224	4.743	*	-1.528	27.801	*
price.period1	-0.313	6.114	*	-1.547	9.163	*
price.period2	-0.347	8.573	*	(1.789)	(22.066)	*
price.period3	-0.340	9.775	*			
price.period4	-0.300	10.084	*			
price.period5	-0.264	9.044	*			
price.period6	-0.209	7.382	*			
wknd.period1	2.469	3.299	*	18.240	14.795	*
wknd.period2	-5.359	7.223	*	-0.634	7.402	*
wknd.period3	1.738	3.367	*	-0.211	0.379	
wknd.period4	1.495	3.441	*	-4.044	7.799	*
wknd.period5	0.620	0.937		-5.987	7.481	*
wknd.period6	2.144	2.990	*	-4.360	6.780	*
No. of observations		820			820	
Rho-squared:		0.7597			0.7733	
Adjusted rho-squared:		**0.7550**			**0.7701**	
Log-likelihood at optimal		-677			-642	
Log-likelihood at zero		-2,816			-2,816	
LL at constant		NA			NA	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.145	-0.880		-0.057	-0.406	
price.period1	-0.228	-1.172		0.693	2.660	*
price.period2	-0.204	-1.059		0.655	2.374	*
price.period3	-0.391	-1.031		0.681	2.358	*
price.period4	-0.171	-0.829		0.655	2.333	*
price.period5	-0.116	-0.552		0.661	2.351	*
price.period6	-0.065	-0.313		0.706	2.552	*
Class Model	**Class1**			**Class2**		
Class Size	0.629			0.372		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	-0.875	-0.792		0.875	0.792	
Monday	2.777	1.342		-2.777	-1.342	
Tuesday	0.894	0.748		-0.894	-0.748	
Wednesday	-0.916	-0.599		0.916	0.599	
Thursday	-1.087	-0.721		1.087	0.721	
Friday	-2.709	-1.421		2.709	1.421	
Saturday	-2.528	-0.933		2.528	0.933	
Early morning	-	-		-	-	
AM peak	-	-		-	-	
AM off peak	4.307	2.002	*	-4.307	-2.002	*
PM off peak	1.305	0.803		-1.305	-0.803	
PM peak	1.678	1.035		-1.678	-1.035	
No. of observations					820	
Rho-squared:					0.7345	
Adjusted rho-squared:					**0.7252**	
Log-likelihood at optimal					-748	
Log-likelihood at zero					-2,816	
LL at constant					NA	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.562	6.055	*
price.adult	-0.528	3.077	*
price.child (2-15)	-	-	-
price.senior (62+)	-1.162	0.190	
price.unacc child (8-11)	-	-	-
price.student advantage	-0.451	2.617	*
price.adultAAA member	-0.429	2.526	*
price.childAAA member	-	-	-
price.military adult	-	-	-
price.disabled adult	-	-	-
price.others	-0.525	3.068	*
price.period6	-0.460	2.679	*
wknd.period1	4.811	5.139	*
wknd.period2	-8.318	1.963	*
wknd.period3	1.614	1.530	
wknd.period4	1.685	1.689	
wknd.period5	1.689	1.547	
wknd.period6	1.538	1.452	
No. of observations		820	
Rho-squared:		0.7185	
Adjusted rho-squared:		**0.7121**	
Log-likelihood at optimal		-797	
Log-likelihood at zero		-2,816	
LL at constant		NA	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-8 Model 8 Result

	MNL			ML		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.266	10.741	*	-0.523	11.787	*
price.period1	0.013	1.050		-4.521	12.959	*
price.period2	-0.015	1.384		(2.356)	(20.432)	*
price.period3	-0.018	2.047	*			
price.period4	-0.017	1.926				
price.period5	-0.005	0.613				
price.period6	-0.002	0.231				
wknd.period1	0.959	1.108		4.005	4.728	*
wknd.period2	1.226	2.361	*	2.291	4.164	*
wknd.period3	0.790	1.654		0.307	0.711	
wknd.period4	0.899	2.520	*	-0.696	2.038	*
wknd.period5	-0.937	1.475		-1.662	3.502	*
wknd.period6	0.065	0.169		-1.247	3.126	*
No. of observations		959			959	
Rho-squared:		0.4190			0.4216	
Adjusted rho-squared:		**0.4151**			**0.4188**	
Log-likelihood at optimal		-1,913			-1,905	
Log-likelihood at zero		-3,293			-3,293	
LL at constant		NA			NA	

LC						
Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.280	-4.923	*	-0.633	-4.482	*
price.period1	0.044	1.243		-0.445	-3.343	*
price.period2	0.006	0.156		-0.421	-3.716	*
price.period3	-0.001	-0.014		-0.400	-3.758	*
price.period4	-0.003	-0.072		-0.399	-3.708	*
price.period5	0.006	0.144		-0.394	-3.601	*
price.period6	0.017	0.455		-0.399	-3.680	*

Class Model	**Class1**			**Class2**		
Class Size	0.723			0.277		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	-1.518	-1.106		1.518	1.106	
Monday	2.646	1.835		-2.646	-1.835	
Tuesday	2.443	1.728		-2.443	-1.728	
Wednesday	1.995	1.440		-1.995	-1.440	
Thursday	2.507	1.646		-2.507	-1.646	
Friday	2.121	1.523		-2.121	-1.523	
Saturday	0.456	0.204		-0.456	-0.204	
Early morning	-1.913	-4.068	*	1.913	4.068	*
AM peak	-1.058	-3.176	*	1.058	3.176	*
AM off peak	-0.330	-0.946		0.330	0.946	
PM off peak	2.270	1.454		-2.270	-1.454	
PM peak	1.420	1.096		-1.420	-1.096	
No. of observations					959	
Rho-squared:					0.4345	
Adjusted rho-squared:					**0.4266**	
Log-likelihood at optimal					-1,862	
Log-likelihood at zero					-3,293	
LL at constant					NA	

MNL with Socioeconomics			
Variable	Est	T-Stat	
advbk	-0.328	6.172	*
price.adult	-0.055	1.186	
price.child (2-15)	-	-	-
price.senior (62+)	-0.350	5.813	*
price.unacc child (8-11)	-	-	-
price.student advantage	-	-	-
price.adultAAA member	-0.034	0.747	
price.childAAA member	-	-	-
price.military adult	-	-	-
price.disabled adult	-	-	-
price.others	-0.302	1.070	
price.period6	-0.044	0.970	
wknd.period1	2.200	2.824	*
wknd.period2	1.025	1.964	*
wknd.period3	0.246	0.579	
wknd.period4	0.423	1.282	
wknd.period5	-0.592	0.986	
wknd.period6	-0.301	0.807	
No. of observations		959	
Rho-squared:		0.4026	
Adjusted rho-squared:		**0.3972**	
Log-likelihood at optimal		-1,967	
Log-likelihood at zero		-3,293	
LL at constant		NA	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-9 Model 9 Result

Variable	MNL Est	MNL T-Stat		ML Est	ML T-Stat	
advbk	-0.130	2.977	*	-0.447	23.425	*
price.period1	-0.031	2.869	*	-23.031	5.192	*
price.period2	-0.029	4.255	*	(20.015)	(5.006)	*
price.period3	-0.027	5.257	*			
price.period4	-0.022	6.002	*			
price.period5	-0.009	3.002	*			
price.period6	-0.004	1.361				
wknd.period1	1.860	3.386	*	1.186	5.814	*
wknd.period2	0.227	0.367		0.452	38.652	*
wknd.period3	0.144	0.343		-1.572	3.221	*
wknd.period4	0.605	2.305	*	8.139	3.249	*
wknd.period5	-0.583	1.550		-18.146	3.635	*
wknd.period6	0.747	3.178	*	12.915	4.613	*
No. of observations		1,307			1,307	
Rho-squared:		0.4843			0.4876	
Adjusted rho-squared:		**0.4814**			**0.4856**	
Log-likelihood at optimal		-2,315			-2,300	
Log-likelihood at zero		-4,488			-4,488	
LL at constant		NA			NA	

LC

Choice Model	Class1			Class2		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-0.381	-4.143	*	-0.167	-2.846	*
price.period1	0.109	2.234	*	-0.019	-0.487	
price.period2	0.036	0.369		-0.016	-0.445	
price.period3	0.082	1.660		-0.020	-0.539	
price.period4	0.076	1.588		-0.018	-0.530	
price.period5	0.064	1.342		-0.002	-0.060	
price.period6	0.092	1.877		-0.010	-0.279	

Class Model	Class1			Class2		
Class Size	0.549			0.451		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	1.348	3.218	*	-1.348	-3.218	*
Monday	0.303	0.694		-0.303	-0.694	
Tuesday	-0.705	-2.311	*	0.705	2.311	*
Wednesday	-0.657	-2.056	*	0.657	2.056	*
Thursday	-0.486	-1.574		0.486	1.574	
Friday	-0.296	-0.945		0.296	0.945	
Saturday	-0.135	-0.258		0.135	0.258	
Early morning	-	-		-	-	
AM peak	-2.757	-1.640		2.757	1.640	
AM off peak	-0.771	-2.522	*	0.771	2.522	*
PM off peak	-0.706	-2.392	*	0.706	2.392	*
PM peak	-0.787	-2.465	*	0.787	2.465	*
No. of observations					1,307	
Rho-squared:					0.4960	
Adjusted rho-squared:					**0.4903**	
Log-likelihood at optimal					-2,262	
Log-likelihood at zero					-4,488	
LL at constant					NA	

MNL with Socioeconomics

Variable	Est	T-Stat	
advbk	-0.556	18.993	*
price.adult	-0.151	13.208	*
price.child (2-15)	-	-	-
price.senior (62+)	-0.128	7.494	*
price.unacc child (8-11)	-	-	-
price.student advantage	-	-	-
price.adultAAA member	-0.122	8.557	*
price.childAAA member	-	-	-
price.military adult	-	-	-
price.disabled adult	-	-	-
price.others	-2.369	14.017	*
price.period6	-0.143	12.346	*
wknd.period1	3.642	5.768	*
wknd.period2	0.398	0.695	
wknd.period3	-0.368	0.823	
wknd.period4	-0.369	1.219	
wknd.period5	-0.725	1.769	
wknd.period6	0.422	1.758	
No. of observations		1,307	
Rho-squared:		0.4709	
Adjusted rho-squared:		**0.4669**	
Log-likelihood at optimal		-2,375	
Log-likelihood at zero		-4,488	
LL at constant		NA	

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Table A-10 Model 10 Result

MNL and ML

Variable	MNL Est	MNL T-Stat		ML Est	ML T-Stat	
advbk	-0.534	20.090	*	-0.602	58.485	*
price.period1	0.056	7.863	*	-6.945	25.872	*
price.period2	0.011	2.760	*	(3.371)	(20.846)	*
price.period3	-0.007	2.626	*			
price.period4	-0.019	11.389	*			
price.period5	-0.016	12.133	*			
price.period6	-0.008	8.293	*			
wknd.period1	2.701	9.230	*	3.390	6.153	*
wknd.period2	1.560	7.202	*	1.483	3.920	*
wknd.period3	-0.268	1.183		-0.255	0.875	
wknd.period4	-0.639	4.205	*	-1.129	4.460	*
wknd.period5	-0.758	4.708	*	-1.108	3.839	*
wknd.period6	0.404	4.267	*	0.618	2.516	*

LC

Choice Model	**Class1**			**Class2**		
Variable	Est	T-Stat		Est	T-Stat	
advbk	-1.164	-6.080	*	-0.327	-10.734	*
price.period1	-0.378	-1.451		-0.066	-2.898	*
price.period2	-0.518	-1.953		-0.083	-3.591	*
price.period3	-0.448	-2.712	*	-0.095	-3.954	*
price.period4	-0.606	-3.504	*	-0.089	-3.597	*
price.period5	-0.595	-3.399	*	-0.081	-3.202	*
price.period6	-0.426	-2.788	*	-0.090	-4.031	*

