EVALUATION OF BENEFITS AND EFFECTIVENESS OF SMART CARDS FOR PUBLIC TRANSPORT

XU HAO

SUBMITTED IN ACCORDANCE WITH THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF LEEDS
INSTITUTE FOR TRANSPORT STUDIES

AUGUST 2007

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

ACKNOWLEDGEMENTS

This work is indebted to a number of people. First of all, I gratefully acknowledge the constant guidance of my supervisors: Prof Mark Wardman and Dr Haibo Chen. Meanwhile, I would like to thank Dr Gerard Whelan and Miss Hui Lu (Lucy) for invaluable help and suggestions on the RP/SP modelling analysis of this thesis. I also would like to thank Prof Xinlian Xie from Dalian Maritime University, who gave me suggestions and encouragement on my PhD research. I also wish to thank Miss Kaushali Dave, Mr James Jackson and Ms Margaret Watson for their help on English proof-reading of this thesis. Finally, I would like to thank PhD students in ITS, Leeds University. Sharing and exchanging ideas with them stimulated me on the research work.

I am particularly grateful to my parents for their support and encouragement along my PhD study.

Once again, thank you all, who offered me invaluable help and concerns. Without you, I could never achieve my PhD.

ABSTRACT

As a new technology, smart card ticketing for public transport has become increasingly popular across the world. Now smart cards and the traditional fare payment methods, cash and paper-based travel cards, have become three major fare payment options in public transport systems. The success of smart card applications across the world led to the realisation of the potential of smart card ticketing by some local governments and public transport service providers in China. For example, in Dalian, China, more than one million public transport smart cards have been issued since the payment application was just introduced in July 2001.

However, the traditional payment methods (i.e. cash and travel cards) are still in use in most Chinese cities. Passengers may choose between smart cards and traditional payment methods, according to their perceptions. Therefore, the aim of this research is to identify the fare payment preferences of passengers based on the existing and prospective situations for three fare payment methods (i.e. cash, travel cards and smart cards), to carry out the user demand analysis and provide an insight into the benefits and effectiveness of smart card ticketing.

The revealed preference (RP) and stated preference (SP) surveys were designed and carried out in Dalian, China, where the smart card project has been successfully implemented. In the data analysis, two different models are discussed: firstly standard logit models are used to analyse the joint RP and SP data. Secondly, two kinds of new techniques: fuzzy logic (FL) and artificial neural network (ANN) methods are introduced as an alternative to model discrete choice data. The motivation for using FL and ANN is that these two models can be non-linear and simulate human's decision making process without any *a priori* assumptions between inputs and outputs. The purpose of using FL and ANN in this research is to explore and compare the forecasting ability in the user demand analysis and model performance between new techniques and logit models.

Finally, results of the analysis, including forecasted market shares, valuation of attributes, fare elasticities, *etc*, indicate the increasing trend of smart card use in future development. Through monetary valuations, the importance of attributes is determined, such as multifunction and top-up/purchase options for smart cards. In addition, relevant policies are suggested to authorities to enhance the smart card payment service.

CONTENTS

ACKNOWLEDGEMENTS I
ABSTRACT II
CONTENTS III
LIST OF TABLES VII
LIST OF FIGURES X
GLOSSARY OF TERMS AND ABBREVIATIONS USED XI
Chapter 1 Introduction .... - 1 -
1.1. Research Background .... - 1 -
1.2. Research Objectives of Thesis .... - 3 -
1.3. Thesis Structure .... - 4 -
Chapter 2 Electronic Fare Payment for Public Transport .... - 7 -
2.1. Introduction .... - 7 -
2.2. Electronic Fare Payment and Its Technology .... - 7 -
2.2.1 Introduction to Electronic Fare Payment (EFP) .... - 7 -
2.2.2 EFP Technology .... - 8 -
2.3 History and Development of EFP in the World .... - 10 -
2.3.1 EFP History .... - 10 -
2.3.2 EFP Applications in the World .... - 12 -
2.4. Advantages and Disadvantages of EFP .... - 24 -
2.4.1 Advantages to Public Transport Operators .... - 24 -
2.4.2 Advantages to Public Transport Users .... - 29 -
2.4.3 Advantages to Other Sectors of the Society .... - 30 -
2.4.4 Disadvantages of EFP .... - 31 -
2.5 Enhancing Public Transport in China .... - 33 -
2.5.1. Background of the Current Chinese Public Transport Situation .... - 33 -
2.5.2. Current EFP Applications for Public Transport in China .... - 34 -
2.5.3. Necessity to Implement Electronic Fare Payment in China .... - 38 -
2.6. Summary .... - 39 -
Chapter 3 Literature Review .... - 40 -
3.1. Introduction .... - 40 -
3.2. Review of Evaluation Studies of EFP .... - 40 -
3.2.1 'Before and After' Studies .... - 41 -
3.2.2. Users' Demand (Preference) Studies .... - 47 -
3.2.3. Evaluation of Operators' Perceptions to EFP .... - 53 -
3.2.4. Cost-Benefit Analysis .... - 55 -
3.3. Key Findings and Significance to this Study .... - 58 -
3.4. Summary .... - 62 -
Chapter 4 Research Methodology .... - 64 -
4.1. Introduction .... - 64 -
4.2. Research Framework .... - 66 -

4.3. Revealed and Stated Preference Survey .... - 68 -
4.3.1 Revealed Preference (RP) Survey .... - 68 -
4.3.2 Stated Preference (SP) Survey .... - 69 -
4.3.3 Pilot Survey .... - 70 -
4.4. Logit Model Analysis .... - 71 -
4.4.1. MNL Models .... - 71 -
4.4.2 Hierarchical Logit Model .... - 73 -
4.4.3. Joint Analysis with RP and SP Data .... - 75 -
4.4.4 Model Validation .... - 79 -
4.5. Fuzzy Logic/Neural Network Techniques in Discrete Choice Analysis .... - 79 -
4.5.1. Fuzzy Logic Methods .... - 79 -
4.5.2. Neural Network Methods .... - 87 -
4.6. Model Application .... - 99 -
4.6.1 Demand Forecasting and Evaluation .... - 99 -
4.6.2 Segmentation Analysis .... - 101 -
4.7. Summary .... - 102 -
Chapter 5 Survey Design .... - 104 -
5.1. Introduction .... - 104 -
5.2. Definition of Population .... - 104 -
5.3. Revealed Preference (RP) Survey Design .... - 105 -
5.3.1. RP Questionnaire Types .... - 105 -
5.3.2. Variables in the RP Survey .... - 106 -
5.3.3. Conditional Questions .... - 107 -
5.3.4. Combinations of Different Payment Methods .... - 108 -
5.3.5. Demographic Questions .... - 109 -
5.4. Stated Preference (SP) Survey Design .... - 112 -
5.4.1. Determination of SP Games .... - 112 -
5.4.2. Determination of Attributes .... - 114 -
5.4.3. Determination of Attribute Levels .... - 117 -
5.4.4. Fractional Factorial Design and Choice Set Generation .... - 123 -
5.4.5. Presentation and Administration of the SP Survey .... - 133 -
5.4.6. Some RP Questions in the SP Survey .... - 134 -
5.4.7. Demographic Questions .... - 134 -
5.5. Pilot Surveys .... - 134 -
5.5.1. Sampling for the Pilot Surveys .... - 134 -
5.5.2. Findings and Lessons from the Pilot Survey .... - 136 -
5.5.3. Modifications of the Survey Design .... - 137 -
5.6. Summary .... - 138 -
Chapter 6 Data Collection .... - 140 -
6.1. Introduction .... - 140 -
6.2. Framework of Data Collection .... - 140 -
6.2.1 Survey Permission .... - 140 -

6.2.2 Survey Location .... - 141 -
6.2.3 Sampling .... - 141 -
6.2.4 Survey Presentation and Administration.................................. - 142 -
6.2.5 Survey Team Recruitment and Training................................... - 142 -
6.3. Survey Methods .... - 143 -
6.3.1 RP Survey Methods .... - 143 -
6.3.2 SP Survey Methods.. - 144 -
6.4. Revealed Preference (RP) Survey .... - 145 -
6.4.1 the RP Survey.. - 145 -
6.4.2 Basic Characteristics of the RP Data .... - 147 -
6.5. Stated Preference (SP) Survey .... - 154 -
6.5.1 the SP Survey .... - 154 -
6.5.2 Basic Characteristics of the SP Data... - 156 -
6.6. Basic Characteristics of the Respondents .... - 159 -
6.6.1 Age .... - 159 -
6.6.2 Gender... - 160 -
6.6.3 Educational Level .... - 160 -
6.6.4 Employment Status .... - 161 -
6.6.5 Household Income .... - 162 -
6.6.6 Availability of Private Transport .... - 163 -
6.6.7 Willingness to prepay PT Fare.. - 164 -
6.7. Issues in the Data Collection.. - 164 -
6.8. Summary .... - 167 -
Chapter 7 Data Analysis with Logit Models.. - 169 -
7.1. Introduction.. - 169 -
7.2. Modelling RP Data.. - 170 -
7.2.1. RP Input Data Preparation and Coding Scheme .... - 170 -
7.2.2. RP Choice Model and Parameter Estimation.......................... - 173 -
7.3. Modelling SP Data .... - 178 -
7.3.1. SP Input Data Preparation and Coding Scheme..................... - 179 -
7.3.2. SP Choice Models and Parameter Estimation......................... - 181 -
7.3.3 Combining SP Data Sets .... - 191 -
7.4. Data Enrichment Analysis with the RP and SP Data......................... - 198 -
7.4.1. Sequential Estimation of the RP and SP Data......................... - 198 -
7.4.2. Simultaneous Estimation of the RP and SP Data.................... - 200 -
7.5. Measuring the RP and SP Choices... - 203 -
7.5.1. Value of Boarding Time Savings.. - 204 -
7.5.2. Value of Other Attributes... - 204 -
7.5.3 Importance of Attribute.. - 208 -
7.6. Market Share Forecasting .... - 209 -
7.6.1. Market Share Forecasting for Different Fare Payment Methods.. - 210 -
7.6.2. Fare Elasticities .... - 216 -

7.7. Model Validation .... - 219 -
7.8. Market Segmentation Analysis .... - 221 -
7.8.1 Age .... - 222 -
7.8.2 Sex .... - 224 -
7.8.3 Household Income .... - 225 -
7.9. Principal Findings and Conclusions .... - 226 -
Chapter 8 Modelling Preference Data with Fuzzy Logic and Neural Network Technique .... - 230 -
8.1. Introduction .... - 230 -
8.2. Fuzzy Logic Analysis .... - 231 -
8.2.1 Inputs and Outputs .... - 231 -
8.2.2 Membership Functions .... - 232 -
8.2.3 Rule Base and Fuzzy Inference System (FIS) .... - 236 -
8.2.4 Estimation Results .... - 240 -
8.2.5 Model Calibration .... - 243 -
8.2.6 Discussions on FL Model Performance .... - 245 -
8.3. Artificial Neural Network Analysis .... - 248 -
8.3.1 Inputs and Outputs .... - 248 -
8.3.2 Network Structure .... - 249 -
8.3.3 Training and Learning Process for ANN Model .... - 250 -
8.3.4 Estimation Results and Model Validation .... - 252 -
8.4. Comparisons between MNL, FL and ANN Models .... - 256 -
8.4.1 General Model Expression of MNL, FL and ANN .... - 256 -
8.4.2 Model Forecasting Ability .... - 258 -
8.4.3 The Goodness of Fit Measures .... - 258 -
8.4.4 Interpretability of Outcomes .... - 259 -
8.5. Findings and Conclusions .... - 261 -
Chapter 9 Conclusions .... - 264 -
9.1. Introduction .... - 264 -
9.2. Summary of Achievements and Findings .... - 264 -
9.2.1 .... - 264 -
Evaluation of Benefits of Electronic Fare Payment .... - 264 -
9.2.2. Findings of Evaluation Methodology .... - 269 -
9.2.3. Contributions to Smart Card Applications from Policy Angle- 273 -
9.3. Recommendations for Future Work .... - 275 -
9.4. Conclusions .... - 278 -
Reference .... - 280 -
Appendix A: Comparisons of Different Features of Main Smart Card Applications in the World .... - 289 -
Appendix B: RP and SP Questionnaires .... - 291 -
Appendix C: Estimation Results in Segmentation Analysis .... - 309 -
Appendix D: Calibrated Membership Functions .... - 315 -
Appendix E: Fuzzy Rules (SP2, 3 and 4) .... - 321 -

LIST OF TABLES

Table 2.1 Characteristics of Card Technologies - 9 -
Table 2.2 Smart Card Transaction Volume, 1998-2003 ($million) - 12 -
Table 2.3 Summary of Features of Smart Cards for Public Transport - 23 -
Table 2.4 Smart Card Applications for Public Transport in China - 37 -
Table 3.1Cross-Attribute Analysis between Satisfactory Degree and Other Factors - 50 -
Table 3.2 Benefits and Costs of Smart Cards - 56 -
Table 5.1 The Layout of RP Questionnaires for Each Fare Payment Methods - 108 -
Table 5.2 Combinations of Fare Payment Methods in the RP Survey - 109 -
Table 5.3 Socio-economic Factors and Segmentations - 110 -
Table 5.4 SP Exercises and Suitability for Different Fare Payment Users - 112 -
Table 5.5 Attributes in the SP Experimental Design - 116 -
Table 5.6-1 Exercise 1: Cash vs. Travel Cards - 119 -
Table 5.7 Travel Cost Difference in SP 2 and SP 4 - 124 -
Table 5.8-1 Skeleton Design of SP 1: Cash vs. Travel Cards - 126 -
Table 5.9 SP Choice Situation-Cash vs. Travel Cards - 133 -
Table 6.1 The Response of the RP Survey - 145 -
Table 6.2 Respondents' Choices on Three Fare Payment Methods - 146 -
Table 6.3 Ticket type used or could be used by the respondents - 147 -
Table 6.4 Travel card type used or could be used by the respondents - 148 -
Table 6.5 Smart card type used or could be used by the respondents - 148 -
Table 6.6 Actual User Travel Cost - 149 -
Table 6.7 Possible User Travel cost - 150 -
Table 6.8 Boarding Time Difference of Actual Users (cash: baseline) - 150 -
Table 6.9 Boarding Time Difference of Possible Users (cash: baseline) - 151 -
Table 6.10 Seat Availability by using Travel Cards and Smart Cards - 152 -
Table 6.11 Whether More Trips (were or could be) Made - 152 -
Table 6.12 Overall Assessment of the Three Payment Methods - 153 -
Table 6.13 Choices on Methods to Buy/renew Travel Cards - 153 -
Table 6.14 Choices on Methods to Buy/top up Smart Cards - 153 -
Table 6.15 Other Functions of Smart Cards (were or could be) Used - 154 -
Table 6.16 The response of the SP survey - 155 -
Table 6.17 Fare Payment User Type (Primary) in All SP Exercises - 156 -
Table 6.18 Cash Fare Payment Choices in SP1 and SP2 - 157 -
Table 6.19 Travel Card Type Choices in SP1, SP2 and SP4 - 157 -
Table 6.20 Smart Card Type Choices in SP2, SP3 and SP4 - 157 -
Table 6.21 Statistic Characteristics on Number of Trips by Cash - 158 -
Table 6.22 Statistic Characteristics on Number of Trips by Travel Cards - 158 -
Table 6.23 Statistic Characteristics on Number of Trips by Smart Cards - 158 -
Table 6.24 Share of Payment Type in the SP Survey - 159 -
Table 6.25 Illogical Background Data in the RP Survey - 166 -
Table 6.26 Summary of the Distribution of Different RP and SP Data - 167 -
Table 7.1 Variables and Codes for Cash Fare Payment - 171 -

Table 7.2 Variables and Codes for Travel Card Fare Payment - 171 -
Table 7.3 Variables and Codes for Smart Card Fare Payment - 172 -
Table 7.4 RP Estimation Results of Standard MNL Model - 175 -
Table 7.5 RP Estimation Results of Two HL Models - 178 -
Table 7.6 Variables and Codes for SP-1: Cash vs. Travel Cards - 179 -
Table 7.7 Variables and Codes for SP-2: Cash vs. Smart Cards (Pay as you go) - 180 -
Table 7.8 Variables and Codes for SP-3: Travel Cards vs. Smart Cards - 180 -
Table 7.9 Variables and Codes for SP-4: Travel Cards vs. Smart Cards - 181 -
Table 7.10 Estimation Results of SP 1: Cash vs. Travel Cards - 187 -
Table 7.11 Estimation Results of SP 2: Cash vs. Smart Cards (Pay as you go) - 188 -
Table 7.12 Estimation Results of SP 3: Travel Cards vs. Smart Cards (Pay as you go) - 189 -
Table 7.13 Estimation Results of SP 4: Travel Cards vs. Smart Cards (Pay monthly) - 190 -
Table 7.14 Estimation Results of Combined SP Data Sets with Three Scale Factors - 193 -
Table 7.15 Estimation Results of Combined SP Data Sets with Scale Factor for SP 3 - 195 -
Table 7.16 Sequential Estimation Results - 199 -
Table 7.17 Simultaneous Estimation Results - 201 -
Table 7.18 Comparison of Estimates of Common Attributes in Different Models - 202 -
Table 7.19 Value of Boarding Time Savings in Different Models - 204 -
Table 7.20 Valuation of Attributes for Mixed RP and SP Data - 205 -
Table 7.21 Demand Forecast Based on Cost Variable - 210 -
Table 7.22 Demand Forecast Based on Boarding Time Variable - 211 -
Table 7.23 Demand Forecast Based on Service Routes Variable - 212 -
Table 7.24 Demand Forecast Based on Overdraft Variable - 213 -
Table 7.25 Demand Forecast Based on Deposit Variable - 213 -
Table 7.26 Demand Forecast Based on Multifunction Variable - 213 -
Table 7.27 Demand Forecast Based on Geographic Areas Variable - 214 -
Table 7.28 Demand Forecast Based on Top-up/purchase Options Variable - 214 -
Table 7.29 Demand Forecast—Multiple Changes of Variables of SC - 215 -
Table 7.30 Own and Cross Fare Elasticities in the Pure RP Data - 216 -
Table 7.31 Own and Cross Fare Elasticities in the Pure SP Data - 216 -
Table 7.32 Own and Cross Fare Elasticities in the Joint RP and SP Data - 216 -
Table 7.33 Demand Forecast Based on Cost Variable by Validation Sample - 219 -
Table 7.34 Demand Forecast Based on Boarding Time Variable by Validation Sample - 219 -
Table 7.35 Demand Forecast Based on Service Routes Variable by Validation Sample - 219 -
Table 7.36 Demand Forecast Based on Overdraft Variable by Validation Sample - 219 -
Table 7.37 Demand Forecast Based on Deposit Variable by Validation Sample - 220 -
Table 7.38 Demand Forecast Based on Multifunction Variable by Validation Sample - 220 -
Table 7.39 Demand Forecast Based on Geographic Areas by Validation Sample - 220 -
Table 7.40 Demand Forecast Based on Top-up/purchase Options Variable by Validation Sample - 220 -
Table 7.41 Forecast Market Share of Multiple Changes in Validation Sample - 220 -
Table 7.42 Correlations of Segmentation Variables - 221 -
Table 7.43 Segmentation Results by Age Factor: Value of Attributes - 223 -

Table 7.44 Smart Card Demand Forecast by Age Factor - 223 -
Table 7.45 Segmentation Results by Sex Factor: Value of Attributes - 224 -
Table 7.46 Smart Card Demand Forecast by Sex Factor - 224 -
Table 7.47 Segmentation Results by Household Income Factor: Value of Attributes - 225 -
Table 7.48 Smart Card Demand Forecast by Household Income Factor - 226 -
Table 8.1 Fuzzy Rules in the RP Survey - 237 -
Table 8.2 Fuzzy Rules in the SP-1 - 238 -
Table 8.3 Predicted Share by three FL Models in the RP Survey - 241 -
Table 8.4 Predicted Share by three FL Models in SP 1 - 241 -
Table 8.5 Predicted Share by three FL Models for SP 2 - 242 -
Table 8.6 Predicted Share by three FL Models for SP 3 - 242 -
Table 8.7 Predicted Share by three FL Models for SP 4 - 243 -
Table 8.8 Comparison between the Basic Model and Calibrated Model-1 - 244 -
Table 8.9 Comparison between the Calibrated Model-1 and Calibrated Model-2 - 245 -
Table 8.10 Uncertainty of Choice Behaviour in FL Models - 246 -
Table 8.11 Uncertainty of Choice Behaviour in Logit Models - 247 -
Table 8.12 Comparison of the Goodness of Fit (RMSE) of FL Models - 248 -
Table 8.13 Number of Neurons for Each Layer in the ANN Model - 250 -
Table 8.14 RP Prediction Results by ANN Model - 254 -
Table 8.15 Prediction Results by ANN Model for SP 1 - 255 -
Table 8.16 Prediction Results by ANN Model for SP 2 - 255 -
Table 8.17 Prediction Results by ANN Model for SP 3 - 255 -
Table 8.18 Prediction Results by ANN Model for SP 4 - 256 -
Table 8.19 Comparison of Goodness of Fit of MNL, FL and ANN Models: RP - 259 -
Table 8.20 Comparison of Goodness of Fit of MNL, FL and ANN Models: SP - 259 -

LIST OF FIGURES

Figure 2.1 2003 Contactless & Dual Interface Smart Cards Delivery ....- 12 -

Figure 2.2 Smart Card Growth in Hong Kong ....- 19 -

Figure 4.1 Research Design of this Thesis..- 66 -

Figure 4.2 Hierarchical Structure of the RP Alternatives (1).......................................- 73 -

Figure 4.3 Hierarchical Structure of the RP Alternatives (2).......................................- 74 -

Figure 4.4 Hierarchical Structure of SP Data Combination ....- 74 -

Figure 4.5 Artificial Tree Structure for Joint RP and SP Estimation ....- 78 -

Figure 4.6 the Fuzzy Decision Making Framework..- 80 -

Figure 4.7 A Three Layer Artificial Neural Network with Biases.................................- 88 -

Figure 4.8Neural Network with Softmax Outputs...- 92 -

Figure 5.1 Flow Chart of the RP Survey Design ....- 111 -

Figure 6.1 Sample Background Distribution by Age..- 160 -

Figure 6.2 Sample Background Distribution by Gender...- 160 -

Figure 6.3 Sample Background Distribution by Educational Level...........................- 161 -

Figure 6.4 Sample Background Distribution by Employment Status.........................- 162 -

Figure 6.5 Sample Background Distribution by Household Income per Month ....- 163 -

Figure 6.6 Sample Background Distribution by Individual Transport Availability .....- 163 -

Figure 6.7 Sample Background Distribution by Willingness to Prepay PT Fare ....- 164 -

Figure 8.1 Input Membership Functions for Cash Fare Payment in RP......................- 233 -

Figure 8.2 Input Membership Functions for Travel Cards in RP ....- 234 -

Figure 8.3 Input Membership Functions for Smart Cards in RP ....- 234 -

Figure 8.4 Output Membership Functions in RP...- 234 -

Figure 8.5 Input Membership Functions of Cash and Travel Cards in SP 1.................- 235 -

Figure 8.6 Output Membership Function in SP 1...- 235 -

Figure 8.7 Example of Fuzzy Inference Diagram for SP 1 ....- 239 -

Figure 8.8 Effects of Travel Cost on Choices...- 261 -

GLOSSARY OF TERMS AND ABBREVIATIONS USED

ANN	Artificial Neural Network
ASC	Alternative Specific Constant
BP	Backpropagation
C	Cash
CBA	Cost Benefit Analysis
EFP	Electronic Fare Payment
FF	Feedforward
FIS	Fuzzy Inference System
FL	Fuzzy Logic
HL	Hierarchical Logit
IIA	Independence of Irrelevant Alternatives
IID	Independently and Identically Distributed
LM	Levenberg-Marquardt Algorithm
LR	Likelihood Ratio
MF	Membership Function
MMS	Mean of Market Share
MNL	Multinomial Logit
MSE	Mean Square Error
NPV	Net Present Value
PT	Public Transport
RMSE	Root Mean Square Error
RP	Revealed Preference
RUT	Random Utility Theory
SC	Smart Cards
SP	Stated Preference
TC	Travel Cards
VOBTS	Value of Boarding Time Savings
VSE	Variance of Square Error
WTPr	Willingness to Prepay

Chapter 1
Introduction

1.1. Research Background

As a new technology for enhancing public transport services, smart card ticketing for public transport is becoming increasingly popular across the world. Now smart cards and traditional fare payment methods, cash and travel cards, have become the three major fare payment options in public transport systems. Smart card technology and its application to public transport have been presented and discussed in many previous relevant studies (Blythe, 2004; Chambers, 1998; Laconte, 1998). One of the advantages of the smart card technique in public transport systems is that passengers can avoid handling cash for payment when boarding, therefore journey time can be reduced. Moreover, smart cards have greater security, higher reliability, and higher resistance to fraud than other payment means. Recently, dual-interface smart cards using a single chip to communicate with smart card readers and other terminals have been developed, because they have a single integrated platform for contact and contactless applications. They may prove to be more popular for multi-application schemes facilitating cooperation across industrial sectors (Casey, 2000).