Class Model	**Class1**			**Class2**		
Class Size	0.595			0.406		
Variable	Est	T-Stat		Est	T-Stat	
Intercept	0.462	4.444	*	-0.462	-4.444	*
Monday	0.135	1.595		-0.135	-1.595	
Tuesday	0.031	0.379		-0.031	-0.379	
Wednesday	0.118	1.432		-0.118	-1.432	
Thursday	-0.062	-0.760		0.062	0.760	
Friday	-0.037	-0.440		0.037	0.440	
Saturday	0.079	0.585		-0.079	-0.585	
Early morning	-1.117	-5.476	*	1.117	5.476	*
AM peak	-0.922	-9.202	*	0.922	9.202	*
AM off peak	-0.297	-3.406	*	0.297	3.406	*
PM off peak	-0.083	-1.007		0.083	1.007	
PM peak	-0.287	-3.548	*	0.287	3.548	*

MNL with Socioeconomics

Variable	Est	T-Stat	
advbk	-0.674	32.119	*
price.adult	-0.215	22.716	*
price.child (2-15)	-	-	-
price.senior (62+)	-0.209	17.927	*
price.unacc child (8-11)	-	-	-
price.student advantage	-0.099	5.940	*
price.adultAAA member	-0.158	11.881	*
price.childAAA member	-	-	-
price.military adult	-1.084	18.308	*
price.disabled adult	-0.227	6.498	*
price.others	-0.238	19.380	*
price.period6	-0.204	21.406	*
wknd.period1	5.381	18.684	*
wknd.period2	1.900	9.105	*
wknd.period3	-0.674	3.033	*
wknd.period4	-1.646	11.372	*
wknd.period5	-1.577	8.975	*
wknd.period6	-0.384	3.444	*

	MNL	ML	LC	MNL with Socioeconomics
No. of observations	4,454	4,454	4,454	4,454
Rho-squared:	0.5637	0.5493	0.5845	0.5611
Adjusted rho-squared:	**0.5628**	**0.5487**	**0.5828**	**0.5599**
Log-likelihood at optimal	-6,674	-6,894	-6,356	-6,713
Log-likelihood at zero	-15,295	-15,295	-15,295	-15,295
LL at constant	-6,478	-6,478	-6,478	-6,478

*Statistically significant at the 5% significance level. Parenthesis indicates standard deviation.

Appendix B. Leg-based capacity consumption

Table B-1 Leg-based capacity consumption (cumulated)

Leg	Bkday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Leg(9-8)	Exist	1	1	2	3	3	3	4	4	5	5	5	5	8	8	10	11	14	14	14	16	16	23	27	34	37	43	59	70	92	123	202
Capacity*=	LC	0	1	1	2	3	4	5	6	10	11	15	15	16	16	17	18	19	21	23	25	28	30	34	38	46	54	64	77	91	131	202
202	MNL	0	0	1	1	1	2	2	3	4	5	7	7	8	8	9	9	10	12	13	15	17	20	24	28	34	40	49	60	71	110	202
Leg(8-7)	Exist	2	2	3	4	4	4	5	5	6	7	7	7	10	10	12	14	18	18	18	20	20	28	32	39	42	48	66	79	102	138	221
Capacity=	LC	1	1	2	3	3	4	6	8	11	13	17	17	18	19	20	22	23	25	27	31	33	36	40	45	53	62	73	86	102	143	221
221	MNL	0	0	1	1	1	2	3	3	5	6	8	8	8	9	10	11	12	13	15	17	19	22	26	31	37	45	54	66	77	120	220
Leg(7-6)	Exist	2	2	3	4	4	4	4	4	5	6	6	6	9	9	10	12	16	16	16	18	18	24	28	34	39	45	62	75	92	124	203
Capacity=	LC	0	1	2	2	3	4	5	7	10	11	14	15	15	16	17	17	18	19	21	24	26	28	30	34	41	47	55	64	76	116	189
203	MNL	0	0	0	1	1	1	2	2	3	4	5	5	6	6	7	7	8	9	10	11	13	15	17	21	25	31	38	47	56	100	190
Leg(6-5)	Exist	1	1	2	3	3	3	3	3	4	5	5	5	9	9	10	12	15	15	15	16	16	22	25	28	33	38	52	64	75	100	162
Capacity=	LC	0	1	1	2	3	3	4	6	9	10	11	11	12	12	13	13	14	15	16	19	20	22	24	27	32	38	44	52	62	95	160
162	MNL	0	0	0	1	1	1	1	2	2	3	4	4	4	5	5	5	6	7	8	9	10	12	14	16	20	25	31	38	46	85	159
Leg(5-4)	Exist	1	1	2	2	2	2	2	2	3	4	4	4	7	7	8	10	13	13	13	14	14	19	22	25	29	31	43	53	65	86	142
Capacity=	LC	0	1	1	2	2	3	4	5	8	8	9	9	10	10	11	11	12	13	14	15	17	18	20	23	27	32	37	44	53	82	142
142	MNL	0	0	0	0	1	1	1	2	2	3	3	3	4	4	4	5	5	6	6	7	8	10	12	14	17	21	27	34	41	76	141
Leg(4-3)	Exist	0	0	0	0	0	1	1	2	4	5	5	7	10	10	11	14	17	18	18	19	20	24	28	32	33	37	47	56	67	85	122
Capacity=	LC	0	0	0	1	1	1	1	1	2	3	4	4	5	6	7	7	8	9	11	12	14	16	18	20	24	28	34	41	49	67	106
122	MNL	0	0	0	0	1	1	1	1	2	2	3	3	3	3	4	4	5	5	6	7	8	10	12	14	16	21	27	34	41	63	106
Leg(3-2)	Exist	0	0	0	0	0	1	1	2	4	5	5	7	10	10	10	13	15	16	16	17	18	20	23	26	27	31	40	50	60	74	103
Capacity=	LC	0	0	0	1	1	1	1	1	2	3	4	4	5	6	7	7	8	9	11	12	13	15	18	20	23	28	33	40	48	65	103
103	MNL	0	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	5	5	6	7	8	10	11	13	16	21	27	34	40	62	103
Leg(2-1)	Exist	0	0	0	0	0	0	0	1	2	3	3	5	7	7	7	9	9	10	10	11	12	13	14	17	17	20	25	31	36	47	73
Capacity=	LC	0	0	0	0	0	1	1	1	1	2	3	3	4	4	5	6	7	8	9	10	11	12	13	15	17	20	24	28	34	46	73
73	MNL	0	0	0	0	0	1	1	1	1	1	2	2	2	2	3	3	3	4	4	5	6	7	8	9	11	14	18	23	28	43	73

*Indicates allowable capacity.

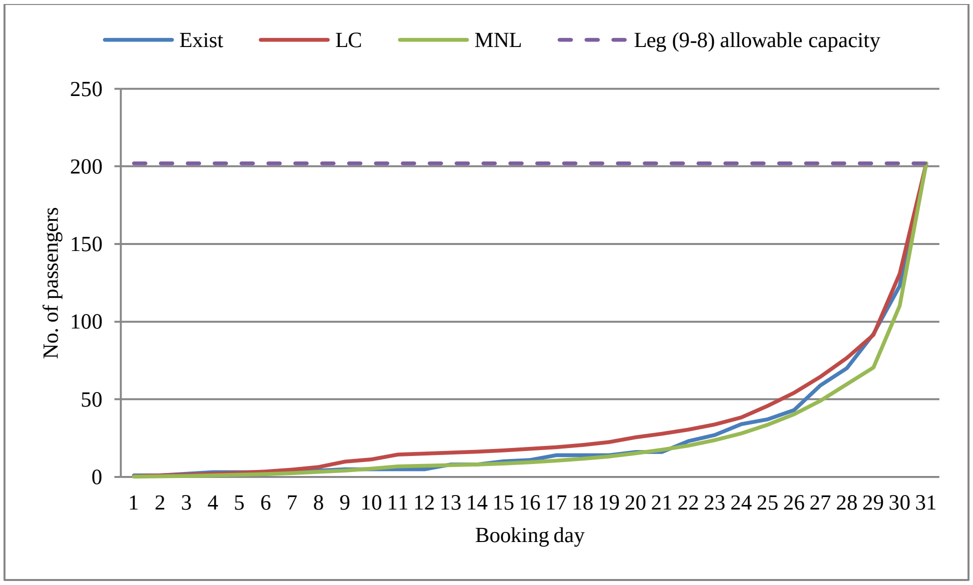


Figure B-1 Leg 9-8 capacity consumption (cumulated)

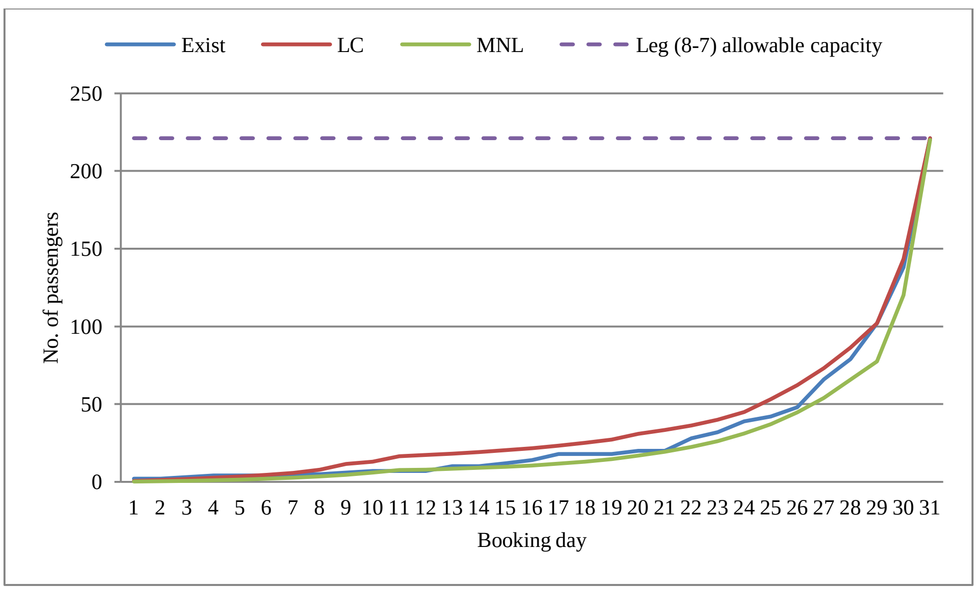


Figure B-2 Leg 8-7 capacity consumption (cumulated)

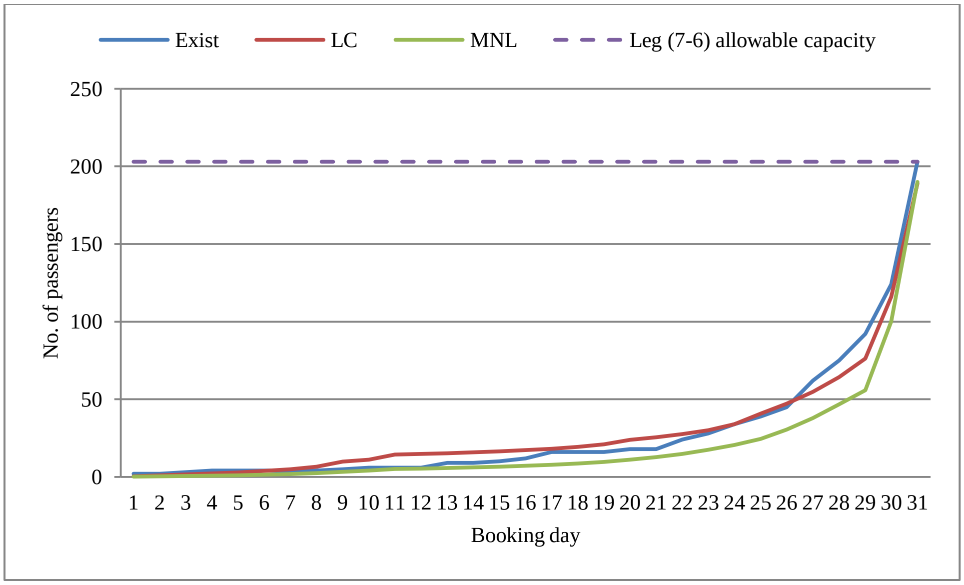


Figure B-3 Leg 7-6 capacity consumption (cumulated)