Some large-scale smart card ticketing schemes have been carried out successfully in North America, Europe and Asia, the leading examples being the London Oyster card, the Paris Claypso card, the Hong Kong Octopus card and the Seoul 'T-money' card. Beyond applications in public transport fare payment/collection, these projects have also suggested the versatile functions of smart cards in other social services, for example, Transport for London, UK, announced a breakthrough in smart card ticketing systems in January 2007 with the news that, Transys, the consortium behind the Oyster card, has signed a deal with Barclaycard Visa enabling the finance house to provide customers with a new 'chip and PIN' credit card that also has an Oyster card facility (TfL, 2007). Another example is that the Hong Kong Octopus card has had other payment functions added, e.g. shopping and parking fee payment, in addition to public transport fare payment functions.

The success of smart card applications throughout the world has led to the realisation of the potential of smart card ticketing by some local governments and public transport service providers in China. Recently, with the implementation of smart card projects for public transport in some Chinese big cities (e.g., Beijing, Shanghai, Shenzhen, Dalian, *etc.*), smart cards have become one of the major payment options in China, in addition to cash and travel cards. For example, in Dalian, the number of smart card owners has been more than 1 million since 2001, approximately 30% of the population in the urban area. Under these circumstances, passengers could have various choices of payment means to suit their different perceptions. To promote new technology for public transport fare payment, it is necessary to understand users' choice behaviours and the users' demand of different fare payment methods so as to explain benefits of fare payment methods to PT users. In the mean time, users' demand forecast of fare payment options can feed back to the further improvement of smart card ticketing, so as to enhance the service quality of public transport. Although some evaluation studies of smart card applications have been carried out in recent years, all of them focused on current applications and a systematic analysis of users' demand, however user preference has not yet been explained clearly, particularly for investigating respondents' trading off behaviours between conventional payment methods (i.e., cash and travel cards) and the smart card ticketing based on the combination of attribute-level of payment alternatives.

In order to capture individual preferences under new choice situations, the stated preference (SP) approach has proven to be successful (Louviere *et al.*, 2000) in transportation studies. The advantages of the SP methods for this research are:

- It allows modelling of new alternatives, attributes or variations in the attributes of existing ones;
- The degree of correlation and variation between attributes of different fare payment methods may be controlled.

Therefore, this research is based on a stated preference (SP) survey with binary-choice situations among public transport users in Dalian, China, where the smart card project has been successfully implemented for more than five years, to collect preference data towards three fare payment applications, i.e. cash, travel cards and smart cards. In addition to the SP survey, the revealed preference (RP) method is suitable for investigating respondents' actual choice behaviour. Currently, cash, travel cards and smart cards are in use in the city, therefore, through using RP and SP survey, we can make use of advantages of each survey method to understand PT users' perceptions towards different fare payment methods when carrying out demand forecast analysis to evaluate benefits of smart cards.

The RP and SP surveys contained current and designed attributes and levels for fare payment options and were used in this study to identify public transport passengers' actual choices and preferences under those hypothetical situations. So this research aims to explain benefits and effectiveness of smart cards to PT users by analysing passengers' fare payment choice behaviour and demand forecast with stated preference and revealed preference data. Furthermore, the evaluation result is expected to suggest the relevant policies to be implemented in the future public transport ticketing applications.

1.2. Research Objectives of Thesis

The focus of this research is to analyse public transport users' choice behaviour and users' demand of fare payment through RP and SP survey in a city where a variety of fare payment methods (i.e., cash, travel cards and smart cards) have been implemented, to reveal the benefits of smart card ticketing to PT users. The objectives of this research which support the aim above are stated as follows:

- First of all, review different evaluation methods about smart card applications in public transport. Through the literature review, determine the analysis method suitable for assessing the benefits of smart cards and some other payment methods from demand side.
- Through the users' preference survey, this research is to explore and develop an understanding of the existing fare payment applications and to ascertain other potential uses of smart cards and how the PT payment system can improve its utilisation and the services provided. The importance of features of smart cards is expected to obtain from the users' preference data so that how these payment attributes were perceived by respondents would be identified.
- Produce a demand forecast analysis based on the RP and SP survey. The RP survey can be used to forecast the users' demand in the short term. Through investigating trading off behaviour in the SP survey, particularly based on payment alternatives with some new features or new levels of current payment attributes, changes of choice behaviour of respondents can be identified. Therefore, in this research benefits of smart cards to PT users are to be explained by these choice behavioural changes (users' demand) when some payment services were improved.
- Finally, through analysing respondents' perceptions to different fare payment methods, the relevant policies and suggestions for enhancing the PT fare payment applications can be raised.

Meanwhile, in the modelling analysis of discrete choice data, two techniques (fuzzy logic and artificial neural network) are introduced to explore the improvement of forecasting ability and model performance. Therefore, another objective of this thesis is to explore and assess whether FL/ANN methods are a feasible alternative to MNL models for forecasting PT fare payment methods shares in the market place through individual preference data.

1.3. Thesis Structure

In total, nine chapters are discussed in this thesis. Except this chapter, the other eight chapters are organised as follows:

Chapter 2 introduces an overview of electronic fare payment (EFP) for public transport, including technological features of smart card ticketing for public transport, the history and current applications of smart cards throughout the world, and advantages and disadvantages of smart cards to PT users, operators, and local governments. In addition to the overview of smart card applications in the world, the Chinese situation of public transport and EFP applications is outlined as the research context in this chapter. It is intended that this chapter can act as the research background that yields the relevant research methodology.

Chapter 3 reviews literature concerned with the relevant evaluation studies on smart cards for public transport. The previous evaluation studies are categorised by four parts, including "before and after" studies, users' preference studies, operators' perception studies and cost-benefit analysis. The purpose of this chapter is to have an insight into previous evaluation studies of benefits and effectiveness of smart cards for public transport and get implications of the research methodology of this thesis. In "before and after" studies, mainly the previous studies focused on evaluating performance of smart cards projects comparing with conventional fare payment methods. However, most smart card applications in evaluation studies (e.g., 'before' and 'after' studies) were only implemented on a small scale for testing. Users' revealed preferences towards smart cards were used in some evaluation studies, such as the Hong Kong Octopus card, however only a basic statistical analysis was carried out for the RP survey data. Further demand forecasting analysis may be done based on the survey data, therefore, this is the motivation to carry out the preference survey and demand analysis in this research. Beyond analysing users' demand, some literature also focused on public transport operators' perceptions of smart card applications, such as the point of view of public transport staff about smart card ticketing.

After the discussion of the research context and literature review, in **Chapter 4** the detailed methodology for this research is presented. First of all, a research design (framework) is outlined. Following the framework, each stage of the research methodology is discussed respectively, including: reasons for using the revealed and stated preference survey particularly in the Chinese context and their advantages and drawbacks; the data analysis methods with the logit model and the fuzzy logic (FL) and artificial neural network (ANN) methods. Finally, the model applications of analysing fare payment choice behaviour are discussed in this chapter

Chapter 5 describes the survey design of revealed preference (RP) and stated preference (SP) prior to the data collection (Chapter 6). First of all, the survey population and location are determined. If the RP and SP were carried out in one questionnaire, the survey would become excessively long, therefore, the RP and SP were separately designed and carried out. The RP survey was basically used to collect respondents' actual choice behaviour. The SP survey was used to collect preference data based on hypothetical situations. In the SP survey design, the first task is to determine the SP games being used to collect preference data from respondents, then the relevant attributes and levels were selected to generate choice profiles by using fractional factorial design technique (Pearmain *et al*. 1991). In order to test the survey design, a pilot survey was discussed before the main survey in this chapter, and findings and lessons from the pilot survey suggested some modification for finalising the survey design.

Following the survey design and the pilot survey in Chapter 5, **Chapter 6** describes the main data collection for this research, which was conducted in Dalian, China during July and August 2005. In both the RP and SP data collection, on-board surveys were mainly used, because this is the best way to approach public transport users for data collection. With the cooperation of the local authority and public transport operators, 869 RP questionnaires and 896 SP questionnaires were returned and could be used in the later data analysis, with a response rate of about 58% overall. Meanwhile, some basic data preparation for the data analysis had been done, including analysing basic characteristics of the RP and SP data, respondents' characteristics and checking the validity of the RP and SP data. Another task carried out in this chapter is to compile the data for the data analysis in following chapters.

After the data compilation in Chapter 6, the next task is to conduct users' demand forecast and measure the benefits of smart cards through outcomes of the modelling analysis. The data analysis is divided into two chapters. First of all, **Chapter 7** discusses MNL models about analysing preference data, including, RP, SP and joint RP and SP models. In addition to the model estimation, the model application is discussed in this chapter, including valuation of attributes, market share forecasting, and fare elasticities.

Finally, effects of socio-economic factors (age, sex and household income) on different fare payment options and segmentation analysis are discussed.

Chapter 8 continues the data analysis based on the RP and SP data. This chapter discusses the fuzzy logic (FL) and artificial neural network (ANN) techniques in modelling discrete choice data. The motivation to introduce FL and ANN techniques in this research is to explore improvements in forecasting ability and model performance because of the non-linearity of these two methods. Moreover, comparisons between MNL models in Chapter 7 and FL, ANN models are made in this chapter. In general, FL and ANN models show the advantages of improvement of forecasting ability over the MNL models. However, the pros and cons of using different modelling techniques on discrete choice data under different conditions are also discussed at the end of this chapter.

Finally, **Chapter 9** provides a summary of the achievements of this research and identifies areas that would benefit from further studies. Meanwhile, the relevant policies concerning future fare payment applications are presented in this chapter.

Chapter 2
Electronic Fare Payment for Public Transport

2.1. Introduction

The purpose of this chapter is to introduce an overview of electronic fare payment (EFP) for public transport, including its development, technology and advantages over conventional fare payment methods. In addition, the Chinese situation of public transport and EFP applications is outlined as the research context for this study. It is intended that this chapter acts as the research background.

The structure of this chapter is as follows: in Section 2.2 the definition and technological features of smart card ticketing for public transport are introduced. Section 2.3 overviews the history and development of EFP applications around the world. In Section 2.4 advantages and disadvantages of EFP are presented from three different angles: to PT operators, to PT passengers, and to other components of society (e.g., local governments, security, environment, *etc*). Section 2.5 specifically discusses the Chinese context for this research, including the general public transport situation, current issues in the public transport system, the current applications of fare payment (including cash, travel cards and smart cards) in China, necessity to implement EFP applications in China and need for this research.

2.2. Electronic Fare Payment and Its Technology

2.2.1 Introduction to Electronic Fare Payment (EFP)

EFP systems for public transport are of two types—those that use magnetic stripe cards and those that use smart cards (SC). Magnetic stripe cards require a contact between the card's stripe and a device that validates the card for the trip taken (e.g., a monthly pass) or a read-write device that can deduct the fare from the value stored on the card and restore the remaining balance. Another type of EFP application is the smart card ticketing. The smart card technology is by no means new, because it was invented more than 30 years ago and implementations have been made with smart cards for almost two decades in social services, such as telecommunication, banking and identification cards, *etc* (Blythe, 2004).

A smart card that contains a microprocessor may interface with the reader by direct contact or by radio frequency. So, most smart cards can have both contact and contactless interfaces. Compared with cash fare payment, the main advantage of contactless smart cards is that passengers only need to pass the card reading device with smart cards instead of handling the exact cash to pay fare, hence the journey time can be effectively saved.

Moreover, smart cards are pre-paid, so card users do not have to pay fare for each trip (like cash). Additionally, smart cards have greater security, higher reliability, and higher resistance to fraud than magnetic stripe cards (Casey, 2000) because of the more advanced technology used than magnetic strip cards. The detailed advantages of smart cards over magnetic strip cards are presented in Section 2.4. Magnetic strip cards have been become obsolete in public transport systems, therefore this research focuses on smart card applications as the research context rather than magnetic stripe cards.

Regarding the definition of smart cards, they can be categorised as follows, according to the form and quantity of memory and the logic capabilities (Blythe, 2004):

1). Memory cards—These are credit-card-sized integrated circuit cards that store information but do not contain onboard microprocessors.

2). Microprocessor cards—These are credit-card-sized integrated circuit cards that have internal logic capabilities because of the presence of a microprocessor; in other words, they are essentially tiny computers. An advanced version is the "super smart card", which includes a miniature keypad and display.

When defined on the basis of communication technique, smart cards can be categorised as follows (Wikipedia, 2006):

1). Contact cards—These cards (memory or microprocessor) require a physical contact between the card and the reader-writer unit.

2). Contactless cards—These "remote coupling" or "close coupling" cards use a contactless interface to provide power to the card and transfer data using inductive and capacitive techniques. Among contactless smart cards, RF identification (RFID) cards or tags—RFID cards transfer data between the card and the reader-writer unit using RFID induction technology. These cards require only close proximity to an antenna to complete transaction. They are often used when transactions must be processed quickly or hands-free, such as on mass transit systems, where smart cards can be used without even removing them from a wallet.

2.2.2 EFP Technology

The term of "smart card" has been used to describe a range of card classifications and technologies in the world. The microchips embedded in the smart cards can be computer chips, capable of both storing and processing information, or memory chips, which are capable only of accessing data already stored on the card (Bagchi and White, 2004).

The manner in which a smart card can be used depends on how the chip in the card interfaces with the card reader machines (e.g. a bus ticket machine). The chips on the card can have either a contact or a contactless interface. With contact smart cards, the chip is connected to the surface of the card. In order to be used, these cards have to come into contact with the device they are required to communicate with. These devices are known as 'read/write' devices or 'readers' or 'terminals'. Contactless cards, theoretically, do not have to come into physical contact with the device they are going to communicate with. Contactless cards that communicate within a range of 10 cm and conform to the international card standard ISO/IEC 14443 are known as 'proximity' cards. Those that can communicate at distances of up to 70 cm and conform to ISO 15693 are known as 'vicinity' cards.

In recent years, hybrid or dual-interface technology has become an important technology for smart card applications. Hybrid or dual-interface cards refer to smart cards that can support both the contact and contactless interface. Cards are described as hybrid when the independent contact and contactless technologies share a single card and do not communicate with one another. Moreover, dual-interface smart cards (also referred to as combi-cards) have emerged in recent years, which have a single chip that can communicate with the smart card readers and other terminals using the contact or contactless interface. Dual-interface cards are cheaper in cost terms than hybrid cards. Also, because they have a single and integrated platform for contact and contactless applications, they may prove more popular than hybrid cards for multi-application schemes facilitating cooperation in multi-application projects across industrial sectors (Bagchi and White, 2004).

Table 2.1 Characteristics of Card Technologies

Criterion	**Magnetic Stripe Card**	**Contact Smart Card**	**Contactless Smart card**
Convenience	Must be inserted or swiped	Must be inserted into card reading devices	Approach card reading devices within 8-10cm
Protection if lost or stolen	Moderate	High	High
Boarding time/through fare gates	Depends on format, lower than contactless cards	Lower than contactless cards	Quicker than contact EFP technology
Standardisation	Standard exists (for stripe cards)	ISO 7816	ISO 14443
Capital Cost	Low	High	High
Operating& Maintenance Cost Impact	Highest equipment maintenance cost	Longer life for cards than magnetic stripe cards; cards can be re-used	Lowest equipment maintenance cost; longer life for cards; cards can be re-used
Data Capacity (i.e., user information)	Up to 0.2 KB	Up to 8 KB	Up to 64 KB

Source: Blythe, 2004; Fleishman, 1996; Wikipedia, 2006

Table 2.1 lists and compares the characteristics of different EFP technology in public transport. It can be seen that contactless smart cards take more advantages on convenience and security, *etc*, over magnetic stripe cards and contact smart cards in practice. For example, magnetic strip cards and contact smart cards must be inserted or swiped in card reading devices to pay fare, but contactless smart cards only need to approach card reading devices with 8-10cm distance and then the transaction can be done. Also, due to the different check-in procedure, the throughputs (boarding time/through fare gates) of contactless smart cards are quicker than contact FEP methods.

Standardisation of EFP is also moving forward with final draft standards for contactless cards now published (ISO 14443) and some application standards emerging in Europe, from the standards committees CEN TC224 and CEN TC278 (Wikipedia, 2006). Inevitably, as with all rapidly developing techniques, some issues remain to be resolved, such as memory, security and interface, *etc*. But as can be seen, based on the standardisation for contactless smart cards, the data capacity has been increased several times than contact EFP methods.

Regarding the operating and maintenance cost of different EFP methods, magnetic strip cards have the highest maintenance cost, particularly on card reading devices and card manufacture costs, because most magnetic strip cards cannot be re-used and the life duration is relatively short. But compared with magnetic strip cards, smart cards can reduce card manufacture cost because of longer life in use.

2.3 History and Development of EFP in the World

2.3.1 EFP History

When reviewing the history and development of EFP for public transport industry across the world, it can be seen that the great changes took place in fare payment media during the last decade. Because of the technological advantages of EFP in practical applications as discussed in the previous section, now the smart card technology is replacing the traditional fare payment types and outdated EFP applications (i.e., magnetic stripe cards) in the worldwide public transport systems.

The potential of smart cards for public transport has received increasing attention during the past ten years. In the early 1990's, some small scale smart card projects for public transport were successfully carried out and tested in some countries, such as US, UK, France, *etc*. (Fleishman, 1996; Ampelas, 1998; Blythe, 2004; ITSO, 2006). For instance, in 1991, London underground implemented contact smart cards as yearly passes. However, most early smart card applications were based on the contact communication technology, which required users to insert or swipe smart cards in card reading devices. The throughput and reliability of the systems were not satisfactory. With the development of smart card technology, the emergence of contactless cards has sparked the interest of the world public transport industry, and the development of multiple use smart card systems that include transit has extended the use of smart cards by public transport operators. Other two examples are the New York MetroCard and Gothenburg smart card. In 1993 the New York MetroCard for the underground system was launched based on contact technology in the initial stage. In 1993, the contactless smart cards application in Gothenburg could be viewed as the first implementation for the contactless technology in large-scale.

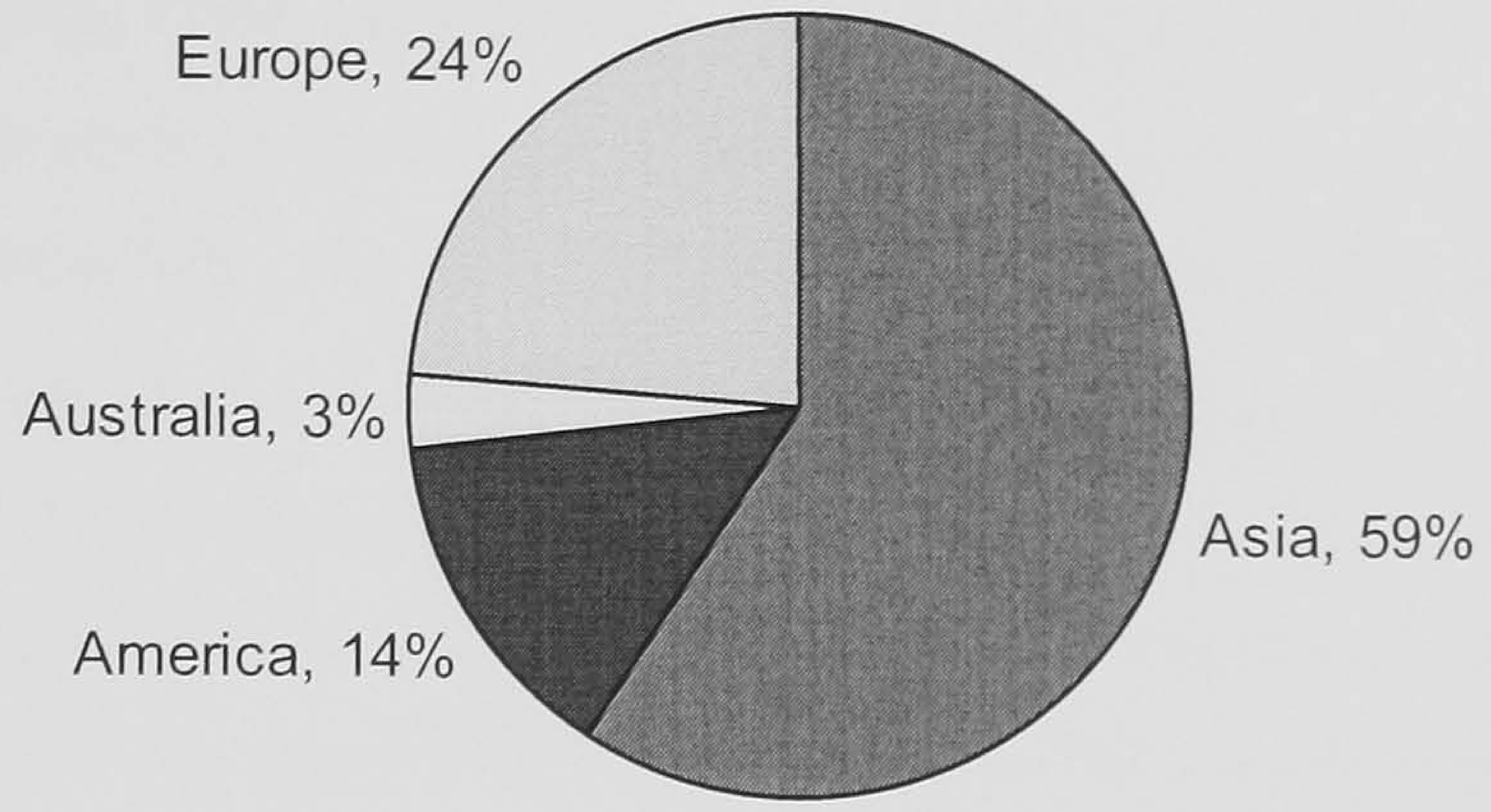
The contactless smart card technology is now becoming the mainstream in electronic fare payment for public transport in the worldwide, just as Laconte (1998) summarised, "*in 1990, the first UITP Conference on Automatic Fare Collection concluded that Smartcards were very promising but not yet mature. By 1994, contactless technology was recognised as the answer to improved passenger flows and integration in multi-modal payment schemes. Our last Bologna Conference in February 1996 reported on some important pilot schemes in Europe and the Far East*". There were several experimental schemes for public transport ticketing with smart cards in the 1990s. Some of them acted as the key drivers in raising the concerns on smart card applications throughout the world, such as the Octopus card in Hong Kong, which multiple functions (e.g. fare payment, shopping, telecommunication and parking fee payment, *etc*) of smart cards have been widely covered by one card, bringing users a great deal of convenience; the London Oyster card has a variety of fare package available for different card users, such as "pay as you go" cards, weekly cards, monthly cards, *etc*; and the MetroCard in New York, which has more convenient purchase/top-up options for card users, such as by vending machine, station booths, mails, *etc*, in addition to various ticket options of smart cards like the Hong Kong Octopus card (Chambers, 2002; Casey, 2000; Blythe, 2004; ITSO, 2006).

Figure 2.1 illustrates the proportions of contactless and dual interface smart cards delivery in the world in 2003. In total, there were 70 million contactless and dual interface smart cards used in transport services in 2003, with a majority deployed in Asian countries (59% of total number of smart cards used). These countries implementing smart cards in public transport, such as China, South Korea, Singapore, Japan, *etc.* are becoming the major users in the smart card market (APSCA, 2004). Following Asia, European countries were in second place with 24% of total number of smart cards.

Several reasons for boosting the smart card use in Asian market are:

- Potential huge market demand: public transport is still the primary transport mode in most Asian countries, especially those developing countries, like China and India. Therefore, a great number of PT users can definitely cause an increase of delivery of smart cards. In addition, it is also because of large population of Asia as well as high reliance of public transport;
- Even for those developed countries with limited land resources and high population density, like Japan and Singapore, governments also focus on

encouraging people to use public transport, so as to reduce the impact of private cars to the environment. So, enhancing PT services by using new fare payment method (e.g., smart card ticketing) is one solution to increase the attractiveness and competitiveness of public transport.



Source: eEuro Smart Cards, 2004

Figure 2.1 2003 Contactless & Dual Interface Smart Cards Delivery

Another study of smart cards by Business Communications Co, Inc. (1999) presented the general trends of smart card transaction volume in the world and U.S. specified. The worldwide transaction volume would increase from $5.3 billion in 1998 to $14.6 billion in 2003 by forecasting, with an average annual growth rate (AAGR) of 22.7% (see Table 2.2). The figures listed in the table involve all sectors of smart card applications in the U.S. and the world, respectively, including transportation, and show the great potential in the use of smart cards in the world. It can be seen that the AAGR in U.S. is much higher than the international and total AAGR. Although from Table 2.2 we cannot capture exact reasons (or influential factors) why so much higher growth in U.S. than other countries, the possible reasons would include, GDP growth, US$ exchange rate and multi-applications of smart cards (e.g. transport, banking and shopping), *etc*.

Table 2.2 Smart Card Transaction Volume, 1998-2003 ($million)

	1998	2003	Average Annual Growth Rate (1998-2003)
U.S.	58	2,015	103.3%
International	5,200	12,600	19.4%
Total	5,258	14,615	22.7%

Source: Franklin, 1999

2.3.2 EFP Applications in the World

Since the early 1990s when smart card ticketing was firstly introduced in America and Europe, large-scale smart card ticketing schemes have been widely implemented throughout the world. In this section, a detailed review of selected EFP (smart card) applications is presented.

America:

In this section, several American smart card projects are selected to illustrate current applications of smart cards and their features in U.S.

Washington: In May 1999, the Washington Metropolitan Area Transit Authority (WMATA) completed a 1-year study of the feasibility of a contactless smart card scheme (using Cubic's Go-Card, called 'SmarTrip') for use on rail and bus, as well as at park and ride lots.

The scheme installed the reader/writer units in 24 rail mezzanines, 21 buses (on 3 routes), one bus depot and five park and ride facilities. This scheme was based on the basic Go-Card technology originally developed and intensively tested by London Underground in 1990 and 1991 in the "Touch and Pass" programme. AVMs (Automated Vending Machines) can read and display the value remained on a Go-Card and top up the cards. On the bus, the maximum fare is deducted on entry by the "target reader". Passengers must check out on leaving, if a one- or two-zone ride is taken and the appropriate value is restored. The same concept is used to pay for parking. Data from rail, bus, and parking subsystems are transmitted via a modem to the WMATA Central Computer System to apportion revenue.

By the end of 2004, over 800,000 of the permanent, rechargeable plastic smart cards, which hold up to $200 in fare value, had been sold. One third of WMATA Metrorail riders use SmarTrip cards regularly (American Public Transportation Association, 2007).

New York: New York MetroCard was firstly implemented in the urban system in January 1994. Now the MetroCard is the current payment method for the underground and bus services (Barrett, 2003).

This smart card scheme had two features when it was launched in the 1990's:

- A variety of fare packages for different users;
- A range of top-up/purchase methods for smart cards.

These two features above were also considered by later smart card applications in other countries, which were proved the popularity of these two features among users, because various price schemes can enrich PT users options based on their own travel habits. The fare packages based on the MetroCard included single ride card, pay-per-ride card, unlimited ride MetroCard (from 1 day to 30 days), student Metrocard, and disable/senior citizen MetroCard. All these packages enrich people's options of smart cards in the market place. In recent years, various purchase options have been gradually added in the smart card application, bringing convenience to card users. The MetroCard has presented a good example on this: card users can purchase/top up MetroCards at ticket offices, vending machines, agencies, on board and by mail.

Southern California: In 1999, as part of Phase II of the Advanced Fare Payment Media Study, funded by FTA (Federal Transit Administration) and the California Department of Transportation, the Echelon Industries developed bus card read-write units and installed them on buses at three transit agencies in Southern California (Gardena, Torrance, and LA DOT). Echelon tested these units with contact cards on some buses and contactless cards on others, in order to evaluate the user acceptance and performance of the two types of cards. In Phase III, the read-write units are being used by seven transit operators in Ventura County, California. The test results have indicated that contactless smart cards had more advantages than contact cards (Business Wire, 2002).

San Francisco: This project involves the development of a regional integrated stored-value card system for transit operators in the San Francisco Bay Area. It was initially intended that the project would use magnetic tickets, similar to the existing BART (Bay Area Rapid Transit) ticket, and the original TransLink ticket was tested at BART and two bus systems (BART Express and Central Contra Costa County) in 1994 and 1995. The Metropolitan Transportation Commission (MTC), the lead agency, commissioned a study to determine the most appropriate technology. This study, completed in late 1995, recommended a contactless card system, and MTC commenced the development of the regional system in mid-1996 (Laezza, 2004).

Two transit agencies currently accept TransLink on all routes: AC Transit (including its subsidiary, Dumbarton Express) and Golden Gate Transit. Of the two transit agencies currently accepting TransLink, AC Transit riders account for about 45% of the riders using TransLink, Golden Gate Transit and Ferry accounts for a similar percentage, and a small number of cardholders ride the Muni Metro system. Use of TransLink on AC Transit has expanded dramatically since October 1, 2007, when AC Transit began offering TransLink cardholders discounted fares ($.25 off when paying with TransLink e-cash; $5 off when loading a local 31-day pass; and $10 off when loading a Transbay 31-day pass) (TransLink, 2007).

Europe:

The UK:

According to the report from Department for Transport, U.K. (MVA, 2003), 23 cities or regions have implemented or piloted smart cards in public transport systems, such as Transport for London (TfL), Southampton, Nottinghamshire, Edinburgh, *etc.* Among these, TfL would become the biggest user of smart cards in the future (1.1 million concessionary cards, 3 million commercial cards). We select a few of these cases as an example to study the applications of smart cards in the country.

London: The London Oyster smart card system, which was launched in 2003, is one of the most advanced, flexible and integrated ticketing systems of its kind in the world. By March 2007 over 10 million Oyster cards had been issued and more than 80% of all journeys on services (underground and bus) run by Transport for London used the Oyster card (Greater London Authority, 2007). Around 22% of all Tube journeys are Oyster Pay as you go, around 4% cash, the rest paid by travel cards (i.e., Oyster travel cards and paper based travel cards) (TfL , 2007).

To encourage people to use the system, in 2003 a discounted ticket of 70p instead of the £1 adult single cash fare was available for the Oyster Pre Pay fare on bus and Tramlink services (TfL, 2004). As of 22 October 2007 a cash bus or tram fare is £2, while the single Oyster fare is £0.90, but capped at £3 for any number of trips in a day. On the PAYG (pay as you go) rail network, a single trip within Zone 1 costs £1.50 (compared to £4 cash), or £1 (£3 cash) within any other single zone (TfL, 2007). Besides the Pre Pay tickets, passengers can also purchase other types of the Oyster cards such as 7-day, monthly and annual. Oyster cards are valid on Underground, Tramlink, DLR (Docklands Light Railway) and National Rail services within travellers' chosen zones and across the entire London bus network. To be more flexible, travellers can buy different tickets using a single Oyster card. For example, a passenger can buy a Monthly ticket for Zones 1 and 2 and Pre Pay ticket for Zone 3 if needed.

Passengers can purchase an Oyster card on-line and the card will be dispatched by mail. But this normally takes a couple of days in the UK. There is a £3 deposit for each card which is refundable when users return the cards, though the whole idea is that you continually top up the card from time to time.

With the increasing use of the Oyster cards in London, TfL (2007) addressed the impact of the smart card ticketing: "*Since the introduction of the Oyster card, the number of customers paying cash fares on buses has dropped dramatically. In addition, usage of station ticket offices has dropped, to the extent that in June 2007, TfL announced that a number of their ticket offices would close, with some others reducing their opening hours. TfL suggested that the staff would be 're-deployed' elsewhere on the network, including as train drivers*".

Milton Keynes, England: The city has used contact smart cards for riders since 1990 (Blythe, 2004). In 1996, approximately 40,000 passengers used smart cards on fixed-route service, and smart cards were also used on demand-responsive service (Fleishman *et al*. 1996). Passengers can pre-purchase a fixed number of rides at a small discount. The system also supports unlimited-ride passes that begin on the first use. The plan is that, eventually, passengers will be able to sign up to have their bank accounts debited, with value transferred to the smart card. A list of rides taken will appear on their monthly bank statements.

The introduction of a Smart Card which can be used for parking and bus journeys enhanced the public transport service quality. This gives more freedom of choice and demonstrates the differential between the cost of using the bus and parking. In due course, the card can also be used to pay for other services throughout the city, making Milton Keynes the leader in using technology to manage its services (Milton Keynes Council, 2007).

Bradford, England: The bus company, First Bradford, runs a smart card scheme (FirstCard) in and around the area in West Yorkshire, UK (Bagchi and White, 2004). The scheme is open to all users of First in Bradford buses (they operate approximately 250 vehicles). The smart cards can be used to purchase a range of bus tickets such as period bus passes and one-day travel cards. The bus passes are valid within an area known as the Rider boundary, surrounding Bradford. The card also has stored value, which can be used to purchase a range of bus tickets (including the passes mentioned). To Dec 2006, about 40,000 cardholders have been registered. However, there is no identification about the cardholder on the FirstCard, which means that it could potentially be used by anyone.

This scheme offers a system of 'Bus Miles' (with 1 bus mile being awarded for every £1 spent and 100 bus miles being converted into £1 of stored value - effectively a discount of 1%) whilst the Cheshire scheme offers 10% extra travel. The majority of bus passengers generally have limited amounts of disposable income and are, therefore, more interested in the fare that they actually pay. Sureline, therefore, decided to first offer a stored value ('electronic purse') smartcard. The first type of SmartRider card to be introduced, therefore, offers regular passengers a 10% discount on cash fares (with cash fares for most journeys already lower than the fares charged by the established operator). It was Sureline's view that a 10% discount on cash fares would be perceived as offering greater value than offering 10% extra travel. This discount applies to all passengers purchasing SmartRider cards – adults, children and concessionary passengers (Beaman, 2004).

Nottingham, England: The EasyRider card is a contactless smart card introduced in September 2000 in Nottingham, for use on Nottingham City Transport Services (Nottingham City Transport, 2007). It was originally named 'BusCard'. Like the London Oyster Card, the EasyRider card also offers a variety of card products for different users:

- EasyRider City: Green card can be purchased in blocks from 7 days up to 1 year. Activated for all buses and trams in that period of time.
- EasyRider Anytime: Easyrider Anytime card can be purchased for 2, 5, 10 or 20 days. However, it is only activated on the days it is used, and is therefore useful for those who travel regularly, but not on consecutive days.

- EasyRider Farecard: Like an electronic purse, ordinary tickets can be bought at a reduced price using credit stored on it.
- EasyRider <16: EasyRider City equivalent for under 16's.
- Student Cards: Students of the Nottingham Trent University can also use their university ID cards on Nottingham City Transport and NET, these are activated in the university at discounted rates. They can also be topped up at the NCT Travel Centre in the usual way at full price.

In February 2006, Trent Barton, a competing bus operator in Greater Nottingham announced a competing contactless smart card system, ToTo (Touch-on Touch-off). While the ToTo system will be more advanced, the Easyrider card is already established and has the backing of the local tram service (Nottingham Express Transit).

Ireland:

In March 2005 the Luas smart card was launched in Dublin. This allows travellers to pay for travel on the urban light railway system in Dublin. A smart card can be purchased at ticket agents or online. The card costs €10, which includes a €3 non-refundable charge for the card, €3 of credit and €4 for a fully refundable 'reserve fund' which allows you to travel even if there is insufficient credit on the card for the journey. However, the card must then be topped up before another journey can be taken. This function is similar to the overdraft function in some other smart card applications, such as Beijing public transport smart cards, the last trip can be guaranteed when the credit in the cards is not sufficient to pay fare. However, compared with other smart card applications, the fare package of the Luas smart card is simple. The current application is only for paying per ride policy, which is based on flat fare. Longer term smart cards, such as weekly cards, monthly cards, cannot be found in this project (Dublin Transport Office, 2006).

Sweden:

Gothenburg presented a good example about the combination of contactless smart cards and magnetic fare. 1993 was the first year for the contactless card application in Gothenburg city. And at the early stage of the smart card application, the project combined the magnetic cards, which had been used in the current fare payment system, and contactless smart cards together. The customer buys a magnetic card, what we call a "loading card" in any ticket outlet. On the bus, the passenger inserts the magnetic card into the card reading device while holding the contactless card near the reading device. Now he/she has added a new validity period into the contactless card. He/she gets the magnetic card back, and the magnetic card is then written by the validity area and the expiry date. This is a self-service operation, performed entirely by the customer. The driver does not normally get involved. Advantages for the customer of this combination: Reloading is fast and normally takes place once a month for the passenger. It is easy for the passenger to do the reloading. The passenger is never in doubt of the validity of the contactless card, since it is written on the magnetic card. Also the contactless card serial number is written on the magnetic card. And of course the cards may remain in the wallet, bag or purse when validating. Magnetic loading cards can be sold anywhere, without any technical equipment. But the main disadvantage is that the customer must hold two cards, which could be inconvenient to users (UITP, 2005).

Other smart card applications in Europe can be found in Germany, Holland, France, *etc*. For example, in Holland, a smart card scheme is proposed on a nationwide level so that the smart card application can provide seamless journey for all card users by integrating different PT fare systems across Holland (Cheung, 2003). A similar application can be found in the Rhein-Ruhr area, Germany. Travellers can use one smart card to travel in different cities and towns within this area rather than a single city. In Paris, with a wide use of the smart card named "Navigo Card", it is planned to replace the magnetic stripe cards in all travel zones shortly. The mutual features of these smart card applications include: the various fare packages combined in smart cards (for different users, different time periods), a combination of contact and contactless technologies, which can be used different ticket check-in systems (contactless card reading devices or physical contact check-in system). Extra services added in smart cards is another new direction in recent years, such as shopping, parking fee payment and telecommunication, *etc*. The details of smart card applications in the world are presented in Appendix A.

Asia and Pacific Area:

Hong Kong: On the 1st September 1997, the Hong Kong contactless smart card system called Octopus went live, following 3-year development and testing. As it provides a sort of novel fare payment method on a variety of PT services (e.g. rail, bus, tram and ferry), the Octopus card was already adopted by almost 50% of Hong Kong's population carrying out 2.5 million transactions per day in 1997. The contactless smart cards used in the Octopus system in Hong Kong have a life of about ten years. Figure 2.2 shows the rapid growth of smart card ownership and usage in year 1997, 1999, 2000 and 2004. As can be seen, there is a dramatic increase of the use of Octopus cards compared with previous periods of time. Particularly by Dec 1997, the first year for the smart card application, the number of cards reached 3 million, indicating the huge potential in the market. Comparing with fast changes in the number of card holders, the daily transaction gradually increased from 1997 to 2004. According to Octopus Cards Limited, operator of the Octopus card system, there are more than 14 million cards in circulation, twice the population of Hong Kong. The cards are used by 95 percent of the population of Hong Kong aged 16 to 65, generating over 10 million daily transactions worth a total of about HK$29 billion (US$3.7 billion) a year (Octopus Cards Limited, 2007).

A deposit of HK$50 for each smart card partially compensates the acquiring cost of the smart card from the card manufacturer. After implementing the Octopus system, fare schemes are now more flexible and sophisticated. Fare schedule is no longer limited to a simple zone-based or distance-based schedule, e.g. tourist Octopus cards. Two versions of this card are offered, a HK$220 card with a free single ride on the Airport Express, the Mass Transit Railway (MTR) train line that runs between the Hong Kong International Airport and the urban areas of Hong Kong, and a HK$300 card with two free single rides included. The airport journeys are valid for 180 days from the date of purchase. Both versions allow three days of unlimited rides on the MTR and include a HK$50 refundable deposit. Usable value on these cards may be added if necessary. These tourist Octopus cards may be used only by tourists staying in Hong Kong for 14 or fewer days; users may be required to produce a passport showing their arrival date in Hong Kong. *Airport Express Tourist Octopus* is available for purchase at all MTR stations (Octopus Cards Limited, 2007).

Another example is the introduction of inter-modal and intra-modal (i.e. discounts between services of different companies) discount schemes. By offering discount for journeys of some specific patterns, promotion of certain 'transportation mix' can be achieved. For example, in order to enlarge the passenger catchment areas of some MTR stations, MTR partnered bus and mini-bus operators to offer discount to some specific 'feeder' bus or mini-bus trips taken after an appropriate MTR ride, or vice versa. These examples illustrate the versatility offered by smart cards, which cannot be provided by the magnetic cards (Chambers, 1998, 2001).

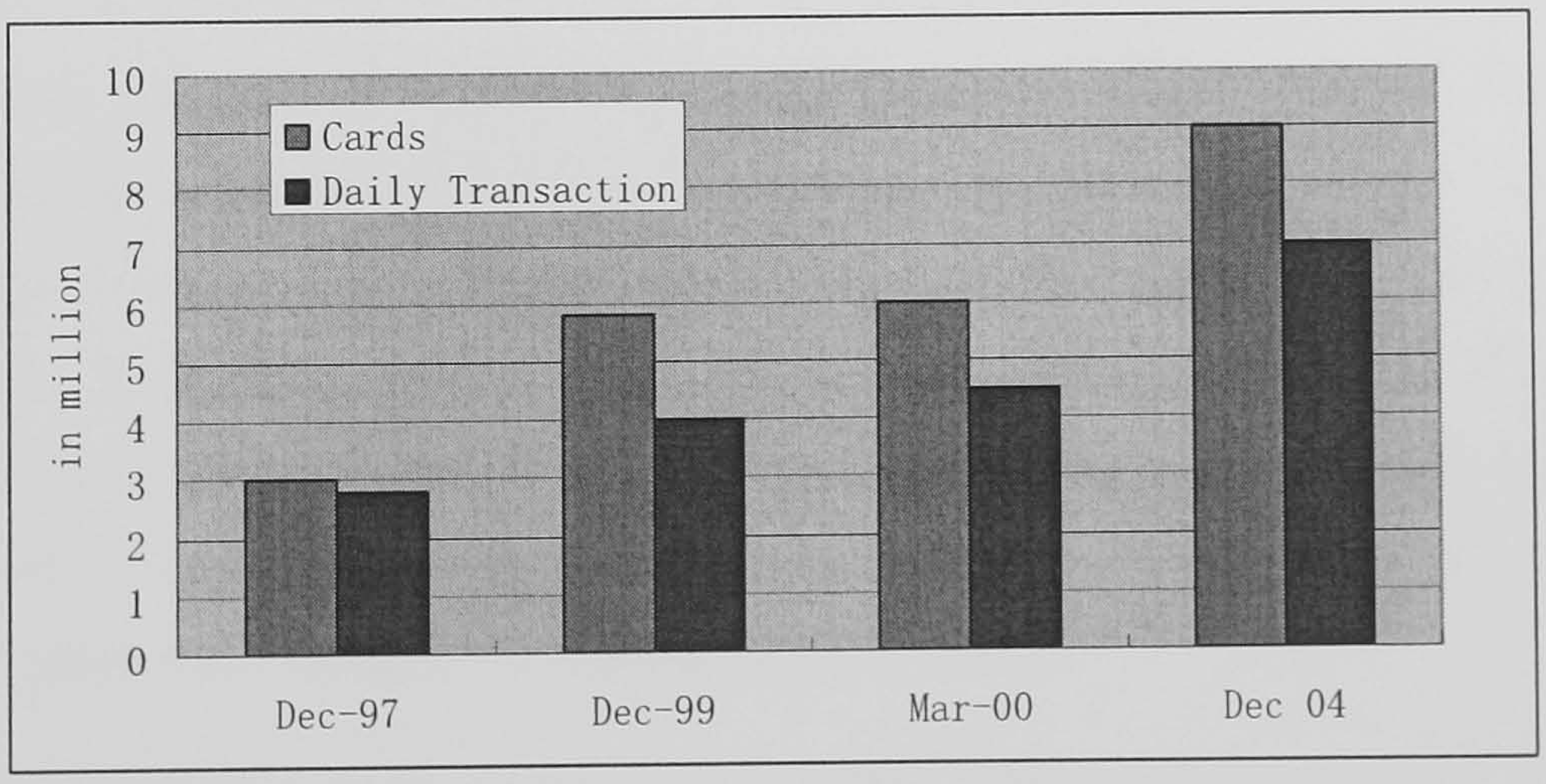


Figure 2.2 Smart Card Growth in Hong Kong

Source: McDonald, 2000, Smart card Alliance, 2005

The fare structure under the Octopus card system is: The MTR charge less for journeys made using an Octopus card instead of conventional single-journey tickets. For example, the adult fare of a single journey from Chai Wan to Tung Chung is HK$23.10 with an Octopus card, and HK$26 with a single journey ticket (MRT report, 2007).Other public transport operators also offer discounts, usually 10 percent, for using Octopus cards on higher fares and round-trip transits (New World First Bus, 2007).

In recent years, usage of the Octopus card was extended to the Chinese cities of Shenzhen and Macau. In collaboration with China UnionPay, Octopus Cards Limited introduced Octopus card usage to two Fairwood restaurants in Shenzhen in August 2006 (Octopus Holdings Limited, 2008). In 2008, five Café de Coral locations in Shenzhen also started accepting Octopus.