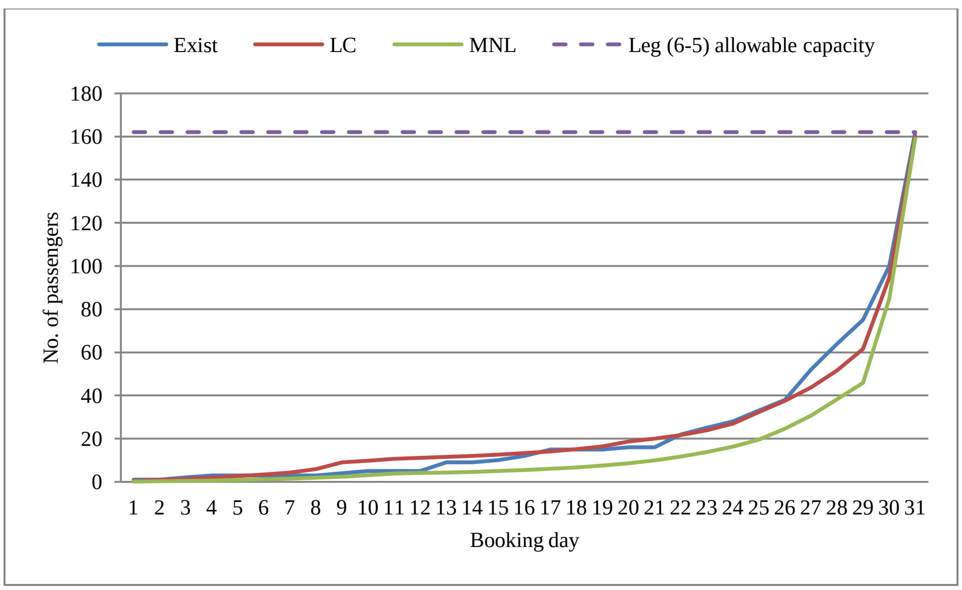


Figure B-4 Leg 6-5 capacity consumption (cumulated)

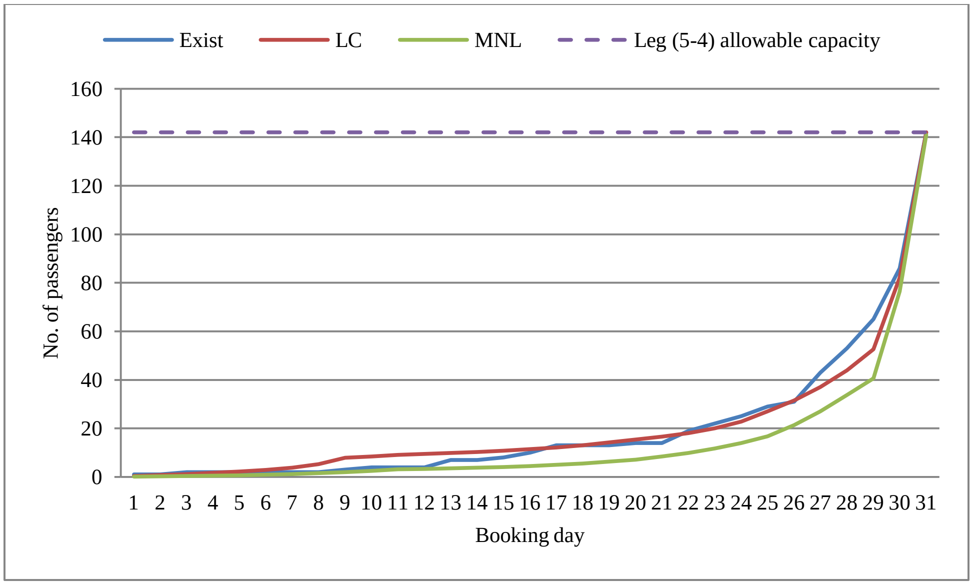


Figure B-5 Leg 5-4 capacity consumption (cumulated)

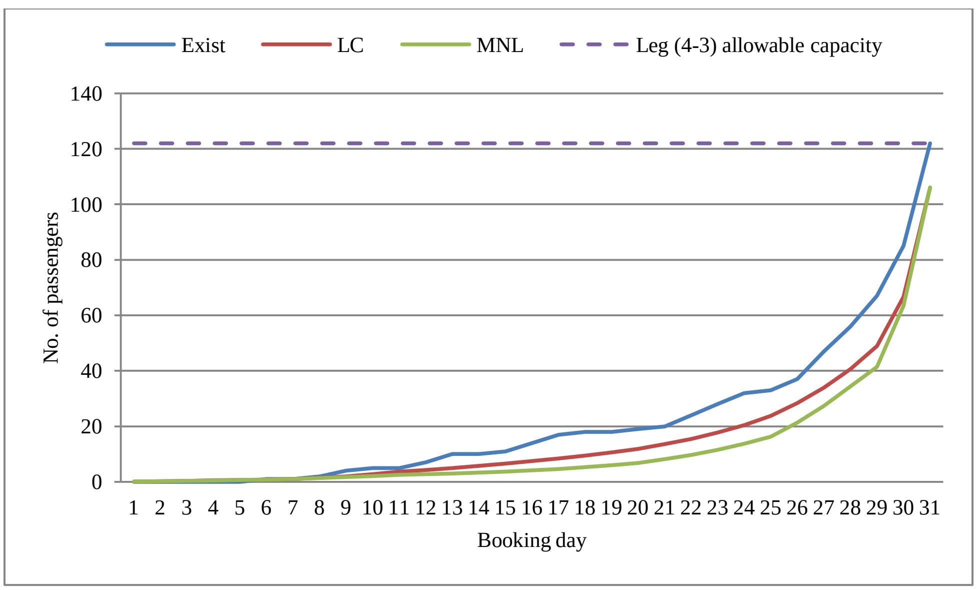


Figure B-6 Leg 4-3 capacity consumption (cumulated)

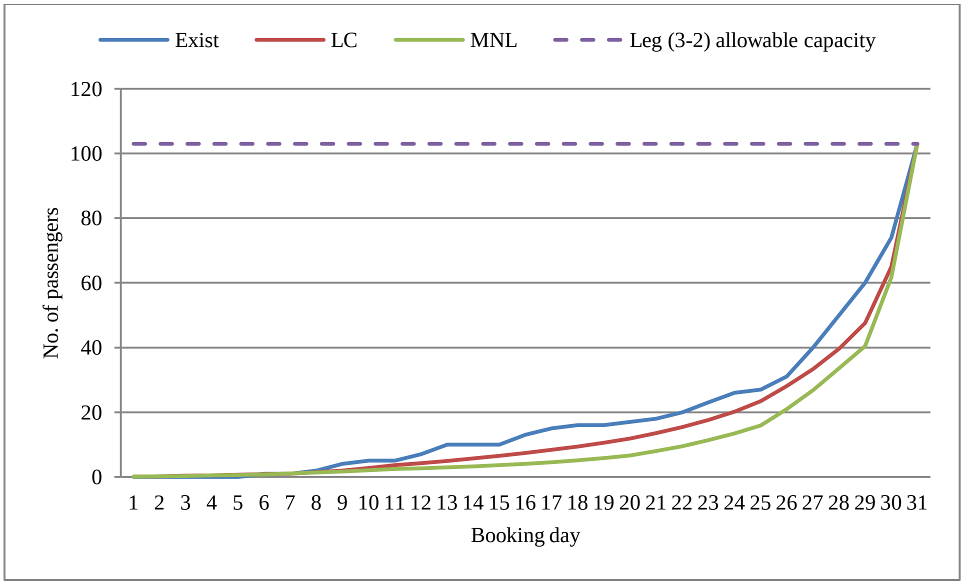


Figure B-7 Leg 3-2 capacity consumption (cumulated)

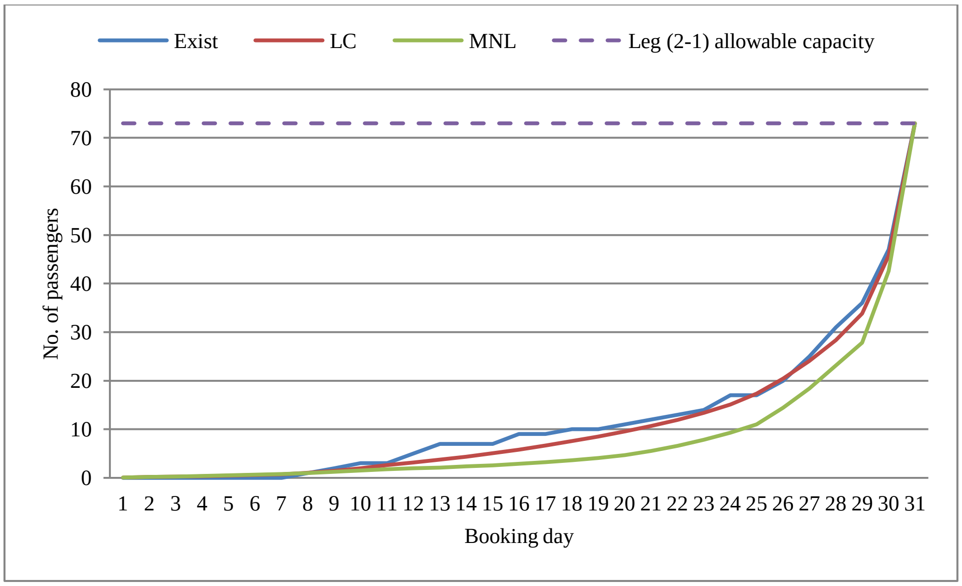


Figure B-8 Leg 2-1 capacity consumption (cumulated)

Appendix C. Choice models with mixed logit using non-parametric B-spline

Table C-1 Choice model with mixed logit non-parametric B-Spline (model1-5)

	Model1			**Model2**			**Model3**			**Model4**			**Model5**		
	Est	T-Stat		Est	T-Stat		Est	T-Stat		Est	T-Stat		Est	T-Stat	
advbk	-0.265	12.150	*	-0.688	126.189	*	-0.543	95.391	*	-0.430	41.250	*	-0.568	72.854	*
price1	-0.361	4.934	*	-0.214	31.776	*	-0.705	16.717	*	-1.187	6.673	*	-0.406	47.219	*
price2	-0.113	1.434		-0.214	21.605	*	-0.240	4.276	*	-0.161	7.862	*	-0.406	37.111	*
price3	-0.113	1.482		-0.214	39.311	*	-0.240	16.017	*	-0.161	16.483	*	-0.406	27.721	*
price4	-0.073	1.605		-0.214	6.655	*	-0.240	2.711	*	-0.161	5.645	*	-0.406	11.498	*
price5	0.202	1.757		-0.214	1.382		2.038	14.856	*	-0.127	0.926		4.569	7.813	*
price6	0.202	0.333		50.164	20.796	*	10.627	2.357	*	3.583	4.304	*	6.698	15.579	*
price7	16.625	1.339		50.164	20.801	*	10.627	2.358	*	74.378	6.066	*	6.698	15.578	*
wknd.period1	2.430	4.531	*	-0.444	7.700	*	1.767	4.518	*	3.918	10.715	*	0.176	1.361	
wknd.period2	1.668	4.504	*	-0.487	10.267	*	-0.711	3.649	*	1.704	5.162	*	0.153	1.206	
wknd.period3	0.743	3.374	*	-0.663	11.469	*	0.061	0.488		1.058	3.323	*	0.245	1.396	
wknd.period4	-0.449	1.008		-0.940	18.327	*	-0.132	1.130		-0.239	0.641		-0.495	4.276	*
wknd.period5	-1.557	4.410	*	-0.987	23.770	*	-1.427	5.745	*	-2.188	5.229	*	-0.603	5.292	*
wknd.period6	-3.132	4.400	*	-0.124	2.748	*	-0.138	1.631		-9.398	8.826	*	0.022	0.177	
No. of observations		1,361			20,000			19,032			4,472			11,536	
Rho-squared:		0.3205			0.5547			0.6070			0.4181			0.6736	
Adjusted rho-squared:		**0.3175**			**0.5545**			**0.6067**			**0.4172**			**0.6733**	
Log-likelihood at optimal		-3,176			-30,582			-25,687			-8,937			-12,929	
Log-likelihood at zero		-4,674			-68,680			-65,356			-15,357			-39,614	
Log-likelihood at constant		-3,988			-47,045			-41,704			-12,330			-22,530	

* Statistically significance at the 5% significance level.