Taiwan: the EasyCard is a contactless smart card system operated by Taipei Smart Card Corporation for use on the Taipei Rapid Transit System and on buses and other public transport services in Taipei since June 2002. It is explicitly modelled after Octopus cards in Hong Kong. Like many electronic fare systems, the card employs RFID technology to operate without contact. In addition to the EasyCard being used on the Taipei Metro and buses, the card is also accepted at public car parks adjacent to Metro stations and in other areas of Taipei. The stored value on the card can be replenished at convenience stores around Taipei, such as 7-11. The simplicity of usage and availability has made EasyCard a household name, being used by most commuters in Taipei. The discount policy also is applied in the EasyCard with 20% off for single fare ticket. The recent development of the EasyCard is to combine the fare payment function of student EasyCards with student ID cards together so that another feature of smart cards, identification, can be taken to full advantage (Wang et al, 2003).

Japan: the Suica card project was launched in Nov 2001. Compared with smart card applications in other countries, the biggest difference of the Suica card is that it can cover a wide area by integrating different fare payment systems in different cities, including Kanto area (including seven counties and the Greater Tokyo). The Suica card also uses contactless technology to collect fare from users. As most smart card applications, 500yen (2.5 pounds equivalent) deposit is required when users buy smart cards. Meanwhile, various purchase/top-up options (e.g., automatic adding value machine, topping up on-line) and multifunction (e.g. shopping and parking fee payment) in the Suica card are other two key features, which have brought convenience to card users.

As of April 2007, over 20 million Suica cards were in circulation (East Japan Railway, 2007).

Seoul, South Korea: From July 1, 2004, a new type of transit card called "T-Money" has been used for public transport on all subways within Seoul metropolitan area and on buses (both intra-city and village) that serve routes in Seoul and local districts in Gyeonggi Province and Incheon.

The "T" in T-Money stands for "top, touch, total, travel, and technology" making it much more than just a fare card, because the cards can also be used for multi-purposes, such as banking, shopping and parking, *etc*. Passengers should have their T-Money cards touch a card reader each time when they embark and disembark from a bus or a subway to get the discount benefits. A passenger who does not intend to transfer to another mode of transit only needs to have the smart card checked by the machine when getting on the bus. After leaving the bus or subway, a passenger must transfer within 30 minutes (except from 21:00 to 07:00, when the transfer period is within 1 hour) to get the transfer discount. Various adding value options, including credit cards, mobile phone, and add-value machine, *etc*., for the smart cards, bring more convenience for the card users (RC, South Korea, 2004; Kim, 2006).

Melbourne, Australia: Melbourne's Public Transport Corporation (PTC) has awarded a contract called "Onelink" to develop, implement, operate, and manage a smart card- and magnetic-ticket based system for the region's network of trams and light rail vehicles, buses, metropolitan trains, and over 1,000 private buses. A major decision by the government was to outsource the entire ticket and allocate revenues for operators. All smart cards are contactless and have a magnetic stripe. The availability of smart cards for the full range of ticket types, i.e., monthly, weekly, multi-ride, have been progressively introduced over an extended period, the rate of introduction being related to customer demand. The new fare system was introduced in two stages, which began in the late 1994 and 2003 respectively. At the beginning stage, riders who were provided with smart cards included disabled persons, long-term riders who normally buy yearly passes and school children (Goulcher and Ashmore, 2004).

Singapore: The EZ-Link card is a contactless smart card based on Sony's FeliCa smart card technology, used for payments in Singapore especially for transportation. Established in 2001, it was promoted as a means for faster travel due to speedier boarding times on buses. As of 2005, there were over 8 million EZ-Link cards in circulation, with 4 million card-based transactions occurring daily. The card is commonly used in Singapore to pay public transport fare, including the city-state's Mass Rapid Transit (MRT), Light Rapid Transit (LRT) and public bus services. The card also serves as a supplementary identification and concession card for students in nationally recognised educational institutes, personnel serving in the Singapore Armed Forces, or senior citizens who are over sixty years old.

The system has since been expanded, with EZ-Link cards being used for payments in Singapore branches of McDonald's, food centres, supermarkets and libraries, and even soft drink purchases from vending machines. Some schools in Singapore have also started to adopt the EZ-Link card as a way to mark the attendance of students and to pay for food served within the school campus. The same system is used by the Octopus card in Hong Kong and will be used for public transit in the Netherlands. On December 3, 2005, EZ-Link Pte Ltd announced that it was working with NETS to create a new hybrid card which will have the functions of both the EZ-Link card and the CashCard. This card would make it possible for one card to be used for payment on three popular modes of land transport in Singapore — ERP (electronic road pricing), bus and MRT. Work on this card is expected to be completed in 2007 (Segaran and Sim, 2004).

Summary of Current Experiences:

To sum up, from the current smart cards applications, we can conclude the following points related to the main success of smart cards, common features and implications for this research:

- The success of smart cards for public transport is that smart cards have become one of major payment methods for public transport users, in addition to cash and travel cards. More and more users would like to accept and use this kind of new payment method. In some China's cities, more than 10 million cards have been issued, such as Shanghai, Guangzhou. The market share of smart card use in public transport systems is presenting an increasing trend in recent years.
- Influential factors (features) of smart cards implied from these applications include:
  - Contactless technology has been widely used in most smart card projects;
  - Based on the smart card ticketing, PT operators can offer a variety of fare packages to different card users, such as student cards, senior/disable citizen cards, weekly cards, *etc*; Secondly, in order to increase attractiveness and encourage people to use smart cards, discounted fare policy is applied in all smart card applications;
  - Deposit policy for initial purchase of smart cards is applied to guarantee good physical condition of card itself when card users cancel and return their cards;
  - Multifunction: the role of smart cards is no longer just for public transport fare payment. Extra services, for instance parking fee payment, shopping, telecommunication, have been added into smart cards in recent years:
  - Various purchasing/adding value options: more convenient options for purchasing/adding value of smart cards can increase accessibility to the smart card use for the vast majority.

The detailed advantages of smart cards to operators, users and authorities are discussed in the next section.

- Finally, implications for this research can be summarised as: faced with increasing demand of smart card use, passengers' choice behaviour and changes of market share, particularly toward different payment options (e.g., smart cards vs. traditional payment methods) need to be identified, and it is necessary to determine the importance of features (or attributes) of smart cards to identify benefits of smart cards. According to the previous applications of smart cards for public transport, features of smart cards which have been widely used throughout the world can be summarised in Table 2.3. Attributes listed in Table 2.3 are proposed to be used in the following survey design and evaluation study.

Table 2.3 Summary of Features of Smart Cards for Public Transport

Features	Countries/Cities	Details of the application
1. Multifunction	Octopus card (Hong Kong); "T-Money" card in Seoul Suica card in Japan	Not only for public transport fare payment, but also for shopping (small value consumption), parking fee payment; tolling fee payment; ID card, etc.
2. Wider geographic areas covered	"T-Money" card (Seoul) Rhein-Ruhr smart card (Germany)	Interoperation between some neighbouring cities by a single card.
3. Top-up/adding value methods	Octopus card (Hong Kong) Oyster card (London) Suica card (Japan)	Online; automatic adding value machine; mobile phone; direct debit, etc.
4. Various options of smart card products	Octopus card (Hong Kong) MetroCard (New York) Oyster card (London)	Weekly cards, monthly cards; annual cards; student cards; the elderly cards; tourist cards, etc.
5. Discounted fare	All smart card scheme, such as Oyster card (London) MetroCard (New York)	Normally 10-40% off for single ticket Or free trip provided (Hong Kong Octopus card)
6. Overdraft	Luas smart card (Dublin) "Yi Ka Tong" card (Beijing)	The last trip can be guaranteed in case the credit of the smart card is not sufficient to a single fare.
7. Deposit of initial purchase of smart cards	All PT smart card scheme	When users purchase smart cards at ticket offices, a mount of deposit is required (such as £3 for London Oyster card, 500yen for Suica card in Japan; 50yuan for Dalian smart cards in China)
8. Personal ID	Singapore EZ-Link card Go-Card, USA	PT smart cards also act as personal identification where required, because personal information also are chipped into the cards.

Looking to the Future:

Except features of smart cards discussed above, the future direction for smart card development will be focusing on the following aspects:

- Interoperation among different public transport fare systems in much wider geographic areas. The Suica cards in Japan, Holland smart cards, the Rhein-Ruhr smart cards in Germany, have presented a good example on this point, because all of these cards integrate different PT operators and fare systems in different geographic areas, so that they can offer convenience to travellers.
- Multi-function, such as in banking, shopping, vending, *etc*., has become the future development direction in smart card applications. Recently, the use of smart cards has grown in the banking sector where a number of "electronic purse" applications have been launched, in particular for the payment of small amounts.
- The level of security of smart cards should be improved so that it is applicable to any low cash value/high volume applications such as public telephones, parking, vending, *etc*.

2.4. Advantages and Disadvantages of EFP

With the increasing applications of smart cards throughout the world, advantages of the smart card ticketing have been gradually realised by its users. Advantages of EFP applications can be summarised from following three aspects:

- to public transport operators;
- to public transport users;
- to other sectors of the society.

As an old saying: 'Every coin has its two sides', without exception, currently EFP applications also have some disadvantages to operators, to users, to the society. For this point, in order to enhance the service quality and increase the accessibility of EFP applications, disadvantages of smart cards are discussed to give direction for improving smart card applications in the near future. This section primarily focuses on qualitative summary of pros and cons from previous applications. Another role of this section is to have a general context for the evaluation analysis of this research (e.g., which aspect of smart card applications could be a focus to evaluate, etc.).

2.4.1 Advantages to Public Transport Operators

Implementation of smart card ticketing in public transport system is not only for the sake of technology. It offers benefits and opportunities to transport operators that other traditional technologies fail to offer. Advantages of EFP to public transport operators can be summarised as follows:

Cost Saving:

The cost saving of smart cards can be discussed from two aspects: Operational/maintenance cost and PT personnel cost, by comparing with those traditional payment methods (cash, travel cards, even magnetic payment cards). For those cities where magnetic payment cards had been widely used when smart card ticketing was introduced, the wide acceptance of smart cards in the urban PT systems reduces the usage of the magnetic ticket transports in check-in gates. For instance in Hong Kong, over two-thirds of the magnetic ticket transports originally installed can be removed from the PT service after the smart card application. This reduces not only the number of magnetic ticket transports for maintenance but also the maintenance frequency per mechanical transport (Chan, 2002).

Compared with cash fare payment, smart cards can reduce the PT personnel cost and improve the staff utilisation. For example, in Beijing, China, under cash payment systems, normally conductors are required to issue and check the validity of ticket on board. After the smart card ticketing was introduced since May 2006, the number of on board personnel was reduced: the number of conductors was reduced from about 10,000 to 5500 (Beijing Public Transport Company, 2006). Instead of conductors, PT drivers and smart card reading devices on board take over the same responsibility as conductors did before.

Thirdly, under smart card applications, higher fare revenues from increased ridership could be realised. In surveys of reported (or intended, in a new system) use of stored-value fare payment media, passengers have indicated the likelihood of making some additional trips because of the convenience of having smart cards (Andrle, 1997; Wang et al, 2003). For example, in a survey in Chicago regarding intended use of the new stored-value smart cards, respondents indicated that they expected to increase their trip-making on local PT system after purchasing the cards; analysis of the results produced an estimate that the smart cards can be expected to induce 2% to 5% increase in trips among these passengers (Chicago Transport Authority, 1995). Also, in Taipei, the survey data indicated that after using smart cards, the average number of passengers' daily trips increased about 29% (Wang et al, 2002).

Better Information Management:

Smart card ticketing systems can lead to improved management information on customers and their journeys which will help in service planning, in managing the operation, for revenue apportionment between routes or operators, and for revenue protection staff who will better be able to control fare fraud.

The inherent flexibility of the smart card, and its great data capacity, will make it possible for operators to introduce new ticketing products in order to skilfully develop their pricing strategy. Smart cards give the ability to implement complex pricing systems if required, which can distinguish between different categories of customer and be used to target specific markets, whilst making the operation of the system on bus or in the station very simple. Integrated ticketing between operators will also be easier to implement, and allow one card to be used on the services of a number of operators, each of whom have their own pricing regime. Customers would be able to use one card on the services of a number of transport operators. This will help public transport to attract new customers, keep existing customers, and encourage them all to make more journeys (Bagchi and White, 2004).

Another advantage on information management is to monitor passengers' travel behaviours on individual level. Through the central data base of the smart card ticketing, the system can record which stop a passenger gets on and where he/she gets off by tracking the smart card ID number. Comparing traditional travel behaviour survey methods (e.g. on board surveys), this is an easy way not only to save survey cost, but also secure the data quality. The key use of smart card data is to conduct demand analysis to see whether passengers' response changed in fares or system characteristics.

Enhanced Operational Efficiency and Performance:

Apart from cost saving, the introduction of EFP also enhances the operational efficiency. For example, the wide acceptance of stored value in EFP systems has reduced coin flow volume attributable to single journey ticket transaction (Chan, 2002). Together with the electronic smart card top-up mechanism such as auto pay (automatic top-up in a transaction from the user's bank account) and top-up with a bank debit card in add value machines in some applications (such as Hong Kong, New York, Japan, *etc*), the expensive cash handling processes are now operated in a much smaller scale (Chan, 2002). Fare accounting is also simplified and is more reliable with the establishment of Central Clearing House System of the smart card ticketing which transfers funds among service providers total automation. Another aspect of operational efficiency is the enhanced gate throughput rate with the use of EFP, for example in underground systems in London and New York.

Compared with magnetic card payment systems, which have been implemented for dozens of years, successful experience on smart card ticketing indicates that the reliability of smart card equipment is much higher than the magnetic alternative. Because of the lack of mechanical parts to process magnetic tickets, smart card equipment seconds a lower fault figure and then maintenance cost can be saved. For example, in Hong Kong and Singapore in which magnetic cards were used prior to the smart card ticketing, it is possible to be able to achieve over 98% availability for smart card machines (such as card reading devices, add value machines, enquiry processors, and gates) during the peak hours. Resources previously spent on maintenance can be better utilised to directly serve PT passengers (Chan, 2002; Chambers, 2001; Segaran and Sim, 2004).

The third aspect about enhanced operational performance is on PT dwelling time saving. As we know, as the number of boarding passengers increases, vehicle dwelling time is expected to increase at a bus stop. In addition, incidents such as jammed fare boxes, handicap boarding, and the bus waiting for passengers running from another bus can also increase vehicle dwelling time significantly. Due to faster boarding time than cash and magnetic cards, Chira-Chavala and Coifman (1996) found boarding times of individual smart card users can lead to reductions of the time that the bus spends at bus stops (called vehicle dwelling time, or vehicle stop time) by up to 60 percent than cash fare payment. They also suggested that under a mixed fare payment system (cash, travel cards and smart cards), the bus dwelling time saving depended on the number of smart card users in passengers. Reduced vehicle dwelling time helps to improve the system efficiency by reducing bus travel time and enhancing bus schedule adherence.

More Flexible Fare Schemes:

Fare schemes can now be more flexible and sophisticated to attract different types of users to use public transport systems under the smart card ticketing. The fare schedule is no longer limited to a simple zone-based or distance-based schedule. An example is the Tourist Octopus Card introduced by MTR in Hong Kong, which is tailor made to meet the needs of the tourists in Hong Kong (Chambers, 2001). Tourist Octopus Cardholders are entitled to have two Airport Express journeys, enjoy three-day unlimited MTR travel and with the same token. Used on other transportation at an initial HK$20 stored value. The versatility of the Octopus card imposes no limit to any possible fare schemes. Another example is the introduction of inter-modal discount. By offering discount for journeys of some specific patterns, promotion of certain 'transportation mix' can be achieved. For example, New York MetroCards can offer free interchange for smart card users (Barrett, 2003).

On the other hand, such flexible fare scheme (price differentiation) can also bring the possibility to increase revenue for operators. For example, in London, the Oyster card separates use of time of smart cards by peak and off-peak period. For peak time users, they are required to pay higher travel costs. In Hong Kong, for frequent PT users, they could choose smart cards with unlimited ridership by paying higher initial fee than normal cards, but for occasional users, they would like to use a "pay as you go" card, which means they need to pay for each single trip but without expiry date for the credit in the card (Chan, 2002).

Reduced Fare Fraud:

Due to the easiness of duplicating travel cards (particularly for those paper-based travel cards), fare fraud and abuse has become a thorny problem for public transport operators. Also, a similar problem always happens on cash fare payment. Meanwhile, for fare box to collect cash, sometimes an insufficient amount of fare payment is also a problem. However, because of their enhanced security characteristics, smart cards are expected to reduce the potential for abuse or fraud and evasion, for example, it is more difficult to duplicate a smart card than a paper-based travel card because of electronic chip bedded in the smart card. In the survey of public transport operators by Andrle (1997), the average amount of revenue reported lost through "theft, fraud and counterfeiting" was approximately 1% for all respondents, or an average of roughly $1 million per year; this amount was significantly higher for the larger PT systems, an average of approximately $1.8 million, or 1.6%, for the heavy rail and commuter rail systems. Counterfeiting of smart cards has not been found to be a significant problem in the transit industry; because advances in protection technology have made smart cards increasingly difficult to duplicate. Another example in the current applications is that after the MetroCard was implemented in New York, the ratio evaders to fare payers dramatically reduced from 7.5% in July 1990 to 0.5% in November 1999 (Savage, 2001).

Increased Subsidies:

Another advantage to both operators and PT users is that smart card applications can produce an increase in subsidy due to better level of service. The aim of local governments is to encourage people to use public transport, to reduce congestion and the impact of private transport to the environment. Smart card ticketing has presented a good application to improve the service quality of public transport and attract more people to use public transport. So, subsidies to operators for promoting and boosting smart card applications have been under way in some cities, particularly in China, such as Beijing, Shanghai, Shenzhen, *etc*. These subsidies are mainly for two aims: the smart card implementation and fare differentiation. In 2004 Shenzhen local government funded the local PT operators 10.5 million yuan (667,000 GBP equivalents) on the development of smart cards and new fare structures based on the smart card ticketing (ITS China, 2004). For the fare differentiation under the smart card ticketing, subsidies can be used on different card products for social welfare purpose, such as the elder cards, student's cards, *etc*.

Reduced Workload for Public Transport Staff:

The use of smart cards could have significant influences on the performance and workload of PT drivers. The following advantages can reduce the workload of PT drivers (ATPA, 2007).

- Relative to cash fare and travel cards, smart card system can make PT drivers' jobs easier. The drivers do not have to count or pay attention to the amount of cash paid and the validity of travel cards used.

- Jammed coins and stuck cash bills, common problems of fare boxes on board, can be stressful to both bus drivers and passengers. Under smart card systems or mixed payment systems, PT drivers do not need to deal with these problems as frequently as before.
- The use of smart cards results in drivers having less interaction with passengers, which in turn has made their job less stressful. These interactions include: fare disputes; passengers paying insufficient fares; assisting passengers to insert cash into farebox; collecting "zone checks" (if applicable, like London public transport system).

2.4.2 Advantages to Public Transport Users

Boarding Time Saving:

Improved ticketing method like smart cards reduces boarding time for individual passengers as transaction is quicker than conventional payment means. For example, boarding time per passenger decreased from 7 seconds to 3.2 seconds on average when smart cards were introduced in California area of US in the late 1990's (Haworth et al, 1995, Chira-Chavala and Cofiman, 1996). Also the whole dwelling time and waiting time of public transport vehicles depends on individual boarding time saving. The latest PT user survey in Taipei showed that the average boarding time per passenger under the smart card ticketing increased by about 5 seconds, from 8.2 seconds of the non-smart card payment system to 3.4 seconds of the smart card payment system (Wang et al, 2003). The more cashless payment methods are used, the more the individual journey duration could be saved. Therefore, average boarding time in a payment system just using smart cards is lower than in a mixed system (e.g. smart cards and cash together). For those 'gate check-in systems' (e.g. underground systems), smart cards also provide easier and quicker entry to stations, trams, underground services for users.

Travel Cost Saving:

Due to lower operational and maintenance cost of smart cards than other fare payment methods, most smart card applications have a discounted fare policy to card users to encourage passengers to use the smart card ticketing. For example, compared with cash fare standard ticket, a 10%-20% off policy is applied in Dalian smart card application, China.

Secondly, a variety of fare packages based on smart card applications also can save travel costs for card users, such as in Hong Kong, short period smart cards are sold specially for visitors (one day, three days cards). Also the London Oyster card has more choices for different users, such as one-day card, weekly cards, and monthly cards (TfL, 2007). The presence of fare differentiation can make passengers choose the most economical fare products for their travelling.

Convenient Purchase of Tickets:

Smart card ticketing is a form to pre-pay fare payment, so card users do not need to pay for each single journey. Moreover, card users avoid handling exact coins/cash to buy tickets, like cash fare payment. Meanwhile, multiple top-up options also bring convenience to passengers, such as automatic adding value machine at stations/but stops, online, telephone, SMS, *etc*.

Integrated Ticketing:

Smart card ticketing provides an opportunity for users to have easier transfers and seamless journeys. Integrated ticketing, which can be used over wider geographic areas and multiple service providers, helps reduce stress in transfers between different services and modes (Gerland, 1996; Blythe, 2000, 2004).

Multifunction:

Smart cards can also produce positive response from users to multiple uses, functioning as an electronic purse, telecommunication, and even personal identification. For example, the Octopus cards in Hong Kong and "T-money" card in Korea can be used to pay fare in different modes and buy goods in stores (Chambers, 2001; Nelson et al, 2001) and ease in transaction where direct debit to bank accounts or deduction of a pre paid amount is carried out (Orin et al, 1997; TfL, 2007).

Also, extra services based on public transport fare payment in smart cards can trigger the demand growth of smart cards and public transport use. Users' attitudes surveys in Hong Kong and Taiwan for the smart card ticketing indicate that multifunction are most welcomed by users and has become one main reason to make passengers use smart cards (Painter and Law, 2003; Wang, 2003).

Protection against loss and theft:

Because smart card passengers do not need to handle cash to buy ticket at stations/on board, the chances of theft in these public areas can also be reduced (Seagram and Sims, 2004; To rode, 1998; TfL, 2007). Furthermore, the prompt reaction of smart card control system can provide sufficient protection against card lost and stolen.

2.4.3 Advantages to Other Sectors of the Society

Integration between Different Transport Modes

The smart card ticketing can not only achieve interoperation between different public transport modes, such as bus, tram, light railway, underground, *etc*, but can also integrate between public transport and private transport modes in society. A good application on this aspect is that the smart card can be used to pay parking fee and motorway tolling for private transport users (Chambers, 2001).

Cross Functionality with the Rest of Social Services

One smart card, which is combined with other function crossing different social service sectors, such as finance, communication, personal identification, *etc*. not only increases convenience for individuals, but also enhances the integrity and cooperation of the whole society (TfL, 2007).

Public Security

As a kind of cashless fare payment method, smart card ticketing can reduce theft at stations, bus stops and on board due to less cash on hand not only for smart card users, but also for other passengers (McDonald, 2000).

Influence to the Environment

Smart card ticketing is one solution to improve the attractiveness and use of public transport. Reduced car usage can relieve traffic congestion and reduce the pollution of private car emission. Moreover, the utilisation of the social resources can be optimised, such as energy, fuel (McDonald, 2000).

2.4.4 Disadvantages of EFP

To Operators:

To operators, the disadvantages of EFP are:

- Relatively higher initial investment cost on the hardware of EFP, including the central control system, card reading device and linkage with the clearing department (i.e., banks) than conventional payment methods.
- Different business interests of operators versus common database and revenue distribution, resulting in the deficiency due to low interoperability among different PT operators;
- Data protection between competitors especially in a deregulated market (one of the potential solutions is the strict legal and organisational framework – contract agreement on the use of the data).