Table C-2 Choice model with mixed logit non-parametric B-spline (model 6-10)

	Model6			Model7			Model8			Model9			Model10		
	Est	T-Stat		Est	T-Stat		Est	T-Stat		Est	T-Stat		Est	T-Stat	
advbk	-0.536	134.316	*	-1.672	26.488	*	-0.753	20.950	*	-0.488	16.048	*	-0.860	0.882	
price1	-0.329	64.833	*	-8.762	8.907	*	-0.474	17.206	*	-0.653	15.617	*	-0.751	0.389	
price2	-0.329	71.807	*	-2.547	4.320	*	-0.474	15.065	*	-0.653	14.902	*	-0.751	0.822	
price3	-0.329	41.721	*	-2.547	5.183	*	-0.474	5.190	*	-0.653	15.386	*	-0.751	0.859	
price4	-0.329	38.874	*	-0.075	1.281		-0.474	3.950	*	-0.653	10.611	*	-0.751	0.972	
price5	-0.329	11.110	*	-0.075	2.846	*	-0.474	1.199		21.804	6.994	*	12.620	1.865	
price6	14.061	19.699	*	-0.075	2.230	*	8.550	4.542	*	21.804	6.996	*	71.768	5.129	*
price7	14.061	19.717	*	13.556	3.169	*	9.061	6.272	*	21.804	6.994	*	71.768	5.130	*
wknd.period1	2.174	82.375	*	18.446	17.721	*	0.406	0.753		-0.939	1.816		4.497	0.637	
wknd.period2	0.828	52.888	*	-1.252	18.542	*	0.432	0.787		-0.951	1.837		0.597	0.067	
wknd.period3	0.487	37.030	*	-1.130	2.528	*	0.418	0.769		-1.174	2.327	*	0.145	0.016	
wknd.period4	0.399	30.604	*	-4.613	6.015	*	-0.315	1.054		-1.559	5.107	*	-0.059	0.006	
wknd.period5	-0.589	38.338	*	-7.825	11.953	*	-0.465	1.778		-1.823	5.886	*	-0.780	0.068	
wknd.period6	-2.184	58.208	*	-3.625	4.415	*	-0.568	2.227	*	1.779	0.744		1.714	1.187	
No. of observations		20,000			820			959			1,307			4,454	
Rho-squared:		0.4677			0.7862			0.5039			0.7200			0.7194	
Adjusted rho-squared:		**0.4675**			**0.7813**			**0.4997**			**0.7168**			**0.7185**	
Log-likelihood at optimal		-36,556			-605			-1,634			-1,257			-4,291	
Log-likelihood at zero		-68,680			-2,830			-3,293			-4,488			-15,295	
Log-likelihood at constant		-49,068			NA			NA			NA			-6,478	

* Statistically significance at the 5% significance level.

Appendix D. Major C code for dynamic discrete choice model estimation

D-1: library.h

```
#ifndef LIBRARY
#define LIBRARY

#ifdef __cplusplus
extern "C" {
#endif
#include <oratio/oratio.h>

typedef struct {
    // individual id
    int id;
    // individual specific variable
    double *indiv;
    } ind;

typedef struct {
    // number of time period
    int tperiod;
    } duration;

typedef struct {
    // original departure time
    int tdepart;
    } schedule;

typedef struct {
    // fare of original ticket at each time
    double **current;
    } cur;

typedef struct {
    // decision variable, 0 or 1 at each time period to each product type
    double **decision;
    // static attributes for potential choice j in time t
    double **stati;
    // dynamic variable for each potential product in each time
    double **dym;
    } poten;

// Define the type of data used for the samplings.
typedef struct {
    // number of individuals
    int indivNum;
    // number of time period
    int time;
    // number of asc
    int numC;
    // number of individual variables
    int numINDIVAR;
    // number of current ticket variables
    int numSTATIC;
    // number of dynamic vars for potential choice
    int numDYNAMIC;
    // number of refund var
    int numREF;
    // number of keep
    int numCUR;
    // number of 'day from issue, exc'
    int numDF;
    // number of 'day from issue, cnl'
    int numDF1;
```

```
    // number of choice
    int numCNL1;
    // number of day 1 cancel
    int numEXC16;
    // number of day 16 exchange
    int numEX;
    // number of early exc
    int numLX;
    // number of late exc
    int numch;
    // discount factor beta
    double discount;
    // flag for number of errors
    int errs;
    // index variable to count the number of iterations (direction changes)
    int iterations;
    } glo;

//put all data files together
    typedef struct {
         ind* in;
         duration* period;
         schedule* depart;
         cur* curr;
         poten* pot;
         glo* glonum;

         double*** prob_matrix; // Probability matrix
         Random *rand_seed;      // Random seed
         double** draw;          //drawn from Random normal distribution
         double ***err1; //error term for i,j,t
         double *err2;  //error term for i
         double **p;    //the random value (0 <= p <= 1) for calculating normal dynamic
variable y.
         double vp;   //the random value (0 <= p <= 1) for calculating MC v
         double ***perror;//the random value (0 <= p <= 1) for calculating gumbel
error_ijt for different i,j,t
         double *perrori;//the random value (0 <= p <= 1) for calculating gumbel
error_ijt for different i

} Aldata;
//check number of parameters (dimensions of the problem)
    int get_dimension(Aldata* d);

//read data from four data structures and allocate memory
    Aldata* format_data();
//generating attributes for scenario tree
    double** draw_random_y(Aldata* d, double** draw);
//mode in scenario tree
    double*** calculate_mode(Aldata* d, double* x);
//mode that is correlated to current situation in each time
    double** calculate_mode_real(Aldata* d, double* x);
//v is for calculating E
    double*** calculate_v(Aldata* d, double* x, double vp);
//recursive process for scenario tree
    double cal_E(Aldata* d, int t, int T, double *v, double current, int n, double* x,
int indiv);

// calculate probability
//probability of specific ticket changing decision
    double*** cal_probcar (Aldata* d, double* x);
//probability of not changing ticket, PI0
    double** cal_prob (Aldata* d, double* x);

// functions of reading four .txt data files
    void read_new_indiv(glo* paraNB, ind* paraIN);
    void read_new_time(glo* paraNB, duration* paraIN);
    void read_new_depart(glo* paraNB, schedule* paraIN);
```

```
    void read_new_current(glo* paraNB, cur* paraC, ind* paraIN);
    void read_new_poten(glo* paraNB, ind* paraIN, poten* paraP);
    void read_new_choice(glo* paraNB, ind* paraIN, poten* paraP);

// function of reading coefficient data files
    void read_new_para(double *x, glo* paraNB);

 // functions of allocating memories to four structures and their elements
     glo* getGlo();
     ind* getIn(glo *glonum);
     duration* getT(glo *glonum);
     schedule* getDT(glo *glonum);
     cur* getC(glo *glonum);
     poten* getP(glo *glonum);

    double*** c_malloc_P(glo* paraNB);

//function of allocating memory to error terms
    double*** c_malloc_e(glo* paraNB);

// function of allocating memory to the array of utility U[i][j][t]
    double*** c_malloc_u(glo* paraNB);

    double**** c_malloc_uy(glo* paraNB);

// function of allocating memory to the array of summation of exp(U[i][j]) for each
person i
    double* c_malloc_w(glo* paraNB);

//function of allocating memory to v_itj (the dimension of j is for 1000 draws)
    double*** c_malloc_v(glo* paraNB);

// free four structures
    void free_ind (ind* in, glo* paraNB);
    void free_duration (duration* period, glo* paraNB);//
    void free_schedule(schedule* depart, glo* paraNB);//
    void free_cur (cur* c, glo* paraNB);
    void free_poten (poten* p, glo* paraNB);
    void free_uy(double**** u, glo* paraNB);
    void free_glo (glo* g);
    void free_w(double* W);
    void free_v(double*** v, glo* paraNB);
    void free_u(double*** u, glo* paraNB);
    void free_p(double*** u, glo* paraNB);
    void free_err (Aldata *d);

// Function to help in calculating t-stats
    double amlet_t_statistics(int n, double *theta, double *hypothetical, double **I,
double alpha, double *t);

void op_matrix_inverse(const enum CBLAS_ORDER Order, const enum CBLAS_UPLO Uplo, double
*I, int npar);

#ifdef __cplusplus
}
#endif

#endif
```

D-2: library.c

```
#include "library.h"
#include <stdlib.h>
#include <string.h>
#include <stdio.h>
#include <oratio/oratio.h>

/*read data from data structures and allocate memory*/
Aldata* format_data(){
  Aldata *d;
  int i,j,t,n;
  int st=2;

  d= malloc(sizeof(Aldata));
  d->glonum = getGlo();
  d->in = getIn(d -> glonum) ;
  d->period = getT(d -> glonum);
  d->depart = getDT(d -> glonum);
  d->pot = getP(d -> glonum);
  d->curr = getC(d -> glonum);

  read_new_indiv(d->glonum, d->in);
  read_new_time(d->glonum, d->period);
  read_new_depart(d->glonum, d->depart);
  read_new_current(d->glonum, d->curr, d->in);
  read_new_poten(d->glonum, d->in, d->pot);
  read_new_choice(d->glonum, d->in, d->pot);

  d->prob_matrix = c_malloc_P(d -> glonum);
  d->rand_seed = ran_random();

 d->p=nt_matrix_new(d->glonum->time+1, st);                 // st
  for (t=0; t<d->glonum->time; t++){
    for (n=0; n<st; n++){                                   // st
    d->p[t][n] = ran_random_get_val(d->rand_seed );
    }
  }

  d->vp = ran_random_get_val(d->rand_seed );

  d->perror=c_malloc_e(d -> glonum);
  for(i = 0; i < d->glonum ->indivNum; i++) {
    for(t = 0; t < d->glonum->time+2; t++) {
      for(j = 0; j < d->glonum ->numch+1; j++) {
       d->perror[i][j][t] = ran_random_get_val(d->rand_seed );
       }
      }
   }

  d->perrori=malloc(d->glonum ->indivNum*sizeof(double));
  for(i = 0; i < d->glonum ->indivNum; i++) {
  d->perrori[i] = ran_random_get_val(d->rand_seed );
  }

  d->draw=nt_matrix_new(d->glonum->time, st);               // st
  for (t=0; t<d->glonum->time; t++){
    for (n=0; n<st; n++){                                   // st
    d->draw[t][n] = st_normal_icdf(d->p[t][n], 0, 16);
    }
  }

 d->err1=c_malloc_e(d->glonum);
 d->err2=malloc(d->glonum ->indivNum*sizeof(double));
    for(i = 0; i < d->glonum ->indivNum; i++) {
        for(t = 0; t < d->glonum->time+2; t++) {
        d->err2[i]=st_gumbel_icdf(d->perrori[i], 0, 1 );
```

```
            for(j = 0; j < d->glonum ->numch+1; j++) {
            d->err1[i][j][t]=st_gumbel_icdf(d->perror[i][j][t], 0, 1 );
        }
      }
    }
    free_p(d->perror, d->glonum);
    free(d->perrori);
    return d;
    }

/*check number of parameters (dimension of the problem)*/
int get_dimension(Aldata* d){

    int l0= d->glonum->numC;         //asc
    int l1= d->glonum->numINDIVAR;   //indiv var
    int l2= d->glonum->numSTATIC;    //static var
    int l3= d->glonum->numREF;       //refund var
    int l4= d->glonum->numCUR;       //keep var
    int l5= d->glonum->numDF;        //exc dfi
    int l6= d->glonum->numDF1;       //cnl dfi
    int l7= d->glonum->numCNL1;      //day 1 cnl
    int l8= d->glonum->numEXC16;     //day 16 exc
    int l9= d->glonum->numEX;        //early exc
    int l10= d->glonum->numLX;       //late exc

   return l0+l1+l2+l3+l4+l5+l6+l7+l8+l9+l10;
}

/* get a glo variable*/
glo* getGlo() {
    glo *glonum;
    glonum = (glo*)malloc(sizeof(glo));
    glonum->indivNum = 696;      // 696 individuals
    glonum->time=16;             // bkday16-bkday31
    glonum->numch=16;            // deptt5-deptt19 (15) +cnl (1) = 16
    glonum->numC=1;              // asc
    glonum->numINDIVAR=7;        // gr size, mor, eve, mon, fri, sta1, sta3
    glonum->numSTATIC=1;         // fare
    glonum->numREF=1;            // refund
    glonum->numCUR=1;            // keep
    glonum->numDF=1;             // exc dfi
    glonum->numDF1=1;            // cnl dfi
    glonum->numCNL1=1;           // day1 cnl
    glonum->numEXC16=1;          // day16 exc
    glonum->numEX=1;             // early exc
    glonum->numLX=0;             // late exc
    return glonum;
}

//get a ind variable
ind* getIn(glo *glonum) {
    ind* in = malloc((glonum->indivNum) * sizeof(ind));
    return in;
}
//get period
duration* getT(glo *glonum) {
    duration* period = malloc((glonum->indivNum) * sizeof(duration));
    return period;
}
//get depart
schedule* getDT(glo *glonum) {
    schedule* depart = malloc((glonum->indivNum) * sizeof(schedule));
    return depart;
}
//get a cur variable
cur* getC(glo* glonum) {
    cur *curr;
    int i;
```

```
    int ttnumCUR=(glonum->time) * glonum->numCUR;
    curr = (cur *)malloc(sizeof(cur));

    curr->current = malloc( glonum->indivNum * sizeof(double *));
    for(i=0;i<glonum->indivNum;i++) {
        curr->current[i]=malloc(ttnumCUR * sizeof(double));
    }
   return curr;
}

//get a poten variable
poten* getP(glo* glonum) {
    poten *pot;
    int i;
    int ttnumSTA = (glonum->time) * glonum->numch * glonum->numSTATIC;
    int ttnumDEC = glonum->time *( glonum->numch+1);
    pot = (poten *)malloc(sizeof(poten));

    pot->stati = malloc(glonum->indivNum * sizeof(double *));
    for(i = 0;i < glonum->indivNum;i++) {
        pot->stati[i]=malloc(ttnumSTA * sizeof(double));
    }
    pot->decision = malloc(glonum->indivNum * sizeof(double *));
    for(i = 0;i < glonum->indivNum;i++) {
        pot->decision[i]=malloc(ttnumDEC * sizeof(double));
    }
    return pot;
}

/*read individual*/
void read_new_indiv(glo* paraNB, ind* paraIN){
    FILE *inn;
    FILE *out;
    int i=0, j=0;

    inn=fopen("indivX.txt","r");
    out=fopen("oput1.txt","w");

    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
        }

    for(i = 0; i < paraNB ->indivNum; i++) {
// read individual ID
        fscanf(inn, "%d", &paraIN[i].id);
        paraIN[i].indiv = malloc( (paraNB ->numINDIVAR) * sizeof(double));  // allocate
memory for individual variable
        for (j=0;j<(paraNB->numINDIVAR);j++) {
// read the x variables for this individual
            fscanf(inn, "%lg", &paraIN[i].indiv[j]);
            }
    }

    for(i=0;i<paraNB->indivNum;i++) {
// write individual ID
        fprintf(out,"%d", paraIN[i].id);
        for (j=0;j<(paraNB->numINDIVAR);j++) {
// write individual variable
            fprintf(out, "%lg", paraIN[i].indiv[j]);
            }
        fprintf(out,"\n");
    }
    fclose(inn);
    fclose(out);
    return;
}
```

```
/*read time*/
void read_new_time(glo* paraNB, duration* paraIN){
    FILE *inn;
    FILE *out;
    int i=0;

    inn=fopen("time.txt","r");
    out=fopen("oput_t.txt","w");

    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
        }

    for(i = 0; i < paraNB ->indivNum; i++) {
// read individual ID
        fscanf(inn, "%d", &paraIN[i].tperiod);
    }

    for(i=0;i<paraNB->indivNum;i++) {
// write individual ID
        fprintf(out,"%d", paraIN[i].tperiod);
        fprintf(out,"\n");
    }
    fclose(inn);
    fclose(out);
    return;
}

/*read departure time*/
void read_new_depart(glo* paraNB, schedule* paraIN){
    FILE *inn;
    FILE *out;
    int i=0;

    inn=fopen("dtime.txt","r");
    out=fopen("oput_dt.txt","w");

    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
        }

    for(i = 0; i < paraNB ->indivNum; i++) {
// read individual ID
        fscanf(inn, "%d", &paraIN[i].tdepart);
    }

    for(i=0;i<paraNB->indivNum;i++) {
// write individual ID
        fprintf(out,"%d", paraIN[i].tdepart);
        fprintf(out,"\n");
    }
    fclose(inn);
    fclose(out);
    return;
}

/*read currrent*/
void read_new_current(glo* paraNB, cur* paraC, ind* paraIN) {
    FILE *inn;
    FILE *out;

    int i=0, j=0;
    int ttnumCURT = paraNB->time * paraNB->numCUR;
    inn=fopen("current.txt","r");
    out=fopen("oput2.txt","w");
```

```
    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
    }
    for(i=0;i<paraNB->indivNum;i++) {
//read individual ID
        fscanf(inn, "%d", &paraIN[i].id);
        for (j=0;j<ttnumCURT;j++) {
//read each individual's total current var
            fscanf(inn, "%lg", &paraC->current[i][j]);
      }
    }
    for(i=0;i<paraNB->indivNum;i++) {
//write individual ID
        fprintf(out,"%d", paraIN[i].id);
        for (j=0;j<ttnumCURT;j++) {
//write each individual's total current var
            fprintf(out,"%lg",paraC->current[i][j]);
            }
        fprintf(out,"\n");
   }
    fclose(inn);
    fclose(out);
}
/*read potent */
void read_new_poten(glo* paraNB, ind* paraIN, poten* paraP) {
    FILE *inn;
    FILE *out;

    int i=0, j=0;
    int ttnumSTAT = (paraNB->time) * (paraNB->numch-1) * paraNB->numSTATIC; //fixed from
'paraNB->numch' because 'cancel' (j=15) no-static
    inn=fopen("poten.txt","r");
    out=fopen("oput3.txt","w");

    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
    }

    for(i = 0;i < paraNB->indivNum;i++) {
//read individual ID
        fscanf(inn, "%d", &paraIN[i].id);
        for (j=0;j<ttnumSTAT;j++) {
//read each individual's total potential static variable
            fscanf(inn, "%lg", &paraP->stati[i][j]);
        }
    }

    for(i=0; i<paraNB->indivNum; i++) {
//write individual ID
        fprintf(out,"%d", paraIN[i].id);
        for (j=0; j<(ttnumSTAT); j++) {
//write each individual's total potential static variable
            fprintf(out,"%lg",paraP->stati[i][j]);
        }
        fprintf(out,"\n");
    }
    fclose(inn);
    fclose(out);
}

/*read choice.txt*/
void read_new_choice(glo* paraNB, ind* paraIN, poten* paraP) {
    FILE *inn;
    FILE *out;
    int i=0, j=0;
    int ttnumDEC = paraNB->time * (paraNB->numch+1);
```

```
    inn=fopen("choice.txt","r");
    out=fopen("oput4.txt","w");
    if (inn == NULL) {
        printf ("File could not be opened\n");
        exit(-1);
    }
    for(i=0; i< paraNB->indivNum; i++) {
        fscanf(inn, "%d", &paraIN[i].id);
        for (j = 0;j<(ttnumDEC);j++) {
            fscanf(inn, "%lg", &paraP->decision[i][j]);
        }
    }
    for(i=0;i<paraNB->indivNum;i++) {
        fprintf(out,"%d", paraIN[i].id);
        for (j=0;j<(ttnumDEC);j++) {
            fprintf(out,"%lg",paraP->decision[i][j]);
        }
        fprintf(out,"\n");
    }
    fclose(inn);
    fclose(out);
}

// read the para.txt file
void read_new_para(double *x, glo* paraNB){
    FILE *inn;
    int i;
    int j = 0;
    float f;
    inn=fopen("parady.txt","r");
        if (inn == NULL) {
            printf ("File could not be opened\n");
            exit(-1);
            }
    // Reads the ASCs
        for(i=0; i< paraNB->numC; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads the individual specific parameters
        for(i=0; i< paraNB->numINDIVAR; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
        fscanf(inn, "\n");
    // Reads the static parameters
        for(i=0; i< paraNB->numSTATIC; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads the refund parameters
        for(i=0; i< paraNB->numREF; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads the keep parameters
        for(i=0; i< paraNB->numCUR; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads 'day from issue, exc' parameters
        for(i=0; i< paraNB->numDF; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads 'day from issue, cnl' parameters
        for(i=0; i< paraNB->numDF1; i++) {
          fscanf(inn, "%f", &f);
```

```
          x[j++] = f;
        }
    // Reads 'cancel day1' parameters
        for(i=0; i< paraNB->numCNL1; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads 'exchange day16' parameters
        for(i=0; i< paraNB->numEXC16; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads 'early exc' parameters
        for(i=0; i< paraNB->numEX; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
    // Reads 'late exc' parameters
        for(i=0; i< paraNB->numLX; i++) {
          fscanf(inn, "%f", &f);
          x[j++] = f;
        }
}

double*** c_malloc_u(glo* paraNB){
       double ***U;
       int i,j;
       U = (double***)malloc(paraNB ->indivNum*sizeof(double**));
       for(i = 0; i < paraNB->indivNum; i++) {
       U[i] = (double**)malloc(paraNB->numch *sizeof(double*));
          for(j=0;j<paraNB->numch;j++){
         U[i][j]=(double*)malloc(paraNB->time*sizeof(double));
          }
       }
      return U;
}
double**** c_malloc_uy(glo* paraNB){
       double ****U;
       int i,j,t;
       int st=2;                                                  // st
       U = (double****)malloc(paraNB ->indivNum*sizeof(double***));
       for(i = 0; i < paraNB->indivNum; i++) {
       U[i] = (double***)malloc(paraNB->numch *sizeof(double**));
          for(j=0;j<paraNB->numch;j++){
         U[i][j]=(double**)malloc((paraNB->time+1)*sizeof(double*));
        for(t=0; t<paraNB->time+1; t++){
        U[i][j][t]=(double*)malloc(st*sizeof(double));                  // st
        }
       }
       }

      return U;
}
double*** c_malloc_v(glo* paraNB){
       double ***v;
       int i,j;
       int st=2;                                                         // st
       v = (double***)malloc(paraNB ->indivNum*sizeof(double**));
       for(i = 0; i < paraNB->indivNum; i++) {
       v[i] = (double**)malloc((paraNB->time+1) *sizeof(double*));
          for(j=0;j<paraNB->time+1;j++){
         v[i][j]=(double*)malloc(st*sizeof(double));                    // st
          }
       }
      return v;
}

double*** c_malloc_e(glo* paraNB){
```

```
        double ***e;
        int i,j;
        int numch=16;
        e = (double***)malloc(paraNB ->indivNum*sizeof(double**));
        for(i = 0; i < paraNB->indivNum; i++) {
        e[i] = (double**)malloc((numch+1) *sizeof(double*));                    // (numch+1)
           for(j=0;j<(numch+1);j++){                                              // (numch+1)
          e[i][j]=(double*)malloc((paraNB->time+2)*sizeof(double));
           }
        }
       return e;
}
double*** c_malloc_P(glo* paraNB){
        double ***P;
        int i,j;
        int numch=16;
        P = (double***)malloc(paraNB ->indivNum*sizeof(double**));
             for(i = 0; i < paraNB->indivNum; i++) {
                 P[i] = (double**)malloc((numch+1)*sizeof(double*));         // (numch+1)
           for(j=0;j<(numch+1);j++){                                              // (numch+1)
          P[i][j]=(double*)malloc(paraNB->time*sizeof(double));
           }
        }
       return P;
}
void free_ind(ind* in, glo* paraNB){
        int i;
        for(i=0;i<paraNB->indivNum;i++) {
        free(in[i].indiv);
        }
        free(in);
}

void free_duration(duration* period, glo* paraNB){
        free(period);
}
void free_schedule(schedule* depart, glo* paraNB){
        free(depart);
}
void free_cur(cur* curr, glo* paraNB){
        int i;
        for(i=0;i<paraNB->indivNum;i++){
        free(curr->current[i]);
        }
        free(curr->current);
        free(curr);
}
void free_poten(poten* p, glo* paraNB){
        int i;
        for(i=0;i<paraNB->indivNum;i++){
         free(p->stati[i]);
         free(p->decision[i]);
         }
       free(p->stati);
       free(p->decision);
       free(p);
}
void free_err (Aldata *d) {
    free_p(d->err1, d->glonum);
    free(d->err2);
}
void free_glo (glo* g){
     free(g);
}
void free_v(double*** v, glo* paraNB){
     int i, t;
    for(i = 0; i < paraNB ->indivNum; i++) {
```

```
       for(t=0; t<paraNB->time+1; t++) {
        free(v[i][t]);
         }
        free(v[i]);
      }
    free(v);
}
void free_u(double*** u, glo* paraNB){
     int i, j;
    for(i = 0; i < paraNB ->indivNum; i++) {

       for(j=0; j<paraNB->numch; j++) {
        free(u[i][j]);
         }
        free(u[i]);
      }
    free(u);
}
void free_uy(double**** u, glo* paraNB){
    int i, j, t;
    for(i = 0; i < paraNB ->indivNum; i++) {
        for(j=0; j<paraNB->numch; j++) {
            for (t=0; t<paraNB->time+1; t++){
            free(u[i][j][t]);
            }
        free(u[i][j]);
        }
    free(u[i]);
    }
    free(u);
}
void free_p(double*** u, glo* paraNB){
     int i, j;
    for(i = 0; i < paraNB ->indivNum; i++) {

       for(j=0; j<paraNB->numch+1; j++) {
        free(u[i][j]);
         }
        free(u[i]);
      }
    free(u);
}
```