To Users and Society:

Although the advantage of smart cards is that it can store users' information and provide sufficient data protection to card users, in recent years, some smart card system are criticised as a threat to the privacy of its users. For instance, personal data of London Oyster card users is stored both on the card and centrally by Transport for London; recent usage can be checked by holders at some ticket machines. Privacy groups consider it a form of mass surveillance and are concerned with how this data will be used (TfL, 2005). Although privacy is an issue for smart cards, the protection to card users when cards are stolen or lost is still higher than magnetic strip cards (most smart card control centre can give prompt reaction to stop the use of smart cards within 24 hours after users report).

Social exclusion would be a potential problem due to the implementation of smart card ticketing. Though an expectation of smart card applications is to increase equity of the whole society, and encourage every member in the society to use public transport services, certain groups of people (e.g., the elderly, tourists, low-income passengers) still thought the smart card ticketing is too far from them because they worried whether they could smoothly use smart cards, such a hi-tech application.

There are also consequences for a user in a multiple fare payment media system (e.g., cash and smart cards). If discounted fare is used as inducements in the smart card programme, anyone who does not opt to purchase the card could potentially be paying a higher price than smart card users. This could become an important equity issue if those that do not purchase the smart cards belong disproportionately to any one racial, ethnic, age or gender group. For example in New York City, studies have shown that women and Latinos are somewhat less likely to use smart cards (Chira-Chavala and Coifman, 1996). Situations like this may mean that PT operators need to monitor what part of passenger population is adopting the card. If certain groups are under-represented and the PT operators care that they have equal access, the PT operators must ascertain why the group is not adopting the card. Is it because they do not know about it, because they cannot afford it or because they do not want it? Each answer has a different remedy from improved marketing to subsidy.

Although multiple options for fare payment (cash, travel cards and smart cards) are applied in most cities, with the development of smart cards, some conventional payment method has been gradually replaced by smart cards through governmental enforcement. It implies that some passengers' payment habits and choices are forced to change. Therefore, another issue for public transport users is the enforcement of smart card use in some cities, for example in Beijing, since May 2006 PT users can only choose between cash and smart cards (ITS China, 2006). The travel card payment has been thoroughly terminated. Unlike travel cards with unlimited number of rides, former travel card users need to pay a given total number of rides for one month use under the smart card ticketing (e.g., 140 rides per month). For excessive trips, users have to pay by standard fare. So, for some users, their travel cost in one month would become higher than before.

To Local Authorities:

As one of advantages that the smart card ticketing can bring to the local authorities is the integration between public transport service and other services, even between different cities if necessary (TfL, 2007). However, in order to achieve such integration, a great deal of work need to be done, because the integrated system is totally new and cannot be referred from previous experiences, such as negotiation between different payment systems of different operators, integrating financial clearing centre for different service sectors. Therefore, for local authorities, the implementation duration and investment cost are the two major issues needing to be planned beforehand.

2.5 Enhancing Public Transport in China

2.5.1. Background of the Current Chinese Public Transport Situation

The problems resulting from the growth in transport demand have led policy makers in transportation to focus on development of sustainable transport systems, and especially in China, which has a population of more than 1.3 billion, public transport is still the major mode for people's travelling in their day-to-day life. At present more than 50% of the population live in the urban areas of China. In addition, economic expansion in urban areas brings many work opportunities that attract a large amount of population from rural regions (Chen and Mao, 2003). High proportion of population with high travel demand in urban areas of China has become one reason, resulting in traffic congestion in China today.

Although public transport is playing a key role in people's life compared with other transport modes, statistics indicated that the bus market share had been turning down in recent years in China because of its lower travel efficiency especially in urban areas. Since the early 1990's, bus use in Chinese major cities decreased from 30% to 10% (Xihua Daily Telegraph, 2006). Average speed of public vehicles (e.g., buses) in many Chinese cities decreased from 17 km/h at the beginning of 1980's to 9 km/h at the end of 1990's. On the other hand, insufficient road capacity and increased car traffic in urban areas also impact the operation of public transport. Another phenomenon worth noting is the increase of the number of private cars in some big cities in China, resulting in the switch of some passengers from public transport to private cars. From 1985 to 2000, private car ownership increased more than 20 times (0.3 million in 1985 to 6.3 million in 2000) in China (Kim, 2002). For the Chinese government, in order to promote sustainable development to increase the attractiveness of public transport, the reduction of road congestion and environmental pollution, it is extremely necessary to enhance the service quality of public transport by means of some new technologies.

As one of the promising techniques in the 21st century, intelligent transport systems (ITS), which combine a broad range of wireless and wire line communications-based information, control and electronics technologies, have been gradually applied to public transport to improve efficiency and effectiveness for operators, service quality for passengers, and environmental influences for society as a whole. Intelligent public transport system applications, including automatic vehicle location (AVL), passenger information systems, electronic fare payment (EFP), traffic signal priority and so forth represent a significant opportunity to improve the efficiency, attractiveness and safety of public transport systems. These applications primarily improve the operation of a transport system by either performing a function quicker or more reliably, or by providing a service that was not previously available. By these means, ITS provide for improved mobility of people and goods on the existing surface systems, and they offer the potential for substantial savings in future construction on transport infrastructure.

2.5.2. Current EFP Applications for Public Transport in China

In Chinese public transport systems, three fare payment types are used in the current stage:

- Cash;
- Travel cards and;
- Smart cards.

The details about these three payment applications in Chinese public transport systems are discussed as follows:

Cash:

As the most conventional fare payment method, cash fares are collected by transit personnel on board. Although cash fare can be flexible, without limits on public transport services and time period used, available all the time for all PT users, the disadvantages have become the main issue in this kind of fare collection/payment method. First of all, excessive transit personnel need to be on board (e.g., conductors) to issue and check tickets. This has impacted the staff utilisation and the operational costs, for example, in Beijing, normally a bus is allocated 1-2 conductors, except the bus driver, because passengers can get on and off through all doors on board in order to reduce bus dwelling time at bus stop and under such circumstances, one bus driver obviously could not take all responsibilities on board. Secondly, fare fraud and evasion always happened in China, for example, in Guangzhou, fare fraud caused about 1 million yuan (equivalent to 67k GBP) loss of public transit revenue per month in 2001 (ITS China, 2004). Thirdly, fare information under cash fare payment applications is not very clear for PT passengers. For example, in some cities applied zonal fare (or distance-based) policy, passengers always asked PT vehicle drivers about fare information, resulting in the increase of the workload of PT staff and a low operational efficiency.

In the last decade, in order to improve the efficiency of cash fare collection/payment, some cities introduced fare boxes to replace conductors to collect cash fare on board, such as in 1995, Dalian public transport system started using fare boxes and by 1997 fare boxes had thoroughly replaced conductors on board. But except that the problems would happen as discussed in the last paragraph, the fare box brought another new problem: passengers must prepare for exact cash or coins to insert the fare box. For the security reason, public transport drivers are not allowed to give passengers change back if passengers cannot pay exact money. So with the improvement of operational efficiency and throughput, inconvenience to PT passengers has become an issue to be solved.

Travel Cards:

Pre-paid monthly/quarterly travel card is another application in public transport fare payment. Several characteristics/issues of travel card applications in China can be summarised:

(1) Currently most travel card applications are pre-paid on the basis of per month or per quarter (unlike some other countries who applied short term travel cards, e.g., daily and weekly). Each month or quarter users have to renew their travel cards at ticket offices when travel cards expire. Travel card is a kind of cashless payment and brings quicker boarding time for passengers and operational efficiency to PT operators.

(2) Limited PT services or modes: most travel card applications in China limit the PT service routes that passenger can take, but without limit on the number of trips per month/quarter. In other words, the travel cost of travel cards is on the basis of PT routes, rather than zones, like the UK applications. Therefore, it is quite common that a passenger buys two travel cards for two different PT services. Also, if occasional travel demand happens on PT services that their travel cards cannot cover, travel card users have to use other more flexible payment methods (e.g. cash) to pay fares.

(3) Fare fraud: this is a serious problem in China, for example, dozens of faked travel cards were found out on one bus route within one month on average in Dalian (Dalian Evening Post, 2006). Because to produce travel cards does not require advanced technique, it seems easy to fake travel cards. In Dalian, public transport operators have taken some legal actions against this behaviour.

(4) Stable customer groups: due to the limited travel card products in China, the current customers are mainly students and commuters, who have high travel frequency and stable travel purpose. For those non-travel card users, high travel cost pre paid for long periods of time (e.g. one month in advance) is the primary reason not to choose travel cards.

Smart Cards:

EFP applications (smart cards) in China can be tracked back to the late 1990's, when the first smart card ticketing for public transport was introduced in Shenzhen in Nov 1996. Over about ten-year development most Chinese major cities have been implementing the smart card ticketing in the public transport systems.

A characteristic of smart card applications in China different from other countries is that unlike UK, USA, Sweden, Singapore, *etc*, which experienced the stage from contact fare cards (e.g. magnetic stripe cards) to contactless smart cards, China is experiencing the change from the traditional fare payment types to the smart card ticketing. Now, the contactless smart card technology is the main pattern for electronic fare payment in Chinese public transport. For instance, in Dalian, 1.2 million contactless smart cards have been issued since July 2001. And 300-400 new cards are demanded everyday. In Shanghai, the daily transaction of smart cards is over 4 million (ITSC, 2003). Since May 2006, smart card payment has thoroughly substituted for monthly travel cards in Beijing through the enforced policy, which can be viewed as the start of smart card implementation in a large-scale. The details about smart card application in Chinese cities are listed in Table 2.4.

As can be seen in Table 2.4, most smart card applications in China were implemented in the recent 6 years. Huge population in Shanghai (17.7 million by Oct 2005) brings the most smart card holders (10 million by 2005) among cities implementing smart card projects. In most cities, smart cards have different products for different users, for example most cities have student smart cards. But compared with applications in other countries (e.g., the London Oyster card, the New York MetroCard, *etc*), options of smart card types for card users in China still seem limited, for example, in Dalian, at the moment there are only two kinds of smart cards products: "pay as you go" cards and "pre-paid monthly" cards. Another feature in Chinese smart card applications is that the deposit policy is implemented in all applications. The aim to require deposit for the initial purchase is to guarantee good physical condition of smart cards when card users cancel and return their cards. But recently some card users questioned the deposit policy with local public transport operators: whether the deposit is really necessary to cover the maintenance costs of PT operators, because operators should present more evidences to explain relationship among the annual profit of PT companies, smart card manufacture cost, card depreciation and smart card life time.

Table 2.4 Smart Card Applications for Public Transport in China

City	Year introduced	Population (in million)	Card ownership	PT modes applied	Ticket Packages	Other features
Beijing	2003	14.93	1.5 million	Bus, LRT, underground and taxi	Pay as you go cards; pay monthly cards	Deposit: 30yuan; topping up at banks and ticket offices only.
Shanghai	Dec 1999	17.78	10 million	Bus, ferry, LRT, underground and taxi	Pay as you go cards and student cards	Deposit: 30yuan; multifunction: shopping; crossing different nearby cities; topping up at banks and ticket offices only.
Tianjin	Jan 2000	10.4	1 million	Bus	Pay monthly cards	Deposit: 18yuan; topping up at banks and ticket offices only.
Dalian	Jul 2001	3.12	1.2 million	Bus, tram, light railway and taxi	Pay as you go cards; student cards and the elder cards	Deposit: 50yuan; shopping and telecommunication function; topping up at banks and ticket offices only.
Guangzhou	Dec 2001	10.5	4.6 million	Bus, ferry, underground and taxi	Pay as you go cards	Deposit: 30yuan; integrate with nearby cities; multifunction: shopping, telecommunication, ID cards, banking.
Shenzhen	Nov 1996	5.9	N/A	Bus	Pay as you go cards and student cards	Deposit: 40yuan; automatic adding value machines applied; integrate with Guanzhou and Hong Kong.
Hangzhou	Sep 2001	6.22	3 million	Bus and taxi	Pay as you go cards, the elder cards, student cards, tourist cards	Deposit: 25yuan, automatic adding value machine
Shenyang	Mar 2003	5.5	1 million	Bus and taxi	Pay monthly cards, student cards	Deposit: 30yuan; topping up at banks and ticket offices only.
Nanjing	Dec 2000	5	2.2 million	Bus, ferry and taxi	Pay as you go cards, student cards,	Deposit: 30yuan; topping up at banks and ticket offices only.
Hong Kong	Sep 1997	6.94	12.4 million	Bus, tram, LRT, ferry, underground and taxi	Adult cards, Elder cards, Child cards, 3-day Hong Kong Transport Pass	Deposit: HK$50; automatic adding value machine; multifunction for shopping, parking fee payment; Interoperation with Shenzhen public transport system; reward scheme for the loyal users.

Two aspects of smart card applications in China that are not as good as applications in other countries and areas are "top-up methods" and "multifunction". In Table 2.3, most projects can only provide limited top-up options to card users, such as adding value at ticket offices and banks only. The variety of top-up methods would be enhanced in the future development by improving the integrity of different top-up options, such as through mobile phones, the Internet, automatic adding value machines, *etc*. As to multifunction, the potential of card integrity among different social service sectors (e.g. banks, retail stores, telecommunication, personal ID, *etc*) has not been fully utilised in China. The expected benefits to card users should not be on public transport fare payment only.

2.5.3. Necessity to Implement Electronic Fare Payment in China

a) Owning a private car is far too expensive for most Chinese households in current China and hence public transport is still the principal mode for the majority. For the Chinese government, in order to develop sustainable transport systems under the limited resources (e.g., land use and energy consumption), some policies must be made to promote public transit over private vehicles, increase the attractiveness of public transport, harmonise ticketing for both occasional users and daily commuters.

b) Fare fraud, including cash fare underpayment/evasion, false bus travel cards, and fare disputes, has resulted in loss of transit revenue in China as we discussed previously, and can be stressful to PT drivers. By successful implementations in other countries, smart cards have effectively decreased the frequencies of such incidents due to the relatively more secure design and fare collection procedure under the smart card ticketing.

c) The cost of public transport personnel is another reason for public transport operators to implement the smart card ticketing. A great amount of public transport personnel in most Chinese public transport companies are conductors. The ticketing duty, such as issuing tickets, checking the validation of fare cards, reporting the information of bus stops and routes, can also be taken by bus drivers under electronic fare payment systems. For example, in 2003, PT fare collection cost in Shanghai was 242.7 million yuan (16.3 million GBP), decreased by 50%, comparing with year 1999 (496.4 million yuan, before the implementation of the smart card ticketing).

d) Using fare boxes to allow passengers to pay the fare is an improvement, comparing manual ticketing by conductors, but it still has many disadvantages. PT drivers often have to "un-jam" fare boxes, or assist passengers in inserting bills. These increase drivers' workload and delay the bus departure. Sometimes the drivers have to answer questions about fare information from passengers during boarding, or even when vehicles are moving. Undoubtedly, these could increase the workload of drivers. Also, they might result in potential incidents when drivers are under work conditions.

2.6. Summary

In this chapter, the development, technology and advantages of EFP in the world are overviewed. Through reviewing smart card applications, we can see that as a new fare payment technique, advantages of smart cards have been realised by PT operators, PT users and authorities, such as boarding time saving, improved operation efficiency and multiple applications, *etc.* The advantages over traditional fare payment methods determine the success of smart cards for public transport throughout the world, becoming one major payment option in public transport systems. Through reviewing pros and cons of smart cards of previous experiences, another purpose of this chapter is to determine the study objective. This research focuses on benefits and effectiveness of smart cards to users. A user demand analysis is proposed to employ so as to examine how PT users assess their current payment applications and future improvements.

Secondly, from reviewing smart card applications, we can see that Asia would be a promising market in smart card applications in the future, in addition to those well-developed markets in North America and Europe. Particularly, the discussion of the current public transport situation in China suggests that due to the unique situation in PT operation, it is important that China should make use of these advantages of smart cards to enhance PT services. Therefore, the unique situations in China, such as high volume of PT users, low efficiency of operation and service quality, etc, determine the reasons for carrying out the evaluation in China. The implications from reviewing smart card applications in the world and China are: to determine the necessity for implementing smart cards in China; to identify which aspects in the current situation of payment applications could be improved in China.

In addition, through reviewing smart card applications, features (attributes) of smart cards, which have been widely used in the current applications, can be determined and are considered as key factors to evaluate benefits of smart cards based on users' preference for this research. Therefore, in the following chapter about the survey design, fare payment features related to cash, travel cards and smart cards are used to collect information we are interested in.

Besides a review of smart card applications throughout the world, a detailed literature review on evaluation studies of EFP applications in the world is discussed in the next chapter to give an insight of evaluation methods used for explaining benefits and effectiveness of the smart card ticketing.

Chapter 3
Literature Review

3.1. Introduction

This chapter describes a detailed literature review about the evaluation studies on smart cards for public transport. The literature review in this research can be divided into four different parts: "before and after" studies; users' preference studies; operators' perception studies and cost-benefit analysis. The purpose of this chapter is to have an insight into previous evaluation studies of benefits and effectiveness of smart cards for public transport and get implications of the research methodology of this thesis.

Section 3.2 categorises and discusses these four aspects of previous evaluation work. In "before and after" evaluation, previous studies mainly focused on evaluating performance of smart cards projects comparing with conventional fare payment methods. Paynter and Law (2003) introduced users' revealed preference surveys for the Hong Kong Octopus card use on individual level. Through users' preference studies, factors (payment features) that have been widely used and could be introduced in fare payment alternatives were determined. Besides analysing demand aspect, some literature also focused on public transport operators' perceptions to smart card applications, such as through face-to-face interviews, on board surveys and observations to investigate attitudes and assessment about the smart card ticketing from PT drivers and managing staff.

Following Section 3.2, the contribution and significance of previous studies to this research is discussed in Section 3.3. It can be helpful to generate the research outline and specific analytical methodology for this research.

3.2. Review of Evaluation Studies of EFP

Evaluation studies of smart card applications can be tracked back to the 1990's, when some trial-run smart card projects were carried out in USA, Europe and Asia. The evaluation of I-110 corridor smart card application in California, US can be viewed as one of the earliest studies for evaluating operators' perceptions to smart cards, which was conducted in 1996 (Chira-Chavala and Coifman, 1996; Giuliano et al, 2000). Except those evaluations on trial-run smart card projects, key studies on large-scale smart card applications have been presenting more mature pictures for the smart card ticketing, such as the evaluation of the Hong Kong Octopus card based on the survey among public transport users (Paynter and Law, 2003). Cheung (2003) discussed lessons after the smart card application in Holland, which would be useful to provide guidelines (or measurements) for implementing smart card ticketing nationwide, such as technical reliability, price principle, user accessibility and acceptability, *etc*. Opurum (2005) evaluated the smart card ticketing in the rail rapid transit system of New York. In Opurum's work, cost-benefit analysis was used in conjunction with the results of elasticity-based transit demand model and ticket choice (legit) models in determining the profitability of the smart card ticketing application.

All these evaluation studies of EFP explained benefits and effectiveness of smart cards from different angles, therefore these previous work can be categorised as following aspects: (1) "before and after" studies; (2) user preference studies; (3) operators' perceptions studies; (4) cost-benefit analysis.

3.2.1 'Before and After' Studies

The definition of 'before and after' analysis of the smart card project is to compare measurements related to public transport operation (e.g. revenue, operational efficiency, *etc*) and passengers' rider ship (e.g. changes on travel frequency, boarding time and convenience, *etc*). However, in order to compare any changes before and after the new payment application, such kind of analysis requires the involvement of fare payment data over long periods of time, particularly for data before the new fare payment application is introduced (Harding, 2006). In "Before and After" studies, the benefits and effectiveness of smart cards were explained by comparisons of selected measurements between before and after the smart card implementation.

Savage (2000) evaluated the New York MetroCard by using "before and after" method. The New York MetroCard was firstly implemented in January 1994. Now the MetroCard is a major payment method for the underground and bus services in New York City. The fare packages based on the MetroCard include single ride cards, pay-per-ride cards, unlimited ride MetroCards (from 1 day to 30 days), student Metrocards, and disable/senior citizen MetroCards. All these packages enrich people's alternatives of smart cards in the market place.

In this study, the author measured the following aspects: ratio of fare fraud and evasion before and after smart cards implementing; change of average weekday ridership; and reliability of the new ticket check-in system. The evaluation data were mainly from the central database of the ticketing system. Key findings about benefits and effectiveness of smart cards can be summarised as follows:

(1) 90% fare fraud and abuse were reduced after using the smart card system, because of more secure design of smart cards and check-in system than paper-based travel cards;

(2) Comparing with situations before the smart card application (i.e. cash and travel card payment methods), the total number of trips after the smart card application (i.e. the

mixed payment system of cash, travel cards and smart cards) increased by 13% for underground service, 41% for bus service in 1999, which indicated that the new fare payment application boosted the travel demand of public transport, because convenience for PT users by using smart cards increased the accessibility to the public transport services. On the other hand, the purchase bonus policy (purchase>$15, receive extra 10% bonus) that had not been applied before the smart card ticketing is another reason to increase the travel demand by using smart cards.

Except the findings in this study, the following two aspects, interoperation and customer acceptance, could be the major two issues after the smart card implementing and need to be addressed in further details. Because of the low interoperability of the smart card in New York in the early stage, the benefits of smart cards were only measured in those major PT operators, while fare payment/collection data with small bus operators were not taken into account. Secondly, user acceptance would be one of reasons to influence the use of smart cards, particularly in the New York underground service. For example, a new check-in system was introduced along with the implementation of the smart card ticketing. Due to unfamiliarity with the new turnstile application for the smart card ticketing, users' and potential users' acceptance of the smart cards would result in the choice behavioural changes. Therefore, under such circumstances, it is obviously not sufficient to study those common measurements that only can be found in both 'before' and 'after' stages. Those new features after the implementation of the smart card ticketing need to be presented to PT users to measure benefits on individual level by comparing changes of individual attitudes and choice behaviour between 'before' and 'after'.

The implication from the study is that: except the survey about the current use of smart cards, it is necessary to investigate and examine new features have been and will be introduced under the implementation of the smart card ticketing (e.g., more extra functions being added, wider geographic areas being covered, a variety of card products associated with fare differentiation, seasonality of payment type, *etc*). In this research all of these concerns can be designed in a survey for individual PT users to examine the current choice preferences and behavioural changes on fare payment options with and without these features, so that the trend of changes in demand, benefits and acceptance of smart card payment may be identified more intensively.

Compared with the 'before' and 'after' analysis based upon operator's level in Savage (2000)'s research, Rainio (1998) surveyed both PT operators and users by face-to-face interviews to compare the performance of the smart card ticketing with conventional payment methods in Tampere, Finland. The smart card ticketing in Tampere started at the beginning of 1994, when electronic ticketing machines and radio modems were installed in local buses. Since July 1997 the new fare collection system has been in full scale operation.

The main ticket types on the smart card are season ticket and stored value ticket. So far 150,000 smart cards have been distributed in Tampere's public transport service (Salonen, 2006).

(1) For operators, three measurements were determined and compared between 'after' and 'before' situations: satisfactory degree, reliability of smart cards and revenue allocation. Except that the revenue allocation before and after the smart card implementation can be compared by statistical data, the satisfactory degree and reliability of smart cards were surveyed via asking five-scale questions (worst, bad, neutral, good and better) to PT operators (managing staff and drivers). Unlike revenue allocation data, the satisfactory degree and reliability about before and after the smart card application were only related to individuals' perceptions.

(2) Regarding the revenue allocation, it was found that after the smart card implementation, the revenue was increased by 20%. Prepayment of smart cards is the reason to cause the increase of revenue.