D-3: main.c

```
#include <math.h>
#include <float.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

#include <oratio/oratio.h>
#include "library.h"

/* (I)
Calculate mode r[i][t]*/

double** calculate_mode_real(Aldata* d, double* x){

    int i=d->glonum->indivNum;
    int j=d->glonum->numch;
    int t;
    int  l,k,m;
    double sum=0;

// put the length of coefficients into  x
    int l0= d->glonum->numC;        //asc
    int l1= d->glonum->numINDIVAR;  //indiv var
    int l2= d->glonum->numSTATIC;   //static var
    int l3= d->glonum->numREF;      //refund var
    int l4= d->glonum->numCUR;      //keep var
    int l5= d->glonum->numDF;       //day from var
    int l6= d->glonum->numDF1;      //day from var
    int l7= d->glonum->numCNL1;     //day 1 cancel
    int l8= d->glonum->numEXC16;    //day 16 exchange
    int l9= d->glonum->numEX;       //early exc
    int l10= d->glonum->numLX;      //late exc

    double  **r_real;
    r_real=  nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
    double  ***U=c_malloc_u(d->glonum);

// calculate utility for 16 choices, from time 0
   for(i = 0; i < d->glonum ->indivNum; i++) {
      for(t=0; t<d->period[i].tperiod; t++) { //*

// exchange decision
// j=0-14 (deptt5-19)
          for(j=0; j<d->glonum->numch-1; j++) {
              sum = 0.0;
               for (l=0; l<l2;l++ ) {
                  sum+=((d->pot->stati[i][l2*j+l+t*(d->glonum->numch-1)*l2])-(d->curr-
>current[i][0]))*x[l0+l1+l];
                    }

                    for (l=0; l<l5;l++ ) {
                  sum+=t*x[l0+l1+l2+l3+l4+l];
                    }
                    if (t==15) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7];
                    }
                    if (d->depart[i].tdepart >j+5) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7+l8];
                    }
              U[i][j][t] = sum;
          }

//  cancel decision
// j=15 (cancel)
              j=15;
```

```
                sum = 0.0;
                sum+=x[0];

                for (k=0; k<l1;k++ ) {
                    sum+=d->in[i].indiv[k]*x[l0+k];
                }
                 for (m=0; m<l3;m++ ) {
                    sum+=(d->curr->current[i][0])*x[l0+l1+l2+m];
                 }
                      for (l=0; l<l6;l++ ) {
                    sum+=t*x[l0+l1+l2+l3+l4+l5+l];
                      }
                      if (t==0) {
                          sum+=x[l0+l1+l2+l3+l4+l5+l6];
                      }
                U[i][j][t] = sum;
            }
        }

// calculate mode
    for(i = 0; i < d->glonum ->indivNum; i++) {
        for(t=0; t<d->period[i].tperiod; t++) { //*
            sum=0;
            for (j = 0;j<d->glonum->numch;j++) {
            sum+=exp(U[i][j][t]);
        }
        r_real[i][t]= log(sum);
        }
    }
       free_u(U,d->glonum);
       return r_real;
}

/* (II) Let n=0,1 only
calculate mode for the scenario tree, r[i][t][n], n refers to the position in the tree;
*/
double*** calculate_mode(Aldata* d, double* x){

    int i=d->glonum->indivNum;
    int j=d->glonum->numch;
    int t;
    int  l,k,m,n;
    double sum=0;

// put the length of coefficients into  x
    int l0= d->glonum->numC;        //asc
    int l1= d->glonum->numINDIVAR;  //indiv var
    int l2= d->glonum->numSTATIC;   //static var
    int l3= d->glonum->numREF;      //refund var
    int l4= d->glonum->numCUR;      //keep var
    int l5= d->glonum->numDF;       //day from var
    int l6= d->glonum->numDF1;      //day from var
    int l7= d->glonum->numCNL1;     //day 1 cancel
    int l8= d->glonum->numEXC16;    //day 16 exchange
    int l9= d->glonum->numEX;       //early exc
    int l10= d->glonum->numLX;      //late exc

    double  ***r;
    r=  c_malloc_v(d->glonum);
    double ****UY = c_malloc_uy(d->glonum);

// calculate utility for 16 choices, from time 0
    for(i = 0; i < d->glonum ->indivNum; i++) {
        for(t=1; t<d->period[i].tperiod+1; t++) { //*

// calculate utilities for the first level of scenario tree, n=0
            for(n=0; n<1; n++){
```

```
// exchange decision
// j=0-14 (deptt5-19)
          for(j=0; j<d->glonum->numch-1; j++) {
              sum = 0.0;
               for (l=0; l<l2;l++ ) {
                  sum+=((d->pot->stati[i][l2*j+l+t*(d->glonum->numch-1)*l2])-(d->curr-
>current[i][0]))*x[l0+l1+l];
              }

               for (l=0; l<l5;l++ ) {
                  sum+=t*x[l0+l1+l2+l3+l4+l];
                    }
                    if (t==15) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7];
                    }
                    if (d->depart[i].tdepart >j+5) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7+l8];
                    }
              UY[i][j][t][n]  = sum;
          }
//  cancel decision
// j=15 (cancel)
              j=15;
              sum = 0.0;
              sum+=x[0];
               for (k=0; k<l1;k++ ) {
                  sum+=d->in[i].indiv[k]*x[l0+k];
              }
               for (m=0; m<l3;m++ ) {
                  sum+=(d->curr->current[i][0])*x[l0+l1+l2+m];
               }
                    for (l=0; l<l6;l++ ) {
                  sum+=t*x[l0+l1+l2+l3+l4+l5+l];
                    }
                    if (t==0) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6];
                    }
              UY[i][j][t][n] = sum;
          }
// calculate utilities for the second level of scenario tree, n=1
            for(n=1; n<2; n++){

// exchange decision
// j=0-14 (deptt5-19)
          for(j=0; j<d->glonum->numch-1; j++) {
              sum = 0.0;
               for (l=0; l<l2;l++ ) {
                  sum+=((d->pot->stati[i][l2*j+l+(t+1)*(d->glonum->numch-1)*l2])-(d-
>curr->current[i][0]))*x[l0+l1+l];
              }
               for (l=0; l<l5;l++ ) {
                  sum+=(t+1)*x[l0+l1+l2+l3+l4+l];
                    }
                    if (t+1==15) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7];
                    }
                    if (d->depart[i].tdepart >j+5) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7+l8];
                    }
              UY[i][j][t][n]  = sum;
          }

//  cancel decision
// j=15 (cancel)
              j=15;
              sum = 0.0;
              sum+=x[0];
               for (k=0; k<l1;k++ ) {
```

```
                    sum+=d->in[i].indiv[k]*x[l0+k];
                }
                 for (m=0; m<l3;m++ ) {
                    sum+=(d->curr->current[i][0])*x[l0+l1+l2+m];
                 }
                      for (l=0; l<l6;l++ ) {
                    sum+=(t+1)*x[l0+l1+l2+l3+l4+l5+l];
                      }
                      if (t+1==0) {
                          sum+=x[l0+l1+l2+l3+l4+l5+l6];
                      }
                UY[i][j][t][n] = sum;
            }
          }
      }
  // calculate mode real
      for(i = 0; i < d->glonum ->indivNum; i++) {
          for(t=1; t<d->period[i].tperiod+1; t++) {  //*
              for(n=0; n<2; n++){
                  sum=0;
                for (j = 0;j<d->glonum->numch;j++)
                  {
                  sum+=exp(UY[i][j][t][n]);
                  }
              r[i][t][n]= log(sum);
              }
         }
      }
         free_uy(UY,d->glonum);
         return r;
  }

  /* (III)
   v is randomly drawn from gumbel distribution with mode r_itn, also in the scenario tree;
   n indicates the position of v in the tree;
   v[i][t][n], n=0 is in the first level; n=1 is in the second level
  */
  double*** calculate_v(Aldata* d, double* x, double vp){
     int t,i,n;

     double*** v=  c_malloc_v(d->glonum);
     double*** r=calculate_mode(d, x);

      for(i = 0; i < d->glonum ->indivNum; i++) {
          for(t=1; t<d->period[i].tperiod+1; t++) {  //*
              for(n=0; n<2; n++){
              v[i][t][n]= st_gumbel_icdf(vp, r[i][t][n], 1 );
              }
          }
    }
      free_v(r, d->glonum);
      return v;
  }

  inline double max(double a, double b) {

        if (a<b) {
          return b;
        }
        return a;
  }

  /* (IV)
  p_ijt is probability of ticket change decision;
  p is calculated through traditional logit way;
  err1 is error term for i,j,t;
  err2 is error term for i;
  */
```

```
double*** cal_probcar (Aldata* d, double* x){

    int i=d->glonum->indivNum;
    int j=d->glonum->numch;
    int t;
    int  l,k,m;
    double sum=0;
    double ***err1, *err2;

    int l0= d->glonum->numC;         //asc
    int l1= d->glonum->numINDIVAR;   //indiv var
    int l2= d->glonum->numSTATIC;    //static var
    int l3= d->glonum->numREF;       //refund var
    int l4= d->glonum->numCUR;       //keep var
    int l5= d->glonum->numDF;        //day from var
    int l6= d->glonum->numDF1;       //day from var
    int l7= d->glonum->numCNL1;      //day 1 cancel
    int l8= d->glonum->numEXC16;     //day 16 exchange
    int l9= d->glonum->numEX;        //early exc
    int l10= d->glonum->numLX;       //late exc


    double ***P =c_malloc_u(d->glonum);

    double **w=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
    double ***U = c_malloc_u(d->glonum);

    err1 = d->err1;
    err2 = d->err2;

// calculate utility for 16 choices, from time 0 same as calculate_mode
    for(i = 0; i < d->glonum ->indivNum; i++) {
    int T=d->period[i].tperiod;
        for(t=0; t<T; t++) {
// exchange decision
// j=0-14 (deptt5-19)
          for(j=0; j<d->glonum->numch-1; j++) {

              sum = 0.0;

               for (l=0; l<l2;l++ ) {
                  sum+=((d->pot->stati[i][l2*j+l+t*(d->glonum->numch-1)*l2])-(d->curr-
>current[i][0]))*x[l0+l1+l];
              }

               for (l=0; l<l5;l++ ) {
                  sum+=t*x[l0+l1+l2+l3+l4+l];
                    }
                    if (t==15) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7];
                    }
                    if (d->depart[i].tdepart >j+5) {
                        sum+=x[l0+l1+l2+l3+l4+l5+l6+l7+l8];
                    }
              U[i][j][t] = sum+err1[i][j][t]+err2[i];
          }

//  cancel decision
// j=15 (cancel)
              j=15;
              sum = 0.0;
              sum+=x[0];

               for (k=0; k<l1;k++ ) {
                  sum+=d->in[i].indiv[k]*x[l0+k];
              }
               for (m=0; m<l3;m++ ) {
                  sum+=(d->curr->current[i][0])*x[l0+l1+l2+m];
```

```
            }
                 for (l=0; l<l6;l++ ) {
               sum+=t*x[l0+l1+l2+l3+l4+l5+l];
                 }
                 if (t==0) {
                     sum+=x[l0+l1+l2+l3+l4+l5+l6];
                 }
           U[i][j][t] = sum+err1[i][j][t]+err2[i];
        }
    }

    for(i = 0; i < d->glonum ->indivNum; i++) {
        for(t=0; t<d->period[i].tperiod; t++) {
                 sum=0;
               for (j = 0;j<d->glonum->numch;j++)
               {
                   sum+= exp(U[i][j][t]);
               }
                   w[i][t]=sum;
               for (j = 0;j<d->glonum->numch;j++)
               {
                   P[i][j][t]= exp(U[i][j][t])/w[i][t];
               }
        }
    }

           nt_matrix_free(w);

          free_u(U,d->glonum);

          return P;
}

/* (V)
recursive process for calculating E_AVE;
 n is the position in the tree
*/

double cal_E(Aldata* d, int t, int T, double *v, double current, int n, double* x, int
id){

    int i;
    int numch=16;
    double e_ave;
    double ***err1, *err2;
    double kp;
    err1 = d->err1;
    err2 = d->err2;

    int l0= d->glonum->numC;         //asc
    int l1= d->glonum->numINDIVAR;   //indiv var
    int l2= d->glonum->numSTATIC;    //static var
    int l3= d->glonum->numREF;       //refund var
    int l4= d->glonum->numCUR;       //keep var
    int l5= d->glonum->numDF;        //day from var
    int l6= d->glonum->numDF1;       //day from var
    int l7= d->glonum->numCNL1;      //day 1 cancel
    int l8= d->glonum->numEXC16;     //day 16 exchange
    int l9= d->glonum->numEX;        //early exc
    int l10= d->glonum->numLX;       //late exc

// Base case to cover the last time period

    if (t==15) {
    kp = x[l0+l1+l2+l3+l4-1]+err1[id][numch][t]+err2[id];
    }
    else {
    kp = err1[id][numch][t]+err2[id];
```

```
    }

    if (t==T) {      // Base Case
      return max(v[n], kp);           // (change, keep)
    }
    else {           // Recursive Step
      e_ave=0;
          for(i=0; i<1; i++){
           e_ave+=cal_E(d, t+1, T, v, d->curr->current[id][t+1],i+1, x, id);// go further
to reach the second level in the tree
          }
      e_ave=e_ave/1;
          return max(v[n], kp+e_ave); // (change, keep)
   }

}

/* (VI)
PI0,PI1 is probability of keep and change;
C_it is utility payoff when keep ticket
E_AVE, the average expectation at each time period;
*/
double** cal_prob (Aldata* d, double* x){

   double **PI0,**PI1, **C, **W, **E_AVE;
   double sum;
   double ***err1, *err2;
   int t, t2, i, n;
   int numch=16;

    int l0= d->glonum->numC;         //asc
    int l1= d->glonum->numINDIVAR;   //indiv var
    int l2= d->glonum->numSTATIC;    //static var
    int l3= d->glonum->numREF;       //refund var
    int l4= d->glonum->numCUR;       //keep var
    int l5= d->glonum->numDF;        //day from var
    int l6= d->glonum->numDF1;       //day from var
    int l7= d->glonum->numCNL1;      //day 1 cancel
    int l8= d->glonum->numEXC16;     //day 16 exchange
    int l9= d->glonum->numEX;        //early exc
    int l10= d->glonum->numLX;       //late exc


    err1 = d->err1;
    err2 = d->err2;

    C=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
    W=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
    E_AVE=nt_matrix_new(d->glonum-> indivNum, d->glonum->time+1);
    PI0=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);
    PI1=nt_matrix_new(d->glonum-> indivNum, d->glonum->time);

   double ***v=calculate_v(d,x,d->vp);
   double **r_real=calculate_mode_real(d,x);

    for(i = 0; i < d->glonum ->indivNum; i++) {

        for(t=0; t<d->period[i].tperiod; t++) { //*

        sum = 0;

        if (t==15) {
        sum+= x[l0+l1+l2+l3+l4-1];
        }
        else {
        sum+= 0;
        }
        C[i][t] = sum+err1[i][numch][t]+err2[i];
```

```
        }
    }

  // calculate expectations E_AVE in the first level of the tree in a recursive way;

    for(i = 0; i < d->glonum ->indivNum; i++)  {
    int T=d->period[i].tperiod;
        for (t=0; t<T; t++){
        t2 = t+2;       // t+(n-look ahead)

    E_AVE[i][t+1]=0;
        for(n=0; n<1; n++){
        E_AVE[i][t+1]+= cal_E(d, t+1, t2, v[i][t+1], d->curr->current[i][t+1], n, x, i);
        }
            E_AVE[i][t+1]=E_AVE[i][t+1]/1;

      W[i][t]= C[i][t]+E_AVE[i][t+1];           // keep reservation utility
        }
    }

 // calculate probabilities of keeping (PI0) with reservation utility W, mode r

    for(i = 0; i < d->glonum ->indivNum; i++) {
        for(t=0; t<d->period[i].tperiod; t++) {  //*
      PI0[i][t]=st_gumbel_cdf(W[i][t], r_real[i][t], 1);
      PI1[i][t]=1-PI0[i][t];
        }
    }

      nt_matrix_free(C);
      nt_matrix_free(W);
      nt_matrix_free(E_AVE);
      free_v(v,d->glonum);
      nt_matrix_free(r_real);
      nt_matrix_free(PI1);
     return PI0;
```

D-4: ll.c

```
#include <math.h>
#include <float.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

#include "library.h"
#include <ophelia/nlp.h>
#include <ophelia/nlp_collection.h>
#include <oratio/oratio.h>

/*PB is the departure time specific probability (j= 0,1,..,14);
PB=PI1*P
j=15 PB is the probability of cancel
*/
double fLL(double* x, int n, void* data) {

   Aldata* d = (Aldata*) data;
   double ***PB = d->prob_matrix;
   double ***P =cal_probcar(d,x);
   double **PI0=cal_prob (d,x);

   int i, j, t;
   int numch =16;
   double ***ch, LL;
   ch=c_malloc_P(d->glonum);


  for(i = 0; i < d->glonum ->indivNum; i++){
      for (t=0; t<d->period[i].tperiod; t++) { //*

       for (j = 0;j<d->glonum->numch+1;j++) {
           ch[i][j][t]=d->pot->decision[i][t*(numch+1)+j];
          }
      }
   }

  for(i = 0; i < d->glonum ->indivNum; i++){
     for (t=0; t<d->period[i].tperiod; t++) { //*

       for (j = 0;j<d->glonum->numch;j++)
        {
          PB[i][j][t]= (1-PI0[i][t])*P[i][j][t];
         }
          PB[i][j][t]= PI0[i][t];
       }

    }

   LL=0;
  for(i = 0; i < d->glonum ->indivNum; i++) {
    for (t=0; t<d->period[i].tperiod; t++) { //*

        for (j = 0;j<d->glonum->numch+1;j++)
          {
            LL+=ch[i][j][t]*log(PB[i][j][t]);
        }
      }
  }

  free_p(ch,  d->glonum);
  nt_matrix_free(PI0);

  free_u(P, d->glonum);
    return -LL;
```

```
}

/*  optimization;
H, preallocated array of size btr->n*btr-btr->n, if the hessian is needed;
I, Hessian matrix;
I1, inversed Hessian matrix;
*/

int btr_unconstrained_opt(NTLog *log, BTR *b, Aldata* d)
{

  double **H;
  double *t, *h;
  int n=get_dimension(d);
  int i;
  double tol=0;  //sets tolerance and scale for hessian derivation
  double scale[n];
  int s;
  FILE *out;
  out=fopen("matrix.txt","w");
  t=malloc(n*sizeof(double));

  double work[100*(b->n)];

  H = nt_matrix_new(n, n);
  nt_matrix_identity(n, n, *H, n);
  nt_log_subsection(log, "optim of Log likelihood");
  op_btr_init(b, n, 0);

  read_new_para(b->x, d->glonum);
  b->printer = btr_print_iteration;

  nlp_btr(b, (C_GENERIC)fLL, NULL, d, log->f, H, work);

  //derive hessian matrix

    for(s = 0; s < n; ++s)
    scale[s]=1.0;
    printf("'reach this part of code\n***********\n");
    h = malloc(n*sizeof(double));

    nt_derive_hess_cd((C_GENERIC)fLL, b->x, H, h, tol, scale, n, NULL, work, (void*) d);
    printf("'estimation completed \n");
    op_matrix_inverse(CblasRowMajor, CblasUpper, *H, n);

  double bugfound[n];
  int bugindex = 0;
  for(bugindex=0; bugindex<n; ++bugindex)
    bugfound[bugindex] = 0;

  amlet_t_statistics(n, b->x, bugfound, H, 0.05, t);
  // Print out inversed hessian matrix
  nt_matrix_print(out, "matrix", H, n, n);

   // Print out the t-statistics
   printf("t:");
   for(i=0; i < n; i++) {
     printf("  %f", t[i]);
   }

  return 0;
}


int main(int argc, char **argv) {

  BTR *b = malloc(sizeof(BTR));
  NTLog *log;
```

```
  Aldata *d = (Aldata*) format_data();

  int n=get_dimension(d);
  op_btr_init(b, n, 0);

  log = nt_log_new(NULL);

  btr_unconstrained_opt(log, b, d);

  nt_log_free(log);
  op_btr_free(b);

  free_ind(d->in, d->glonum);
  free_duration(d->period, d->glonum);
  free_schedule(d->depart, d->glonum);
  free_cur(d->curr, d->glonum);
  free_poten(d->pot, d->glonum);
  free_err(d);

  free_glo(d->glonum);
  free(d);

  return 0;
}
;
```

Glossary

CDLP	Choice-based Deterministic Linear Programming Problem
DCA	Discrete Choice Analysis
DDCM	Dynamic Discrete Choice Model
LC	Latent Class
MNL	Multinomial Logit
ML	Mixed Logit
RM	Revenue Management
RP	Revealed Preference
SP	Stated Preference

Bibliography

Algers, S. and M. Besser. 2001. Modeling choice of flight and booking class: A study using stated preference and revealed preference data. *International Journal of Services Technology and Management*, 2:28–45, 2001.

ALOGIT, 2007. User Manual ALOGIT version 4.2. ALOGIT Software & Analysis Ltd.

Amtrak. 2009. Audited consolidated financial statements - fiscal year 2009. http://www.amtrak.com/servlet/ContentServer/Page/1241245669222/1237608345018.

Amtrak 2010. The Northeast Corridor Infrastructure Master Plan. The NEC Master Plan Working Group. May, 2010.http://www.amtrak.com/ccurl/870/270/Northeast-Corridor-Infrastructure-Master-Plan.pdf

Amtrak 2012. The Amtrak Vision for the Northeast Corridor. 2012 Update Report. July, 2012. http://www.amtrak.com/ccurl/453/325/Amtrak-Vision-for-the-Northeast-Corridor.pdf

Andersson, S. 1998. Passenger choice analysis for seat capacity control: A pilot project in Scandinavian Airlines. International Transactions in Operation Research, 5(6): 471-486, 1998.

Association of American Railroads. 2009. Railroad statistics. 2009. http://www.aar.org/~/media/AAR/Industry%20Info/Statistics%2020090910.ashx.

Aviv, Y. and A. Pazgal. 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing Service Oper. Management* 10(3) 339–359.

Bastin F., C. Cirillo, and P. Toint. 2010. *Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations*, Transportation Science 44(4), pp. 537-549, 2010.

Bastin, F. 2011. Another Mixed Logit Estimation Tool (AMLET) Reference Manual. http://amlet.slashbin.net/

Belobaba, P. P. and C. Hopperstad. 1999. Boeing/MIT simulation study: PODS results update. *1999 AGIFORS Reservations Yield Management Study Group Sympos.* London, U.K.

Ben-Akiva, M., Improving Airline Passenger Forecasting Using Reservation Data, in *ORSA/TIMS Proc.*, St. Louis, MO, 1987.

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

Bertsimas, D. and S. de Boer. 2002. Joint network pricing and resource allocation, 2002.