(3) For card users, the author mainly surveyed individual boarding time changes before and after the smart card ticketing. In an interview with card users, a great majority of passengers reported that the new system was better (64%) or as good (27%) as the old one. Usually (72%) the interviewed passengers estimated that boarding times were shortened after using the smart cards; only some (6%) thought they were lengthened. However, the measured boarding times were only slightly shorter than those of the old system.

The key finding and contribution to this research is that: under different fare payment applications, boarding time savings and convenience for PT users can be selected as one of measurements to identify benefits of smart cards, because this research focuses on the evaluation analysis from demand side (PT users). Although in this literature, individual boarding time savings had been reported through the interviews, how much boarding time savings could influence the changes of choice behaviour and demand of the smart card ticketing, needs to be identified in further analysis.

On the other hand, surveying individual users is a direct way to collect the performance of the smart card ticketing. Face-to-face interview method, which was used in this literature, can achieve higher response rate and better respondents' understanding to questions asked, however the relatively high survey cost must be taken into account, particularly for this PhD research, which is proposed a large scale sample size to collect preference data from individual PT users, some other cost-saving survey methods are more suitable to use.

US DOT (2003) evaluated an operational test of ORANGES smart card in USA.

ORANGES smart card project is a joint smart card ticketing application among several public and private sectors, including Florida, the Orlando-Orange County Expressway Authority (OOCEA), the City of Orlando and the Public Agency Partners, *etc*. The smart card application is not only on public transport fare payment, but also on parking fee payment and expressway tolling, *etc*. Phase I of the evaluation study was mainly to analyse the "before" data of the smart card application.

The quantitative and qualitative data required by the test plans were collected and used for qualitative assessments, for comparison with the testable hypotheses, and for quantitative goals where the assessment of "before" data was applicable. The assessment of qualitative data was presented by the "before" and "after" discussion groups. The role of the initial data collection was to gather "baseline" data about initial conditions before the FOT (Field Operation Test).

In the quantitative evaluation of Phase I, five goals were determined to assess the "before" data:

- Reduced transaction times, measured by average payment transaction duration, for each mode and type of equipment. The test hypothesis is prepaid payment transactions will be quicker than cash payment, so the average duration will decrease if the 'percentage prepaid' increases;
- Increase prepaid revenue share, measured by 'percentage of transactions' that use a prepaid revenue payment method. The test hypothesis is percentage prepaid transactions will increase for equipment accepting the ORANGES card;
- Automated payment equipment uptime, measured by percentage operating hours with cash processing available (coins for toll Automatic Coin Machines (ACMs); coins and bills for fareboxes). The test hypothesis is the frequency and severity of planned and unplanned maintenance for unattended devices relates to the amount of cash processed. Cash processing availability should increase as % prepaid increases;
- Current travel card distribution costs, measured by costs for distributing conventional weekly and monthly travel cards;
- Current processing cost per cash transaction, measured by costs for processing cash, for each mode.

Most measures were based on a sample and statistical analysis was performed by evaluation team. When doing statistical analysis, it is important to note that unforeseen circumstances may cause the variations in data. For example, the duration for a set of boarding transactions varied due to factors such as how long people take to pay with cash or whether the driver is asked for directions or fare information, *etc*.

In the qualitative evaluation, the analysis was carried out through the "after" data, presented by group discussion. The qualitative goal of the evaluation was to understand perceptions of system users by different user categories. Two groups were studied: card holder group (user) and operator group. Discussion groups focused on and collected information about the following general topics:

Card holders:

- Convenience of use;
- Trust and comfort level of use;
- Reporting, informational needs (statements, *etc*);
- Discounts and incentives;
- Attitudinal perceptions regarding investment of effort by agency as compared with focusing on core functions (e.g. does a multipurpose smart card has benefits to users and is this worthwhile effort of the agencies?).

Operators:

- Perceived convenience of use to customer;
- Convenience of use to operators;
- Perceived trust and comfort level of use by customer;
- Trust and comfort level of use by the employee (are there concerns that employers will monitoring employees, for example);
- Trust and comfort level of use by the operator (are there management, concerns such as privacy, liability, monitoring employees, *etc*?) Reporting and informational needs (data collection, reports, statements, data storage, record-keeping, market research, marketing, *etc*)
- Discounts and incentives (planning, management, marketing, recordkeeping)
- Reliability and quality control (operations, maintenance, planning, management issues)
- Attitudinal perceptions regarding investment of effort by operators as compared with focusing on core functions (e.g., does a multipurpose smart card have benefits to users and is this a worthwhile effort of the agencies?)

Key findings and implications from the evaluation study are summarised below:

- Quantitative analysis:
  (1) Average transaction time was 10-13 seconds after using smart cards, much better than fare payment method (20-30 seconds on average);
  (2) At the 95% confidence level, the average prepaid ridership share is about 58% in the market place, higher than “before” data (39%);
  (3) At the 95% confidence level, the average farebox % availability is about 99.12%, presenting the similar result with “before” data (99%).

(4) Salary/benefits cost for the customer service staff that sell the travel cards reduced from $727 to $582 per 1000 travel card sold after the smart card implementation.

(5) Salary/benefit cost for the accounting clerks in the money room that process cash revenue from both travel card sales and fareboxes reduced from $13.42 to $12.55 per $1000 cash revenue after the smart card implementation.

- Qualitative analysis is helpful to determine the attributes of the users' perceptions toward the smart card ticketing for the coming Phase II analysis. Particularly for the attitudes to extra services of smart cards, it may be presented to respondents by introducing some new variations in the "after" stage so that it can provide more detailed description to users.

Cheung (2003) more concerned lessons and limitations of the trial-run smart card project in Holland by comparing card users' assessment and experience 'before' the smart card application ("Tripperpas") with 'after'. The objectives of the experiment were to determine the efficacy of the contactless smart card technology and to assess the acceptability of the new ticketing method in a working urban environment. Another aim was that the practical experience gained could be used to guide the development and implementation of a nationwide electronic ticketing system. The evaluation work was based on four surveys (one before and three after) via questionnaires, group discussion and interviews among smart card users. Meanwhile, additional surveys were conducted to investigate the responses of nonusers and ex-users. Each survey consisted of a series of focus group discussions with a stratified sample of existing and potential Tripperpas users as well as face-to-face interviews with members of the operating staff. Particular attention was paid to the mobility needs of senior citizens and students. In addition, in each and every after study, 2,000 questionnaires were mailed with return-paid envelopes to the addresses of Tripperpas holders to survey passenger reactions. To determine the views of the nonusers, 2,400 questionnaires (of a slightly different design) were distributed at a stratified sample of bus stations to passengers who did not have Tripperpas cards.

The research findings presented a comprehensive picture of the strengths and weaknesses of Tripperpas ('after' study) compared with the conventional strip ticket ('before' study). They also provided a valuable insight into travel behaviour as well as staff and passenger preferences after the smart card ticketing. The author investigated the following aspects to examine the passenger preference:

1) Reasons for purchase: what features had attracted the passengers to change from the conventional ticket system to the new card were ease of use and convenience; for non-users, 30% of respondents thought they did not perceive the advantages of smart cards,

and only 8% considered trying out the smart cards in practice.

2) Combinations of different fare payment means: although smart cards can be used in the urban area, there were nearly 25% respondents still used smart cards and traditional payment methods jointly. Limited coverage of smart cards in different regions resulted in this phenomenon.

3) Overall assessment of card performance: 90% of respondents replied that the experience of using the card was positive and that the system lived up to their expectations in this study. But about one third respondents gave negative responses on check-out procedure and sensitivity of malfunction of apparatus in the smart card ticketing. For nonusers, the lack of transparency concerning how the travel costs were being calculated and the lack of an overview on the travel costs were the prime concerns, particularly among season ticket holders.

4) In the 'after' study, price principle (discounted fare on smart cards), extra services in one card and card accessibility in the nationwide level, were perceived by the respondents after the smart card application.

All in all, these evaluation studies provide evidences on smart card attributes respondents perceived in practice and these attributes can also be included in the survey design of this research. These key factors include: boarding time saving, travel cost, changes of ridership, individual assessment of payment convenience, etc. However, regarding the evaluation methodology, all outcomes in the paper seemed to focus on the 'after' surveys. What respondents experienced in the 'before' survey and how to compare with the 'after' were not presented by the author to illustrate the benefits and advantages of smart cards over the conventional fare payment methods. Another issue in 'before and after' methods is that analysis results cannot provide importance of factors of smart cards by comparing across these selected attributes.

Compared with the 'before' and 'after' analysis, several evaluation studies focused on users' preference studies only after the smart card implementation. Because all of these previous studies were based on passengers' actual use of smart cards through other more cost-saving survey approaches than face-to-face interviews widely used in investigating 'before and after' smart card applications, such as self-completion surveys on board, thereafter, pros and cons of different survey methods used on surveying users' preference data can be reviewed and referred for this research.

3.2.2. Users' Demand (Preference) Studies

To promote new technology for public transport fare payment, it is necessary to understand users' preferences, because benefits and effectiveness of smart cards can also be evaluated by considering how passengers choose and what their perceptions would be based on new fare payment situations. Although there are very few evaluation studies on users' preferences about public transport smart card applications in recent years, these limited number of studies have explained users' preferences based on individual surveys to some extent, including determining the importance of fare payment attributes, different fare schemes, and extra services could be added on the current applications, *etc*. Therefore, benefits and effectiveness of smart cards were assessed by respondents' individual choices in users' demand analyses, based on attributes of fare payment applications.

Paynter and Law (2003) evaluated users' preference about the Hong Kong Octopus card. Since the Hong Kong Octopus card was first implemented in 1997, 7.4 million cards had been issued by the moment of their study. However, relevant studies based on individual perceptions had not been done before Paynter and Law to provide an insight into peoples' responses to fare payment attributes in detail.

The aim of this evaluation study was to explore and develop an understanding of the existing Hong Kong Octopus smart card use and to ascertain other potential uses and how the system can improve its utilisation and the services provided. A multi-methodology approach was used in this study based on a revealed preference survey, in which they sampled 800 actual and potential card users. The use of the Octopus card was widespread with 94.3% of those 800 sampled respondents. The evaluation method used in this study was to statistically analyse the survey data by categories and segmentations to find the main response to different attributes of the smart card application.

Information on card ownership, usage of cards, failure rates, top-up habits, results from promotion schemes and loyalty programmes, satisfaction on the number of services supported and suggestions on potential applications, success factors of the Octopus card, reliance on the system and factors of how the system can be improved, were examined by the survey. Through the descriptive statistical analysis (e.g., mean value, distribution by different attributes with respect to socio-economic backgrounds) for the data, it was found that security feature, multifunction and accuracy rate of card use were three most important factors to be improved in the future application.

Some findings and implications from this study can be summarised as follows:

- Survey methods: convenience sampling process with self-completion questionnaire was used in this study, yielding 507 questionnaires returned with the response rate of 63%. The definition of convenience sampling is described as choosing individuals that are easier to reach. One of advantages of convenience sampling is that it can save survey cost and secure good response. The major disadvantages of this sampling technique is that how representative the information collected about the sample is to the population as a whole may not be explained by the authors (Joppe, 2006). To reduce the bias coming from convenience sampling and increase the representativeness of the surveyed data, a potential solution is to survey in clusters as many as possible (e.g., different geographic areas, communities, *etc*), each of which contains a certain proportion of sample capable to be surveyed (Hensher, 2000). In addition, segmenting data is also helpful to solve the issue of non-representativeness of data.

- The study not only focused on existing smart card users, but also surveyed those potential card users. It may be a good way to identify attraction of smart cards to non-users, but how questions in questionnaires were addressed specially for such group of people and how they responded to these questions differently from existing card users have not been explained by the authors. Because current card users and non-users are different two components in the market place, which could generate different perceptions towards the smart card ticketing, it was necessary to examine such differences of perceptions between card users and non-card users.

Comparing with the evaluation study by Paynter and Law in Hong Kong, Wang *et al* (2003) also evaluated benefits and effectiveness of the EasyCard in Taipei by focusing on existing smart card users and non-users. But a more detailed analysis for different user type was conducted in the evaluation study in Taipei. Three objectives were determined for Wang's study:

- Understanding users' preference to the EasyCard;
- Assessment of satisfactory degree of customers;
- Suggestion on policy making about the EasyCard application: e.g. whether the EasyCard should thoroughly substitute other conventional fare payment methods in practice.

The research methodology was based on the questionnaire and statistical analysis. 1200 public transport users were sampled in 10 different bus stops/tram stations, each of which was distributed 120 questionnaires. The following variables were selected in the questionnaire as measurements by presenting multiple-choice questions: card type; public services (including public transport) can be used; discount levels; attitude to the deposit policy; reasons for complaining; overall assessment, *etc*. The survey method was at bus stop/tram station survey, because this survey method can easily target the survey population specific for this research context (public transport related). In order to survey at bus/tram station, first of all 62 public transit stations were numbered in Taipei. Then 10 of 62 stations were randomly selected as survey locations for this study. Regarding the survey time planning, weekday and weekend were balanced to include different travel purposes (home-work and home-school on weekday; leisure on weekend). Meanwhile, from 6am, the start time of operation, to 12pm, the end time of operation, these 18 hours were divided into six time slots, each having 3 hours. For these time slots, they were also randomly selected as the survey time. All these actions not only secured the randomness of the sample, but also increased the representativeness of respondents in the sample. Finally 1110 questionnaire papers were collected back with a very high response rate of 92.5%.

The key findings are summarised as follows:

- Through statistical analysis for each single attribute of the smart card application, the majority responded positively on the use of smart cards. For example, 65.88% and 15.5% of respondents felt satisfied and very satisfied with the smart card ticketing, respectively.
- When asking whether extra services had been used by respondents, 67.01% of respondents selected telephone card function, and 37.42% usage on shopping payment. Telecommunication and shopping have become the two major extra services in addition to the payment function for PT fare.
- The motivations to use smart cards are: 77.56% responses were because of convenience and 57.39% of respondents answered quicker payment and boarding. Because multiple choice question style was used in this question, the total percentage of different responses was not necessarily required to be 100%.
- Through cross-attribute analysis between satisfactory degree and some other factors, such as socio-economic factors, the causality between users' satisfactory and influential factors can be captured. The detailed causal relationship is described in Table 3.1.

Table 3.1Cross-Attribute Analysis between Satisfactory Degree and Other Factors

	Very Unsatisfied and Unsatisfied	Satisfied and Very Satisfied
Convenience—Satisfactory	3.34%	74.56%
Deposit—Satisfactory	3.34%	40.8%
Smart card ticketing policy—satisfactory	2.34%	70.31%
Educational Level—Satisfactory	2.3% college degree or above	52.83% college degree or above
Residential Area—Satisfactory	0.94% in Taipei urban area	33.18% in Taipei urban area
Transaction Speed—Satisfactory	2.1%	62.98%

In Table 3.1, it can be seen that convenience of the smart card payment is the main reason to make users satisfied (about 74.56% respondents felt satisfied or very satisfied with the smart card ticketing, while only 3.34% were unsatisfied or very unsatisfied). Meanwhile, because discounted fare policy was applied in the smart card ticketing, another factor to achieve users' satisfaction is the smart card ticketing policy, such as the discounted fare policy, subsidy for smart card users, *etc*. The third aspect we can see the benefits of smart cards is the transaction speed: about 62.98% respondents were satisfied or very satisfied with the transaction speed (i.e., quicker boarding time). Compared with three factors mentioned above, users' satisfaction about deposit and residential area covered is relatively low, only reaching 40.8% and 33.2% respectively. That is to say deposit policy would influence PT users' payment choices. In addition, if smart cards can cover wider areas, users' satisfaction and acceptance also would be increased. In the mean time, authors analysed influence of educational background on the assessment of smart card use. The result shows that users with higher educational level more would like to accept and use smart cards than those with low educational level.

Although the authors analysed the causality between satisfactory degree and other influential factors, all of these analyses were based on satisfactory degree and a single factor. How the combination of these factors (fare payment attributes) could influence respondents' assessment and satisfaction has not been explained. For example, if considering deposit, convenience, transaction speed and fare discount level, four factors, then what could the respondents' decision and assessment be? Because economic consumer theory states that individuals make choices between alternatives by evaluating their attributes combined in some way (McFadden, 1981), simple causality between respondents' decision and a single attribute may not be sufficient to explain card users' choice behaviour and perceptions. Therefore, it is necessary to analyse the influence of the combination of payment attributes on users' satisfaction.

Foote and Darwin (2002) discussed the results of a survey of Chicago transit riders participating in the pilot programme designed to test the technological feasibility and customer acceptance of the smart card fare payment. The smart card differed from the magnetic stripe fare card deployed across the Chicago Transit Authority (CTA) system in 1997. 3,500 CTA customers purchased the $5 cards and participated in the pilot programme. A mail-in survey of all 3,500 programme participants yielded 1,300 responses for a 37% response rate. Results of the survey should be taken to represent the views of pilot programme participants rather than CTA riders as a whole, given the self-selected nature of the group opting to participate in the pilot programme. Geographic analysis of the survey data did indicate that respondents resided in all City of Chicago postal codes and most suburban CTA service area postal codes.

Some results of the survey effort included in the paper indicate that:

- Features related to convenience, rail use and speed were most liked by programme participants. 21% respondents rated convenience over the magnetic stripe card as their single favourite feature of the system, 15% liked being able to use the cards for train travel, 13% the time to register rail fare and 8% the convenience of the system over using cash to pay fares.

- The least liked features were the $5 fee (deposit), the need to add value to the card after paying the $5 fee, and inaccuracies in calculating bonus fare when adding $10 or more to the card.
- Features that would simplify adding value to the card were the most popular potential additional features of the smart card ticketing. Most desired were the ability to recharge via the Internet and credit card (desired most by 17% of respondents), use to pay fares on Metro as well as CTA (11%), auto-recharge via credit card (8%), recharge at ATMs (8%), and ability to move value from a magnetic fare card to the smart card (7%).

Bryan and Blythe (2007) studied smart card users' demand through analysing users' data gathered during operation (e.g., passenger behaviour, boarding/alighting location and times, etc.). The purpose of this study was to consider if practical information about the customer could be derived from smart card data and used as a tool by transport planners to create a service which better meets the users' needs, potentially attracting more customers. The card types and mifare numbers were analysed to get an idea of exactly who was using the system and how frequently. More than 80% of all the transactions over all the routes were carried out elderly passengers and by studying the number of transactions carried out by each card holder, more than 76% of holders used their cards 1 to 5 times during the 2 month period.

In this paper, the study of boarding points was viewed as one of the most significant due to the fact that it was possible to clarify which bus stops were frequented the most and when. It was also feasible to determine at which times and stops different card types were used the most. With information at this level it is easy to get an idea of how specific routes are used or extend it to a network level for monitoring route performance. The significance of obtaining this type of information at boarding point, route and network level is that the service provider knows when and where the service is being used. For example, the majority of people use this service in the morning and from 3pm onwards it is barely used. The benefit of smart card use here is that should a change be made to the service to meet different users' needs.

It also was found that the number of journeys involving more than one route was fairly insignificant and it was difficult in some cases to determine if an interchange had taken place because of potentially inaccurate alighting point assignment due, in part, to the inability of the ticketing machines to identify stops uniquely.

Bagchi and White (2004) also looked at the smart card users' data from the central database to investigate and understand smart card users' demand and travel behaviour. Similar with Bryan and Blythe's work, which mainly explained users' demand through analysing passengers' travel behaviour (boarding/alighting points and times) and graphically visualised the outcomes, this paper evaluated benefits of smart cards through investigating PT users' (including traditional fare users and smart card users) route choice and travel behaviour.

Through analysing two case studies (Southport and Bradford), the results suggested that elderly concessionary travellers in Southport make a smaller proportion of linked trips than elderly concessionary travellers in Bradford. One reason for this could be because the Southport concessionary travellers were allowed to use their existing free travel pass that also allowed travel on buses and trains, at the same time as being allowed to use their trial smart-card. A proportion of period travelcard users may use their cards for the routine journey to work, entailing a change of bus. In addition to this, the small area over which the period travel cards are valid means that average journey times may be short (compared to other areas), and that should a transfer from one bus to another be required as part of a person's journey, then it would be undertaken within a shorter time than may be the case in other areas.

The findings of the analyses indicated the importance of the main parameters of a smart card scheme, in addition to a range of other factors, in both the explanation of analysis results, and more generically for the quality and utility of smart card data. The generic stages of the design and implementation of a smart card scheme were presented, highlighting the stages where influence on the end smart card data would be the greatest.

Therefore, the examination of the nature of smart card data, the findings of the analysis and the examination of the wider components of smart card development and implementation allowed a set of factors affecting quality and utility of smart card data to be identified, and these will be presented. These factors can be used as a useful checklist that smart card industry practitioners can incorporate into the design, implementation and evaluation stages of their public transport smart card schemes (e.g., boarding time, travel demand changes, fare payment behaviour, etc.). This study helped to ensure that the smart card data produced are of relatively good quality, and are useful for a range of applications.

3.2.3. Evaluation of Operators' Perceptions to EFP

As well as evaluation studies on users' preference toward smart cards, some studies focused on operational test of smart cards to identify performance, market acceptance and benefits to public transport operators. For operators' perception studies, the benefits and effectiveness of smart cards were analysed by measurements related to PT operation, such as operational cost and productivity, operation efficiency, driver workload, fare fraud, etc.

The first public transport smart card demonstration project in California, US, was evaluated in 1996 by Chira-Chavala and Coifman (1996) and Giuliano and Moore (2000). This study assessed:

- The cost and productivity implications of the smart card system to PT operators, relative to fare boxes, the conventional fare payment method.
- The perceptions of various personnel in PT companies toward the smart card system relative to fare boxes.

The research methodology used in this study was based on data obtained from interviews of transit personnel, independent on-board observations, and personal communications with the smart card company. Public transport operation productivity is usually expressed as system efficiency and effectiveness. Efficiency is the extent to which system inputs are employed to produce outputs. For smart card use, efficiency was measured by the following criteria in this evaluation study:

- Costs of fare collection and related activities, which include both fixed and variable costs.
- Productive (or efficient) use of vehicle, which includes: passenger boarding time; vehicle dwelling time at bus stops; and vehicle down-time due to malfunctions, failures and repairs of the fare collection system.
- Amount and quality of data available from the fare collection system.
- Driver performance and workload.
- Perceptions of transit personnel.
- Fare fraud.

Meanwhile, the author compared with other conventional fare payment methods (i.e., cash and travel cards). The purpose of the comparison between conventional payment methods and smart cards in this study was to examine the following issues: 1) willingness to implement the smart card for operators; 2) experiences using the card, including problems encountered; 3) overall satisfaction with the card; 4) possible differences between the two card technologies (contact and contactless); and 5) relationships between social and demographic characteristics and response to the fare cards. The evaluation results from the following aspects revealed the performance of this trial-run smart card application and suggested the future development:

1) Boarding time savings and convenience for passengers were the main advantages over traditional fare payment (particularly compared with cash). However, such advantages were not significant between smart cards and travel cards due to the similar characteristics of the two cashless payment methods;
2) Public transport vehicle dwelling time can be reduced effectively when smart cards were widely used among passengers;
3) Opportunities of interoperation of smart card applications among different PT operators

and geographic areas provided the potential in the future application;

4) Nearly all PT drivers reported that their work load reduced after the smart card ticketing system comparing with cash fare collection;
5) Through the interview with PT staff, it can be seen that most PT managers said that they would like to see the smart card system with the following capabilities:
   - Automated passenger counting that captures details such as time of use, origin-destination of trip, the number of users by bus line and user demographics, *etc.*
   - Smart cards showing the cash amount, as opposed to the number of trips.
   - Integration of the smart card with the electronic fare box because there will always be a need for a fare box to accommodate cash riders in a foreseeable future.