Besanko, D. and W. L. Winston. 1990. Optimal price skimming by a monopolist facing rational consumers. *Management Science* **36**(5) 555–567.

Bhat, C.R. 1998. Accommodating Variations in Responsiveness to Level-of-Service Variables in Travel Mode Choice Modeling, *Transportation Research Part A,* Vol. 32, No. 7, 1998, pp. 495-507.

Bhat, C.R. 2000. Incorporating Observed and Unobserved Heterogeneity in Urban Work Mode Choice Modeling, *Transportation Science*, Vol. 34, No. 2, 2000, pp. 228-238.

Carrier, E. 2008. *Modeling the Choice of an Airline Itinerary and Fare Product using Booking and Seat Availability Data*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Castillo, M., L. Garrow and M. Lee. 2008. *When to Buy a Ticket*. Paper to the Institute for Operations Research and the Management Science Revenue Management and Pricing Section, Montreal, CA.

Chew, E.P., et al., Joint inventory allocation and pricing decisions for perishable products. International Journal of Production Economics (2008), doi:10.1016/j.ijpe.2008.07.018

Ciancimino, A., G. Inzerillo, S. Lucidi, and L. Palagi. 1999. A Mathematical Programming Approach for the Solution of the Railway Yield Management Problem. *Transportation Science* 33(2) 168.

Cirillo, C., P. Hetrakul, and S.Toobaie. 2011. Discrete choice model for Amtrak Acela Express revenue management , *Journal of Revenue and Pricing Management* 10, 492-513

Cizaire, C. 2011. Optimization Models for Joint Airline Pricing and Seat Inventory Control: Multiple Products, Multiple Periods. Ph.D. Dissertation, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology.

Cote, J., P. Marcotte and G. Savard. 2003. A bilevel modeling approach to pricing and fare optimization in the airline industry. Journal of Revenue and Pricing Management, 2(1):23, 2003.

Dana, J. 1998. Advance purchase discounts and price discrimination in competitive markets. *Journal of Political Economics.* 106(2) 395–422.

Davis, P. 1994. “Airline Ties Profitability to Yield Management,” *SIAM News—Mathematics that Counts,* 12 (1994).

Erdelyi, A. and H. Topaloglu. 2010. Using decomposition methods to solve pricing problems in network revenue management. J. Revenue Pricing Management 10(4) 325–343.

Gale, I. L. and T. Holmes. 1992. The efficiency of advance-purchase discounts in the presence of aggregate demand uncertainty. *International Journal of Organization* 10(3) 413–437.

Gallego, G. and O. Sahin. 2010. Revenue Management with Partially Refundable Fares OPERATIONS RESEARCH, Volume: 58 Issue: 4 Pages: 817-833, 2010

Gallego, G. 1996. A Demand Model for Yield Management, Technical Report, Columbia University, Dept. of Industrial Engineering and Operations Research, New York, NY, 1996.

Gallego, G., G. Iyengar, R. Phillips, and A. Dubey. 2004. Revenue management of flexible product. CORC Technical Report TR-2004-01, Department of Industrial Engineering and Operations Research, Columbia University, New York.

Gallego, G. and G. van Ryzin. 1997. A multi-product, multi-resource pricing problem and its applications to network yield management. *Oper. Res.* 45 24–41.

Gallego, G. and R. Phillips. 2004. Revenue management of flexible products. *Manufacturing Service Oper. Management* 6(4) 321–337.

Garrow, L. A. 2010. *Discrete Choice Modeling and Air Travel Demand: Theory and Applications, Ashgate Publishing*, U.K., 2010

Garrow, L. and F. Koppelman. 2004a. Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information. Journal of Air Transport Management 10, 401–411.

Garrow, L. and Koppelman. 2004b. Multinomial and nested logit models of airline passengers' no-show and standby behavior. Journal of Revenue and Pricing Management 3 (3), 237–253.

Glover, F., R. Glover, J. Lorenzo, and C. Mcmillan. 1982. The Passengers-Mix Problem in the Scheduled Airlines, *Interfaces* 12, 73–79 (1982).

Gowrisankaran, G. and M. Rysman. 2009. Dynamics of Consumer Demand for New Durable Goods, working paper, University of Arizona, Boston University.

Graham, R., L. Garrow, and J. Leonard. 2010. Business travelers' ticketing, refund, and exchange behavior. Journal of Air Transport Management 16 (2010) 196–201

Greene, W.H. and D.A. Hensher. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research B* Vol. 37, 2003, pp. 681–698.

Hetrakul, P. and C. Cirillo. Forthcoming. Accommodating taste heterogeneity in railway passenger choice models based on internet booking data, *Journal of Choice Modeling*

Hood, ISA. 2000. *MERLIN: Model to Evaluate Revenue and Loadings for Intercity*. 2nd ed. Yield Management: Strategies for the Service Industries, Thomson London, 98–110.

Hopperstad, C., "The Application of Path Preference and Stochastic Demand Modelling to Market Based Forecasting," in *AGIFORS Reservations and Yield Management Study Group Proc.*, Hong Kong, 1994.

Horowitz, J. 1982. An evaluation of the usefulness of two standard goodness-of fit indicators for comparing non-nested random utility models. *Transportation Research Record*, 874, 19‐25.

Iliescu, D.C. 2008. Customer Based Time-to-Event Models for Cancellation Behavior: A Revenue Management Integrated Approach, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Georgia Institute of Technology, 2008.

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

Jin, X. 2007. Toward Time-of-Day Modeling for Long Distance Trips, Doctoral Dissertation, University of Wisconsin, Milwaukee, 2007

Johnston, B., The scoop on sleeper pricing. *Trains Magazine*, 2003

Koppelman, F. and C. Bhat. 2006. *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. Report prepared for the U.S. Department of Transportation and Federal Transit Authority. Available at: http://www.civil.northwestern.edu/people/koppelman/PDFs/LM_Draft_060131Final-060630.pdf [accessed: 21 September 2008].

Koenigsberg, O., E. Muller, and N. J. Vicassim. 2006. Should Easy Jet offer last minute deals? Working paper, Columbia University, New York

Kuyumcu, A. and A., Garcia-Diaz. 2000. A polyhedral graph theory approach to revenue management in the airline industry. Computers and Industrial Engineering, 38(3):375-395, 10 2000.

Landsberger, M. and I. Meilijson. 1985. Intertemporal price discrimination and sales strategy under incomplete information. RAND J.Econom. 16(3) 424–430.

Levin, Y., J. McGill, and M. Nediak. 2009. Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. *Production Oper. Management* 19(1) 40–60.

Li, T., F. Hofker, and F. Jansma. 2006. Passenger Travel Behavior Model in Railway Network Simulation. Proceeding of the 2006 Winter Simulation Conference, 2006 IEEE

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

Liu, Q. and G. J. van Ryzin. 2005. Strategic capacity rationing to induce early purchases. Working paper, Graduate School of Business, Columbia University, New York.

McFadden, D. and K. Train. 2000. Mixed MNL models of discrete response. Journal of Applied Econometrics, 15:447-470, 2000.

McGill, J. and G. J. van Ryzin. 1999. Revenue management: Research overview and prospects. *Transportation Sci.* 33 233–256.

Melnikov, O. 2000. Demand for differentiated products: The case of the U.S. computer printer market. Unpublished manuscript, Cornell University.

Méndez-Díaz, I., JJ. Miranda-Bront, G. Vulcano, and P. Zabala. 2012. A branch-and-cut algorithm for the latent-class logit assortment problem. Discrete Applied Mathematics, doi:10.1016/j.dam.2012.03.003

Miller, R. 1984. Job Matching and Occupational Choice *The Journal of Political Economy*, Vol. 92, No. 6. (Dec., 1984), pp. 1086-1120.

Nair, H. 2006. Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video games. Quantitative Marketing and Economics, in press.

Neuling, S., S. Riedel, et al. 2004. New approaches to origin and destination and no-show forecasting: excavating the passenger name records treasure. Journal of Revenue and Pricing Management **3**(1): 62-72.

Office of Rail Regulation. 2009. National rail trends. http://www.rail-reg.gov.uk/server/show/nav.1862.

Ongprasert, S. 2006. Passenger Behavior on Revenue Management Systems of Inter-city Transportation. Ph.D. Dissertation. Graduate School of Engineering . Kochi University of Technology, Japan.

Pakes, A. 1986. Patents as options: Some estimates of the value of holding European patent stocks. Econometrica, 54:755-785, 1986.

Png, I. P. L. 1989. Reservations: Customer insurance in the marketing of capacity. *Marketing Sci.* 8(3) 248–264.

Railway Technology. 2011. Acela Express Specification. Available Online at http://www.railway-technology.com/projects/amtrak/specs.html

Ringbom, S. and O. Shy. 2004. Advance booking, cancellations, and partial refunds. *Econom. Bull.* 13(1) 1–7.

Rusmevichientong, P., D. Shmoys, and H. Topaloglu. 2012. Assortment optimization with mixtures of logits, Operation Research 60(4) 865–882.

Rust, J. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica, 55, 999–1033.

Rust, J. 1994. Numerical Dynamic Programming in Economics. Revised November1994 draft for Handbook of Computational Economics, 1994.

Shen, J. 2009. Latent class model or mixed logit model? A comparison by transport mode choice data, *Applied Economics* Vol. 41, No. 22, 2009, pp. 2915 - 2924

Shugan, S. and J. Xie. 2000. Advance pricing of services and other implications of separating purchase and consumption. *J. Service Res.* 2(3) 227–239.

Sibdari S., K.Y. Lin, K. Y., and S. Chellappan. 2008. Multiproduct revenue management: An empirical study of Auto train at Amtrak. Journal of Revenue and Pricing Management 7(2)172-184

Smith, W.R. 1956. Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing* Vol. 21, pp. 3–8.

Spinler, S., A. Huchzermeier, and P. Kleindorfer. 2002. An options approach to enhance economic efficiency in a dyadic supply chain. S. Seuring, M. Goldbach, eds. *Cost Management in Supply Chains*. Physica, Heidelberg, 350–360.

Spinler, S., A. Huchzermeier, and P. Kleindorfer. 2003. Risk hedging via options contracts for physical delivery. *OR Spectrum* 25(3) 379–395.

Stata 9.0 Stata Corp. College Station Texas. http://www.stata.com/

Stokey, N. 1979. Intertemporal price discrimination. *Quart.J.Econom.* 93(3) 355–371.

Su, X. 2007. Inter-temporal pricing with strategic customer behavior. *Management Sci.* 53(5) 726–741.

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

Talluri K. T. and G.J. van Ryzin. 2004a. The Theory and Practice of Revenue Management. Springer

Teichert, T., E. Shehu, and I. von Wartburg. 2008. Customer segmentation revisited: the case of the airline industry. *Transportation Research Part A*, Vol 42, 2008, pp. 227–242.

Topaloglu, H. 2009. Using Lagrangian relaxation to compute capacity-dependent bid prices in network revenue management. Operation Research 57(3) 637–649.

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

Vermunt J.K. and J. Magdison. 2005. *Technical Guide for Latent Gold Choice 4.0: Basic and Advanced.* Statistical Innovations Inc., Belmont, MA, 2005.

Vulcano G., G.J. van Ryzin , and C. Wassim. 2008. Choice based revenue management: An empirical study of estimation and optimization. January6, 2008

Weatherford, L. 1997. Using prices more realistically as decision variables in perishable-asset revenue management problems. Journal of Combinatorial Optimization, 1(3):277-304, 1997.

Wen, C.H. and S.C. Lai. 2010. Latent class models of international air carrier choice, *Transportation Research Part E* Vol. 46, 2010, pp.211–221

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

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

Wolpin, K. 1984. An estimable dynamic stochastic model of fertility and child mortality. Journal of Political Economy, 92:852-874, 1984.

Xie, J. and S. Shugan. 2001. Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Sci.* 20(3) 219–243.

Yu, M., R. Kapuscinski, and H. S. Ahn. 2008. Advance selling with limited capacity. Working paper, Ross School of Business, University of Michigan, Ann Arbor.

Zhang, D. 2005. Revenue Management with Customer Choice. Ph.D. Dissertation, Graduate School, University of Minnesota, 2005.

Zhang, D. and D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. Transportation Sci. 43(3) 381–394.