Implications and contributions to this research can be summarised as follows:

- When evaluating efficiency of smart cards, it would be good to compare with some other fare payment methods. Therefore, in this research, when designing the RP and SP survey, smart cards as well as those conventional payment methods (cash payment and travel cards) will be used as choice alternatives to examine respondents' trading off in the current and hypothetical choice situations.
- Observed boarding time under different payment methods in this study can provide a reference for designing different boarding time variations in the survey design of this research.
- Although face-to-face interviews can guarantee higher response rate than self-administration survey, it would not be recommended to survey individual preference of PT users in a large sample for this research.

3.2.4. Cost-Benefit Analysis

One of research area is on appraising the value of a smart card project based on economic efficiency in resource allocation. The relevant evaluation techniques, such as multi-criteria analysis, goals achievement matrix and cost effectiveness appraisal and cost – benefit analysis (CBA), are capable of achieving this goal. CBA is a method for appraising a project from the society's point of view and taking account of costs and benefits whether or not they pass through the market (Opurum, 2005). Some previous studies looked at the CBA evaluation on the EFP applications in recent years.

Cheung (2005) explored a Cost-Benefit Analysis for the smart card application in Holland to guide the development and implementation of the smart card technology in the nationwide. The research aimed to establish the financial viability of introducing smart cards and to determine the potential benefits and possible costs from the community's point of view. The research methodology can be summarised as two aspects: the framework of

CBA and two-stage analysis method in the CBA.

- Framework of CBA: in accordance to the conceptual framework given in the Evaluation of Infrastructure Project, the OEI (Overview Effects Infrastructure) guidelines, effects divided into direct, indirect, external and redistribution effects. Applying the OEI guidelines in the appraisal of the smart card project has demonstrated the potentials of using the CBA methodology as a technical tool to assist decision-making. Taken as a whole, the information provided by the OEI evaluation offer the stakeholders not only a comprehensive overview where costs and benefits lie but also provide the raw data for additional analyses to formulate implementation strategy and to design concrete schemes for different geographical areas and for different types of services. The detailed effects categorised in the CBA as direct, indirect, external effects and redistribution cost, are listed in Table 3.2.

Table 3.2 Benefits and Costs of Smart Cards

Benefits and Costs	Factors	
Direct Effects	Passengers:	Reduction in ticket purchase time Reduction in molestation Value of extra mobility
	Operators:	Fewer fraudulent travel Fares differentiation Other cost savings Molestation of vehicles/at stations
Indirect Effects	Other applications of smart cards Reduced purchase time for employers Improved location climate	
External Effects	Reduction in molestation Environmental effects Relief to congestion	
Costs	Introduced costs Chip card costs Extra capacity costs	

Based on the framework of CBA, the analysis was divided into two stages as follows:

- The first stage: direct, indirect and external benefits by introducing smart cards nationwide, which were evaluated by net present value (NPV) of costs and benefits.
- The second stage: actor analysis for re-distribution effects. The study not only examined the total effects but also appraised the differential impacts on the stakeholders under various scenarios' by an actor analysis. The actor analysis is an integrated part of the evaluation to determine which groupings of the community, in what way and by how much are the different stakeholders likely to be affected. Actors included passengers, PT operators, central service units, concession grating authorities, central government, other business sectors and social partners. In the actor analysis, the researcher regarded effects on central service units as costs and on all the rest of actors as benefits, and

compared them in terms of money value.

Key findings from the CBA method were:

- In the NPV analysis, the expected total benefits derived from an integrated programme in implementing the smart card system nationwide were higher than the expected total costs. This result suggested that the project was a profitable investment for the nation.
- Implementation of the smart card would bring positive indirect effects to the national economy for an amount equivalent to €80-100 million. However, the external effects would be relatively limited in scope and size. Less molestation of passengers effectively would mean a reduction in the costs for employers of the passengers affected, to an amount of €40 million.
- In the actor analysis, benefits to different social sectors were identified: Public transport operators together would enjoy the most benefits with the nationwide implementation of the smart card. They would gain between €0.4-0.9 billion. Passengers as a group would benefit substantially to an amount €0.3-0.4 billion. Employers of passengers would also profit—because of faster and safer trips for their employees—but by a smaller amount estimated to be €20-30 million.

Implications from Cheung (2005)'s study are as follows:

- Influential factors: by using the OEI guideline, measurements were split into direct, indirect, external and redistribution benefits. Particularly for passengers, ticket-purchase time saving and value of extra mobility were the main two aspects of direct benefits, which were most perceived. Therefore, these two variables can be used in analysing passengers' benefits in this research.
- Conflicts between partial benefits to dominant components (e.g., PT operators and investors) and benefits to the society as a whole: benefits, when viewed separately with respect to the total project cost, suggest that they were significantly lower; but, the cumulative benefits when added together indicate that the project is profitable. However, in real situations, the introduction of the smart card will be feasible only if public transport companies and other stakeholders genuinely have confidence that the benefits associated with fare differentiation will be realised and that passengers will actually purchase and pay for the cost of the card. If it is not the case, for example for some companies the introduction costs weighed higher than the expected benefits, then there would be a lack of economic incentive to spur the companies (e.g. PT operators) to implement the smart card system.

Opurum (2005) investigated, to a feasible extent, the influence of the new fare collection method (smart card payment, called 'MetroCard') in New York rapid transit ridership, fare revenue and service on-time performance. Meanwhile, to extent possible, this study assessed the transit patrons' reactions towards new fare collection relative to the traditional token method.

Theoretical framework was based on the concept of demand elasticity and, to a large extent, on consumer choice or preference theory. Cost-benefit analysis was used in conjunction with the results of elasticity-based transit demand models and ticket choice (logit) models in determining the profitability and influence of the smart card ticketing system. Choice models (ticket choice and transit demand) used in this research aimed to determine if the New York City subway and bus ridership would improve as the result of the smart card ticketing.

The main criterion to evaluate the CBA was net present value (NPV). Two cost-benefit analyses were taken into account: commercial CBA for PT operators and social CBA for PT users, non-users and governments. Benefits for different CBAs are as follows:

- Commercial CBA: Incremental revenue (additional rides, unused, residual value); Improved cash flow (admin./labour cost saving, etc.); travel time savings.
- Social CBA: Consumer surplus due to discounted fare, convenience, ticket purchasing time, travel time, etc.

Research findings indicated that the investment in the New York Transit automated fare collection system was worthwhile. Its benefits were far greater than its cost to the society. The investment appraisal results also showed that the society, at large, would be at lest $2.5 million better off over the projected 30-year period life compared to a do-nothing scenario.

However, the author also pointed out some un-quantified effects of the smart card payment, also called '***soft benefits***' in other evaluation studies (Mulley, et al, 2004), such as convenience of card payment options to users (travel cards and smart cards). Another large benefit, which cannot be easily quantified without good origin and destination data, is the extent to which the same trip can be made more quickly, by optimal choice of route and reduced waiting time under the smart card payment. Therefore, regarding users' perceptions towards these 'soft benefits', preference study based on the existing situations and hypothetical scenarios (with some changes of payment situations) would be more appropriate to investigate consumers' psychological reactions.

Regarding implications and significance of all these previous studies discussed above to this research, the later section (Section 3.3) discusses in detail.

3.3. Key Findings and Significance to this Study

Through reviewing evaluation studies of smart card payment, and applications of FL and ANN technique in modelling discrete choice data, we can conclude that the following findings and implications may help to determine the gap between precious studies and this research:

- **Understanding choice behaviour of PT users**

For EFP applications, different fare payment schemes could help users to trade off and choose fare payment methods, which can minimise their travel costs and maximise convenience on fare payment. On the other hand, under different fare payment types, particularly after some new payment method is introduced, whether passengers understand and would like to accept and use this novel fare payment method, how their choice behaviour can be varying between traditional payment methods and smart card payment is necessary to be examined. Moreover, understanding payment choice behaviour is helpful to improve the equity issue of the smart card ticketing as discussed before (in Chapter 2). Under this circumstance, proper demand forecasting for the fare payment market in China can suggest relevant policies and reform on PT fare payment applications so as to enhance the level of services of public transport as a whole.

- **Determination of the research objectives**

Benefits and effectiveness of smart cards can be analysed by different methods in previous studies, such as "before" and "after" evaluation, users' preference studies (demand aspect), operators' preference studies (supply aspect), and cost-benefit analysis. But some evaluation studies were only based on suppliers (public transport operators) and the relevant benefits and effectiveness of the smart card ticketing to operators have been explained. However, from users' point of view, such benefits and how different fare payment methods could influence people's choice behaviour have not been clearly examined in the previous studies. Therefore, the principal objective of this research is to evaluate benefits and effectiveness of smart cards through analysing and forecasting respondents' choice behaviour toward different payment means.

When analysing choice behaviour, another aim of this research is to identify respondents' perceptions to fare payment attributes and levels being implemented, particularly those attributes of smart cards. As we discussed about smart card applications throughout the world and China, smart cards have some new features different from those conventional fare payment, such as deposit, geographic areas can be covered, multifunction, *etc.* Through studying passengers' perceptions about these attributes, it can help us to determine factors how the system can be improved.

Secondly, under the public transport-oriented policy, how to benefit public transport users and attract non-users to public transport is the primary task for policy makers and operators. Forecasting demand about fare payment use in the future not only can reflect such benefits and influence of the advanced fare payment application in the long term, but can also contribute the policy-making (e.g., establishing new fare structure and subsidy, introducing extra services, *etc*) in fare payment applications.

Thirdly, the analysis results of payment choice behaviour from user's angle can also provide useful information to PT operators, such as what the market share could be when a new smart card application was introduced, what kinds of features could be perceived by PT users and whether it is worth implementing, *etc*.

- **Determination of RP and SP survey methods in this research**

Previous studies have suggested that investigating users' actual choice behaviour should be a straightforward way to analyse PT users' preferences, perceptions and assessment of different payment applications. Therefore, in this research, revealed preference is used to collect respondents' actual choice behaviour toward different payment methods. However, if only RP survey method was used, we might overlook new possible features of payment applications and new variations of payment attributes. That is one of drawbacks of RP survey. In order to capture behavioural changes of respondents' payment choices toward new features, new variations of payment applications, another kind of survey method: stated preference (SP) is used in this research. The key advantage of SP survey is to forecast user demand based on controlled hypothetical situations.

As discussed in Section 3.2.2, users' preference toward smart card fare payment could be influenced by some other payment means. Particularly in the current market place, cash, travel cards and smart cards are the three major fare payment applications. So in this research, first of all, cash, travel cards and smart cards are viewed as three alternatives in the RP survey to identify people's actual choice behaviours.

In addition to investigating respondents' actual choices, previous studies have not looked at changes of individual choices when some new payment situations are introduced, such as new features, new variations of attribute, *etc*. Whereas RP data describes actual choices in terms of a set of market-based measurements of attributes of alternatives (which by definition are restricted to the currently available feasible set), the SP data describe potential choices in terms of a set of constructed measures of combinatorial mixes of attributes of real and/or hypothetical alternatives. Some researchers have been aware that extra features in smart cards, like multifunction, interoperability, *etc*, would play an important role in the future development of the smart card ticketing, but the relevant evaluation from the demand angle has not been addressed in the current stage. Therefore, three payment options we use in the RP survey (cash, travel cards and smart cards) are still considered in the SP survey, however, new attributes and variations will be introduced.

- **Determination of attributes and levels in the survey**

Individual travel cost as a monetary variable was primarily used in most smart card evaluation studies. Especially in the smart card ticketing, discounted fare policy has been widely applied, therefore, introducing individual travel cost is essential to measure valuation of attributes in the both preference surveys. Another benefits may be measured is how much quicker the boarding time could be under smart card applications, comparing with other fare payment methods.

In previous studies, such as "before" and "after" analysis, only may those attributes, which can be found in "before" and "after" both, be measured. This method did not account for those new features (particularly for smart cards). In users' preference studies, some new features or new variations were addressed, such as multifunction, adding value options, geographic areas covered, *etc.*, but only basic statistical results (e.g., distribution of different responses) were presented in these studies. The further studies, such as valuation of these attributes, behavioural analysis, demand forecasting, have not been clearly explained. So it is necessary to take into account these new features or new variations based on existing attributes in the SP survey to examine respondents' choice behaviour under hypothetical situations. Regarding the detailed attributes and levels in previous applications, Table 2.3 lists the key features of smart cards, which can be used in the questionnaire design of this research. Regarding criteria of each selected attributes and levels used in this research, Section 5.3.2, 5.4.2 and 5.4.3 discuss in details.

- **Survey methods**

In previous studies, the following survey methods have been used. Pros and cons of different survey methods, plus Chinese unique situation, are helpful to determine a suitable method for this research.

- On board survey and at bus stop survey once were adopted in previous evaluation studies when investigating users' or operators' preference and assessment. The advantage of this survey method is that it can directly and easily capture the survey population (PT users or operators) to collect data related to public transport. In addition, cost of on board survey is relatively low, compared with face to face interview. The drawback of this method is possibly low response rate and poor data quality, because of the relative uncomfortable survey environment.
- Face-to-face interview was preferred when asking about public transport operators' point of view in some literatures, but the sample size was limited to a relatively small scale (e.g., less than 100) when considering the higher survey cost of face-to-face interview method. But the advantage of face-to-face interview is higher response rate and better data quality than any other methods (on board survey, mail-back survey, *etc*).
- Mail back survey once was considered to collect preference data in some studies. The advantage of mail back survey is that it can guarantee good data quality,

because respondents can take time to consider answering questions, unlike on board survey. But the disadvantage of this method may be the possibly low response rate.

For this research, because of the low survey budget, first of all face-to-face interview is not preferred. Secondly, due to the poor performance of mail back survey in some market research in China previously (very low response rate), this method also will not be considered. Finally, on board survey is chosen, because it is suitable to target the survey population (PT users), to carry out a large-scale data collection when the survey budget is not sufficient and high response rate is required.

- **Data analysis (evaluation) methods**

To measure discrete choice data from RP and SP survey and forecast respondents' choice behaviour, logit models have been widely used in transport studies and proven to be successful based on the random utility theory (Ben-Akiva and Lerman., 1985). On the other hand, the recognition of the relative strengths and weaknesses of both RP and SP data suggested that the joint utilisation of both data should enrich the modelling activity and further our understanding of choice behaviour. Therefore, this research will explore the feasibility of the combination of RP and SP data in the data analysis stage.

3.4. Summary

To sum up, during the course of the literature review of chapter 3, we have seen evaluation studies of benefits and effectiveness of smart cards from four different folds: "before" and "after" analysis; user's preference studies; operator's perception studies; and cost-benefit analysis. From the literature review, the primary task is to help determine the research objectives and relevant evaluation methodology:

First of all, among these previous studies, the key issue is that there are not many studies related to measure features of smart cards and compare the smart card ticketing with other conventional payment methods from individuals' point of view. Therefore, measuring users' preference and carrying out demand forecast is proposed as the objective of this research, because as measures for evaluating benefits of smart cards, user demand and valuation of attribute have not been examined clearly in previous studies. Hence, benefits and effectiveness of smart cards are to be assessed by users' demand and choice behaviour. In order to forecast the market share of different fare payment methods and identify individual choice behaviour, and get the best features in smart card applications, it is necessary to evaluate benefits of PT fare payment from user's angle.

Secondly, in this research because the benefits and effectiveness of smart cards need to be measured by some attributes (features) related to fare payment itself, another implication of this chapter is to determine fare payment attributes as indicators to reflect benefits to PT users. In Chapter 2, some key features of smart cards have been short listed, and the relevant evaluation studies also used them to evaluate the performance of smart cards, such as boarding time saving, travel cost, etc. Moreover, some payment features (e.g., payment convenience as a result of quicker boarding time, multiple top-up options, a variety of ticket packages in smart cards), called 'soft benefits', are also used in the later modelling analysis, because these soft benefits could be measured by passengers' preference/perceptions toward different payment options.

Therefore, thirdly, for the purpose of this research, preference survey is determined to carry out in the data collection stage. Two types of preference surveys are to be used: revealed preference and stated preference. It is expected that benefits and effectiveness of smart cards can be presented by respondents' decision making of fare payment options (existing and hypothetical situations), in which payment attributes are combined together as well as respondents' socio-economic background.

Finally, in order to analyse preference data, it is found that among techniques to evaluate benefits of smart cards, user demand models (based on discrete choices) are more suitable for this purpose. Eventually, benefits of smart cards to PT users can be explained by forecast market shares of payment options when the future situations would change, valuation of attributes of fare payments in terms of monetary value, etc.

Therefore, based on the implications from the literature review in this chapter, the research methodology of this thesis is generated and addressed in details in the next chapter.

Chapter 4
Research Methodology

4.1. Introduction

Through reviewing smart card applications in Chapter 2 and evaluation studies in Chapter 3, the following issues from previous work can be determined. The research context in Chapter 2 and these issues of evaluation studies also can help to generate the relevant research methodology of this thesis:

- Features of smart card ticketing application: smart card ticketing is a novel fare payment application in public transport systems. Some new features have been gradually introduced in recent years. When reviewing smart card applications throughout the world, these features, such as multifunction, wider geographic areas covered, multiple adding value options, etc, are regarded as the future direction of the smart card development. However, how PT users would perceive these new features and how these features would benefit users have not been explained in previous studies.
- Evaluation techniques: most previous studies focused on PT users' (or operators') current perceptions (attitudes) about the smart card ticketing. Basic statistical analysis was primarily used. But users' demand based on attributes (features) related to smart card applications can directly explain benefits of the smart card ticketing. Therefore this research looks at individual choice behaviours so as to capture people's willingness to pay, value of time (or attributes) and forecast market share of smart card applications. Regarding the choice behavioural analysis, previous evaluation studies have not provided a systematic methodology, but logit models have suggested the feasibility and suitability in modelling discrete choice problems (Ben-Akiva and Lerman, 1985).
- Data collection methods: in order to investigate PT users' preference, particularly based on features which have not been introduced in smart card applications, stated preference (SP) technique has been proved its success by past studies. Compared with the data collection methods in previous studies, which collected users' actual behaviour of smart card use, SP data is more suitable for forecasting choice behaviour in a long term, providing a big range of variation (Louviere, Hensher and Swait, 2000).

Therefore, according to the gaps between previous studies and this research, the relevant research objectives can be determined. In this research, the benefits of smart cards are evaluated from the PT users' side (demand side), because users' perceptions can directly reflect whether the smart card ticketing has advantages on improving service quality of PT, compared with other conventional payment options (i.e., cash and travel cards). In order to capture benefits of smart cards, features of smart card ticketing are introduced in the data collection. Through evaluating these features (attributes), the importance of attributes can be obtained. Furthermore, this result can feed back to the relevant policy making (e.g., where and how the smart card ticketing would be enhanced in the future development).

The overall research hypothesis of this thesis is that benefits of smart cards to PT users can be reflected by their own choices among different fare payment options (alternatives). The reaction between smart cards and conventional payment methods (cash and travel cards) was taken into account, because the use of smart card ticketing may be influenced by other payment methods. So, this research assumes that people make their decisions (which payment method they used or would use) after considering three alternatives, cash, travel cards and smart cards. Each alternative is assigned a utility. The relative utility is related to a combination of a series of attributes of fare payment alternative. The fare payment method that a respondent preferred must have a maximum relative utility.

Therefore, based on the hypothesis of this research, in the data collection stage, first of all, respondents' actual choice behaviours among three payment options should be investigated. However, because in China, smart card applications are still in the early stage, features and some extra services based on these features have not been fully implemented in reality, users' preferences based on hypothetical trade-off situations also should be examined. Revealed preference (RP) and stated preference (SP) methods are suitable for collecting different individual preference data aforementioned.

Regarding the evaluation techniques, discrete choice models have been widely used to analyse preference data. Among discrete choice models, logit models are proposed to employ in this research. The outputs of logit models can be used to explain respondents' perceptions about fare payment attributes, such as estimated coefficients, value of attributes, elasticities, etc. Another aspect in the data analysis is to explore some other techniques (fuzzy logic-FL and artificial neural network-ANN) to model preference data. The motivation of using FL and ANN techniques is to make use of their abilities of modelling the non-linearity and uncertainty of the preference data to improve the model performance.

According to the discussion above, the relevant methodology is generated to have an insight into the evaluation of smart card benefits to PT users. First of all, a research design (framework) is outlined in Section 4.2. Following Section 4.2, each stage of the research methodology is discussed respectively. In Section 4.3, reasons of using the revealed and stated preference survey particularly for Chinese context and their pros and cons are discussed. The data analysis stage is split into two parts: Section 4.4 addresses the logit model analysis for the RP and SP data; Section 4.5 introduces methodologies of the fuzzy logic (FL) and artificial neural network (ANN) as an extension of the standard discrete choice model to improve the performance of the model. Finally, the model applications on analysing fare payment choice behaviour are discussed in Section 4.6.

4.2. Research Framework

Prior to the discussion of the whole research work for this thesis, a research framework needs to be outlined as follows (See Figure 4.1).

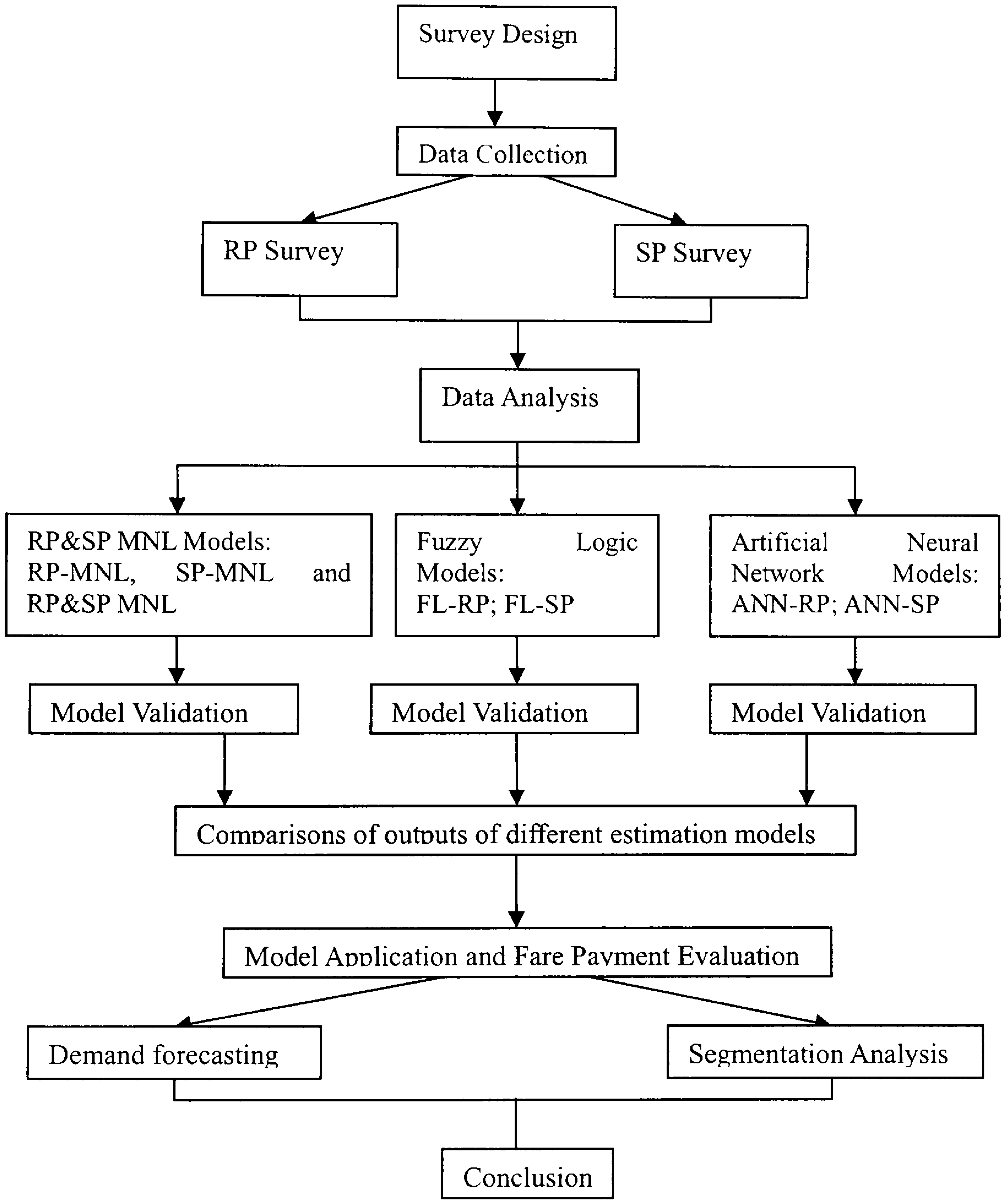


Figure 4.1 Research Design of this Thesis

As can be seen in Figure 4.1, this research is based on the survey of public transport users to collect individual preference data about fare payment methods. Several reasons can explain why we carry out the research based on the public transport passenger survey. First of all, very few previous studies have been done on PT users' preference analysis of fare payment choices in Chinese context. Particularly when alternative fare payment situations may change, there is no evidence about passengers' preference and valuation of attributes of fare payment alternatives under such stated choice profiles. So, it is necessary to carry out individual preference survey among public transport users to collect their choice behavioural response. Secondly, so far in most Chinese cities applying smart card applications, evidences of the market share of three payment methods (i.e., cash, travel cards and smart cards) have not been available, therefore, through investigating passengers' choices (actual and hypothetical if some situations would change) based on questionnaires, it is helpful to estimate the market share for the current and future fare payment applications.

After the survey design, the data collection is carried out by means of questionnaire. Questionnaires have advantages over some other types of surveys in that they are cheap, do not require as much effort from the surveyor as verbal or telephone surveys, and often have standardised answers that make it simple to compile data (Wikipedia, 2006). In the data collection, we split the survey into two independent parts: revealed preference survey (RP) and stated preference (SP) survey, because if the RP and SP surveys were combined, the questionnaire became very long and some negative response from respondents could impact the data quality.

The RP survey is used to collect people's actual choices based on existing fare payment applications. The attraction of using RP data is because they can tell what people actually do. However, RP data cannot look at new attributes. The SP survey investigates stated choice behaviour based on hypothetical situations, particularly for the smart card ticketing, some new features could be added on the current application. Compared with RP data, SP data can generate more data per person. Moreover, the design can control trade offs so that better quality data than with RP data where RP does cover attributes of interests. Therefore, the RP and SP surveys are conducted to collect different types of data.

In the data analysis stage, three actions are taken for modelling the preference data we obtained in the data collection, including MNL model analysis, fuzzy logic analysis and artificial neural network analysis. In the MNL model analysis, MNL models for the RP and SP data are firstly used. Besides the RP and SP MNL models, the data enrichment for the RP and SP data is carried out by MNL model. The second and third part of the data analysis is to use fuzzy logic (FL) and artificial neural network (ANN) models to explore the improvement on model performance and forecasting ability with discrete choice data. Feasibility and contribution of FL and ANN techniques to modelling discrete choice data have been discussed in previous studies of the literature review. The aim of using FL and ANN in this research is to account for uncertainty and linguistic ambiguity of decision-making of human being, and non-linear relationship between inputs (attributes) and outputs (payment choices). In addition to the separate estimation in different models, the comparisons of outcomes among these different data analysis models (including, MNL-RP, MNL-SP, MNL-RP&SP, FL-RP, FL-SP, ANN-RP and ANN-SP models) are made to present the performance on forecasting ability. Meanwhile, the model validation (about 10% of total data retained for validation) is carried out in the data analysis to check the choice behavioural validity by using the models we develop.

Finally, the model application for analysing fare payment choice behaviour in this research are divided into two sections: (1) Segmentation analysis by using different demographic data to support forecasting market segments for the future population and to examine the choice behavioural heterogeneity; (2) Demand forecasting for fare payment choices and measuring choice behaviour, including valuation of attribute and elasticities analysis. Valuation of attribute can reflect people's willingness to pay for different payment services. Through analysing fare elasticities, respondents' fare payment demand with respect to changes of their travel cost can be identified.

4.3. Revealed and Stated Preference Survey

4.3.1 Revealed Preference (RP) Survey

In this research, the survey design consists of two parts: revealed preference (RP) survey and stated preference (SP) survey. By definition, RP data describe only those alternatives that exist. In this RP survey, the existing fare payment options are involved: cash, travel cards and smart cards, according to the current Chinese situation.

The primary advantage of RP techniques is the reliance on actual choices (what people actually do in reality), avoiding the potential problems associated with hypothetical response such as strategic responses or a failure to properly consider behavioural constraints (Kroes and Sheldon, 1988). Hence, RP data are particularly well suitable to short-term forecasting of small departure from the current state of fare payment applications (Louviere, Hensher and Swait, 2000). However, inflexibility of RP data is the main disadvantage if we wish to forecast to a market other than the historical one, such characteristics on RP data: limited variations and situations, could make the RP data inflexible and inappropriate. Shifts of some variations of attributes under perceived situations call for other kinds of data sources.

This RP survey is proposed to be based on self-completion questionnaire, because in the RP survey, a large survey sample is required, and compared with other survey types (mail back, telephone, face-to-face interview, *etc*), self-completion questionnaire is feasible when the survey budget and member of surveyors is limited. The questionnaire is sent to respondents according to their primary fare payment used (i.e. cash, travel cards and smart cards). Meanwhile, conditional questions about fare payment methods the respondents would use also are included, so the respondents are asked about the travel costs and some other features of the un-chosen as well as the chosen fare payment options. Before they commence to answer the RP questionnaire, the respondents are required to report their user types (based on fare payment method they primarily used in the last month) and then questionnaires suitable for their own user types personally are given. As to the details about the RP survey design, we discuss in Chapter 5. This chapter is regarded as an outline for the whole research design.

4.3.2 Stated Preference (SP) Survey

In addition to investigating passenger actual choice behaviour, their responses and preferences based on hypothetical situations also is to be taken into account in this research, because fare payment applications, particularly for the smart card ticketing, are changing to enhance the service quality of public transport systems in China. For example, most cities in China propose to terminate travel card payment, replaced by the smart card ticketing thoroughly. For smart card applications, many attributes/or new variations, which have been implemented successfully in other countries and areas, could be added in the current applications in China (e.g. more extra services, interoperation of different cities' PT systems, *etc*).

Compared with the RP survey, the SP method has presented advantages in its use, especially on predicting responses to changes, controlling correlation and variation between attributes, *etc* (Wardman, 1988). For this research context, the advantages of the SP survey to investigate PT users' preferences can be addressed as follows:

The SP data can capture a wider and broader array of preference-driven behaviours on fare payment choices than the RP survey in this research. Variations of attributes of the existing fare payment alternatives are quite limited, while SP data are particularly rich in attribute trade-off information because wider attribute ranges can be built into experiments, allowing model estimates from SP data more robust than the RP data (Swait, Louviere and Williams, 1994). But in China or in the regions where smart card projects (or some other fare payment applications) have been and will be carried out, users' preferences have not yet been evaluated and explained. Thus through introducing the SP survey in this study, users' preferences under hypothetical payment situations that would be implemented can be identified.

The RP data may be suitable for predicting well in an existing market. However, the long-term response changes of PT users can be forecasted with providing new attributes or new variations in the SP data. Some new features would be proposed to add in the current Chinese smart card applications, such as more multifunction, more top-up/purchase options, a variety of fare packages, *etc*. All of these could not be examined in the RP survey. In addition, collinearity between two explanatory variables (e.g., travel time and travel cost), and lack of data per person due to non-response and unavailability of alternative being surveyed are two other problems in RP data.

However, the principal drawback of the SP survey is the potential inconsistency between respondents' intention based on hypothetical situations and their actual behaviour when they really happened. That is to say, individuals' stated preferences may not correspond closely to their actual preferences (Wardman, 1988). Faced with the drawback of SP methods, some researchers suggested that the use of SP methods in conjunction with RP methods offers an attractive solution which avoids the problem of stated intention/revealed behaviour (Kores and Sheldon, 1988).

Among three SP techniques: ranking data; rating data and choice-based experiments, binary choice-based experiment is more suitable for this research, because:

- More than three attributes are considered for each alternatives in this research and a great number of payment situations for each single alternatives may cause the respondents to become fatigue, if ranking response is used;
- For rating responses, although such responses can provide the richest type of response data by using the strength of preferences, in pair-wise choice situation such response scale could be repetitive among different choice situations. Moreover, for rating response, in order to secure the quality data, we must assume that respondents can provide a reliable and valid measure of their degree of preference for each option. Thirdly, for rating data, except those most and least preferred alternatives, it seems difficult that researchers interpret difference on preference about those alternatives rated in the middle level. In this research, we are more concerned as to the reliability of such likelihood of responses, so choice-based experiment is used by requiring respondents to state a preference for one option over the competing alternative.

Considering the existing payment applications in China, three alternatives (cash, travel cards and smart cards) are also used in the SP experiments, but new attributes and variations are designed for these payment means. Because this research is aiming to the analysis of fare payment choice behaviour, particularly for the smart card ticketing, the smart card is separated into 'Pay monthly cards' and 'Pay as you go cards' to explain any changes on choice behaviour in more details. As to the design of the SP survey, including the attributes, levels and binary choice games we use, Chapter 5 gives the detailed information.

4.3.3 Pilot Survey

Before finalising the survey design, a pilot survey (including the RP and SP survey) needs to be carried out to examine the validation of the questionnaire design. The aim to carry out the pilot survey specific for this research can be explained as follows:

(1) The pilot survey can provide reliable guidance on the way respondents will actually respond. Especially for the SP survey with new attributes and wider variations, because they have not been implemented in the real life, respondents' understanding, familiarity and perception toward hypothetical trading-off situations may be a potential issue, which could influence their responses. In the pilot survey, it is helpful to target problems due to these new features so as to minimise misunderstanding in the main survey.

(2) The pilot survey not only can test the suitability of the SP experimental design, but also the adequacy of the way in which it is presented;

(3) The pilot survey will also highlight practical management issues during the main survey, such as individual survey duration, likely response rates that will be achieved and the proficiency of any respondent taking part in.

4.4. Logit Model Analysis

4.4.1. MNL Models

A widely adopted approach for discrete choice analysis is the logit model (Ben-Akiva and Lerman, 1985), which is used for modelling a choice from a set of mutually exclusive and exhaustive alternatives. It is based on the Random Utility Theory (RUT) by McFadden (1981), which is assumed that the decision-maker chooses the alternative with the highest utility among the set of alternatives. The utility of an alternative is determined by a utility function, consisting of independent attributes of the alternative concerned and the relevant parameters. The RUT considers that the analyst does not include the whole range of factors influencing the choice and introduces a random error to account for them. The random aspect is represented by decomposing the utility into two components: systematic term and error term, the former one can be observed but the later one indicates all unknown factors could influence decision makers' choices. Therefore the individual relative utility function can be written as:

$$U_{in} = V_{in} + \varepsilon_{in} = \sum \beta_{imn} X_{imn} + \varepsilon_{in} \qquad (4.1)$$

where: U_{in}: the utility of alternative i for individual n;

V_{in}: systematic term of attributes related to alternative i for individual n;

β_{imn}: coefficients to be estimated;

X_{imn}: deterministic variables (attributes);

ε_{in}: a random disturbance term.

The key assumption in the RUT is that individual *n* will choose alternative *i* if and only if (M is a choice set):

$$U_{in} > U_{jn} \qquad i \neq j,\ i \text{ and } j \in M \qquad (4.2)$$

From equation (4.1) and (4.2), alternative i is chosen if

$$(V_{in} + \varepsilon_{in}) > (V_{jn} + \varepsilon_{jn}) \tag{4.3}$$

Rearranging to place the observable and unobservable terms together yields:

$$(V_{in} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{in}) \tag{4.4}$$

Because the analyst doest not observe $(\varepsilon_{jn} - \varepsilon_{in})$, hence cannot determine exactly if $(V_{in} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{in})$. One can only make statements about choice outcomes up to a probability of occurrence. Thus, the analyst has to calculate the probability that $(\varepsilon_{jn} - \varepsilon_{in})$ will be less than $(V_{in} - V_{jn})$. This leads to the following equations:

$$P_{in} = \mathrm{Prob}\{(\varepsilon_{jn} - \varepsilon_{in}) < (V_{in} - V_{jn})\} \tag{4.5}$$

A common assumption (the reasons for making it are largely practical) is to assume that the error terms (ε) are independently and identically distributed (IID) with the Weibull (or called Gumbel) distribution. Therefore, Equation (4.5) can be written as:

$$P_{in} \mid \varepsilon_{in} = \prod_{i \neq j} e^{-e^{-(\varepsilon_{jn} + V_{in} - V_{jn})}} \tag{4.6}$$

ε_{in} is not given, and so the choice probability is the integral of $P_{in}|\varepsilon_{in}$ over all values of ε_{in} weighted by its density:

$$P_{in} = \int (\prod_{i \neq j} e^{-e^{-(\varepsilon_{jn} + V_{in} - V_{jn})}}) e^{-\varepsilon_{in}} e^{-e^{\varepsilon in}} d\varepsilon_{in} \tag{4.7}$$

Finally the multinomial logit (MNL) model can be written as:

$$P_{in} = \frac{\exp^{V_{in}}}{\sum_{1}^{j} \exp^{V_{jn}}} = \frac{\exp^{\sum \beta_{imn} X_{imn}}}{\sum_{1}^{j} \exp^{\sum \beta_{imn} X_{imn}}} \tag{4.8}$$

Because Equation (4.8) represents the individual choice probability, it can be aggregated and used to forecast market share of different alternatives, for example by using sample enumeration method, forecasted market shares (the average choice probability) can be obtained (DfT, 2004).

The advantages of the MNL are that it is relatively easy to estimate, the coefficients are easy to interpret and the forecasts are generally quite robust. However, the limitation of MNL is that it assumes that the choice options are independent and therefore fails to take account of correlation between alternatives (Bierlaire, 1997; Whelan, 2003). An important property of MNL, the Independence from Irrelevant Alternatives (IIA), causes this limitation, which means that for any two alternatives, the ratio of their choice probabilities is unaffected by the presence or absence of any other alternatives in the choice set.

In this research, because responses in the RP and SP survey are presented by choice-based data, firstly MNL models are used to estimate the RP and SP data respectively. The purpose of using MNL models for separately analysing the RP and SP data is to obtain parameters in pure RP and SP models and prepare for the data enrichment in the later stage.

4.4.2 Hierarchical Logit Model

In MNL model, we assume IID (independently and identically distributed) Gumbel. The IID assumption implies that cross-substitution between pairs of alternatives are equal and unaffected by the presence/absence of other alternatives (Ben-Akiva and Lerman 1985). However, this assumption may give rise to problems when alternatives are not independent (i.e. cash payment and cashless payment). Therefore after the standard MNL model for analysing the RP data with three alternatives, a hierarchical logit model is employed to relax IID assumption associated with the random components of each alternative in standard MNL models. The hierarchical logit (HL) model is to partition choice sets so that richer substitution patterns can be accommodated to reflect differential degrees of similarity and dissimilarity (Louviere, Hensher and Swait, 2000).

For the SP data, because separate MNL models are used for four different SP games, data-merging for four different SP data sets also calls for hierarchical logit models to achieve the data combination. And through combining all SP data sets, choice behavioural homogeneity with crossing different SP games can be identified.

Hierarchical Logit Model for the RP Data

The hierarchical structure for the RP data is characterised by grouping some alternatives, which are similar with each other in a nest. Each nest, in turn, is represented by a composite alternative which competes with the others available to the individual. The detailed hierarchical structure is illustrated in Figure 4.2 and 4.3 for two different grouping schemes.

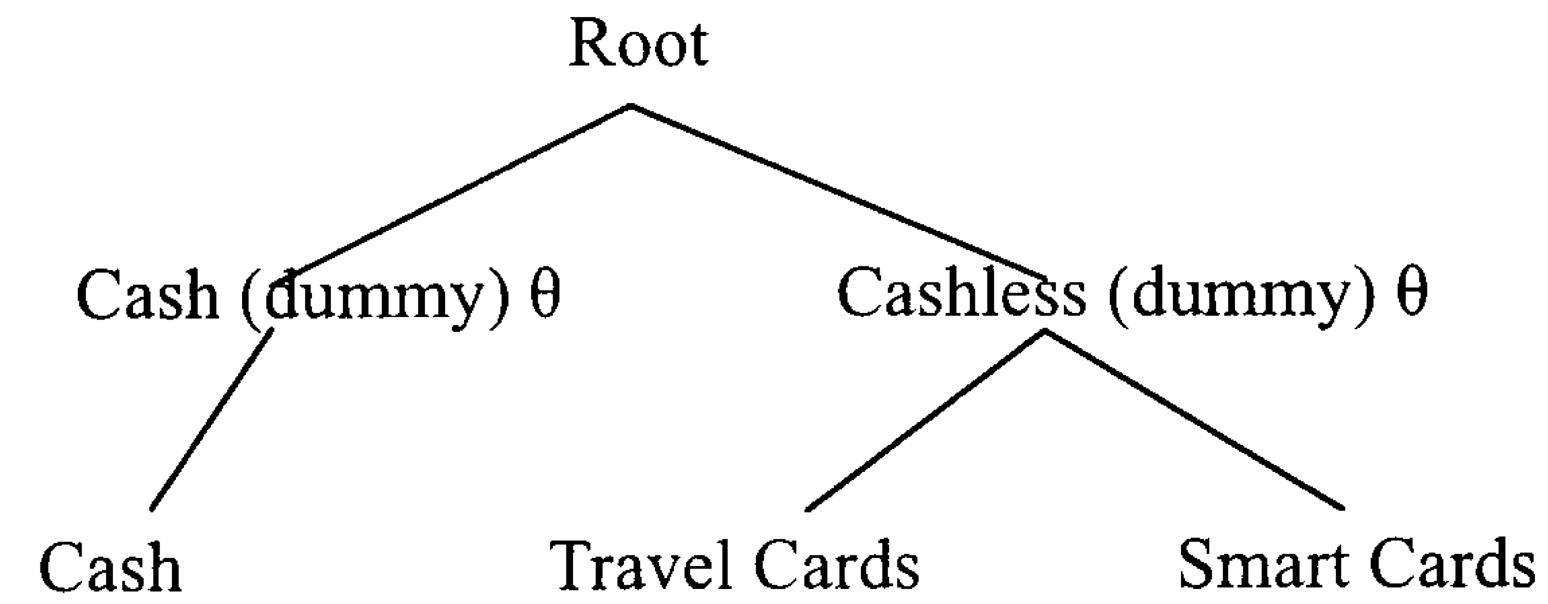


Figure 4.2 Hierarchical Structure of the RP Alternatives (1)

As can be seen in Figure 4.2, first of all, in the upper level, two dummy alternatives, cash and cashless payment methods, are separated into two different sub-nests according to their different payment features (cash is paid for each trip, which cashless payment like travel cards and smart cards is pre-paid/stored value). In the sub-nest of the cashless payment methods, travel cards and smart cards are set in the lower level. In the sub-nest of cash, only cash payment is allocated in the lower level. The reason to do this is to avoid the model ending up being of the 'non-standardised' class (i.e., it does not satisfy the requirement that when adding a constant to each utility the choice probabilities do not change) (Ortuzar and Willumsen, 2001). Therefore, in this case, cash is specified as belonging to a single-element nest (Bradley and Daly, 1997). And in the nest structure, all alternatives in two different sub-nests are scaled by the same factor of θ.

A key feature of the HL model is its flexibility in designing the hierarchical structure. Except the HL structure as shown in Figure 4.2, another structure can be proposed as Figure 4.3. Figure 4.3 is structured according to the payment technology. As a new payment application, the smart card ticketing is set in a sub-nest with single option, and other two payment methods, cash and travel cards, are put in another sub-nest, called conventional payment.

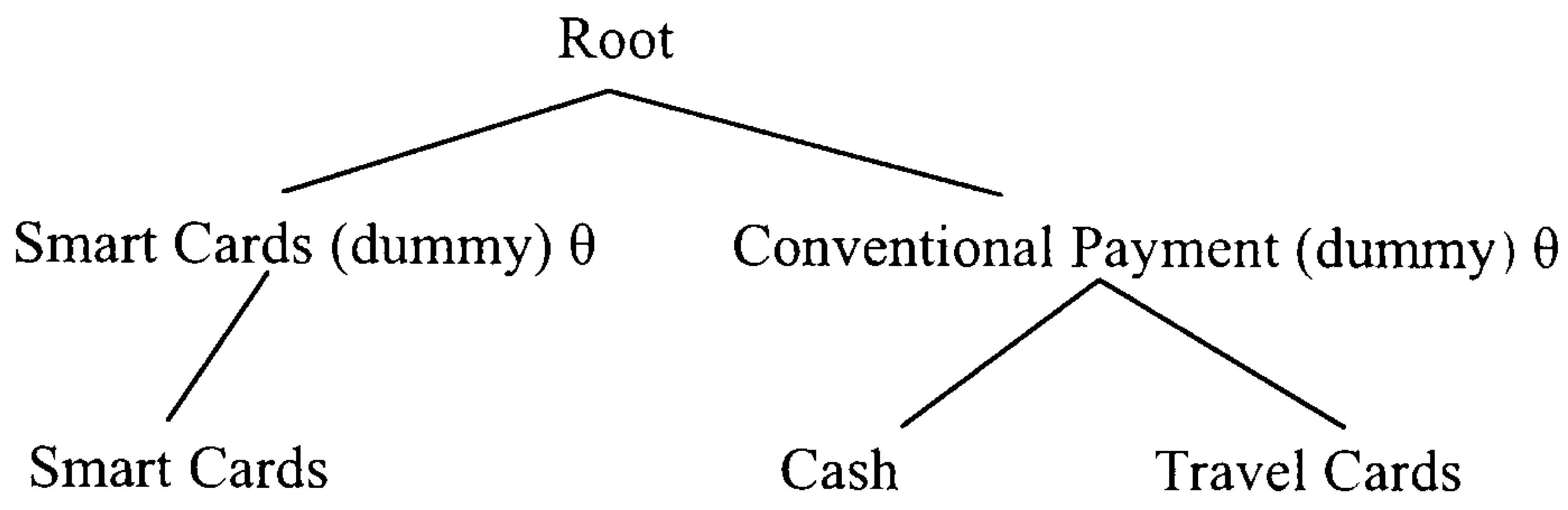


Figure 4.3 Hierarchical Structure of the RP Alternatives (2)

Both hierarchical structures above are tested and compared in Chapter 7: Data Analysis.

Hierarchical Logit Model for the SP Data

The hierarchical logit model for the SP data is used for combining different SP data sets, different from the RP HL model, which aims to relax IID assumption in the model estimation.

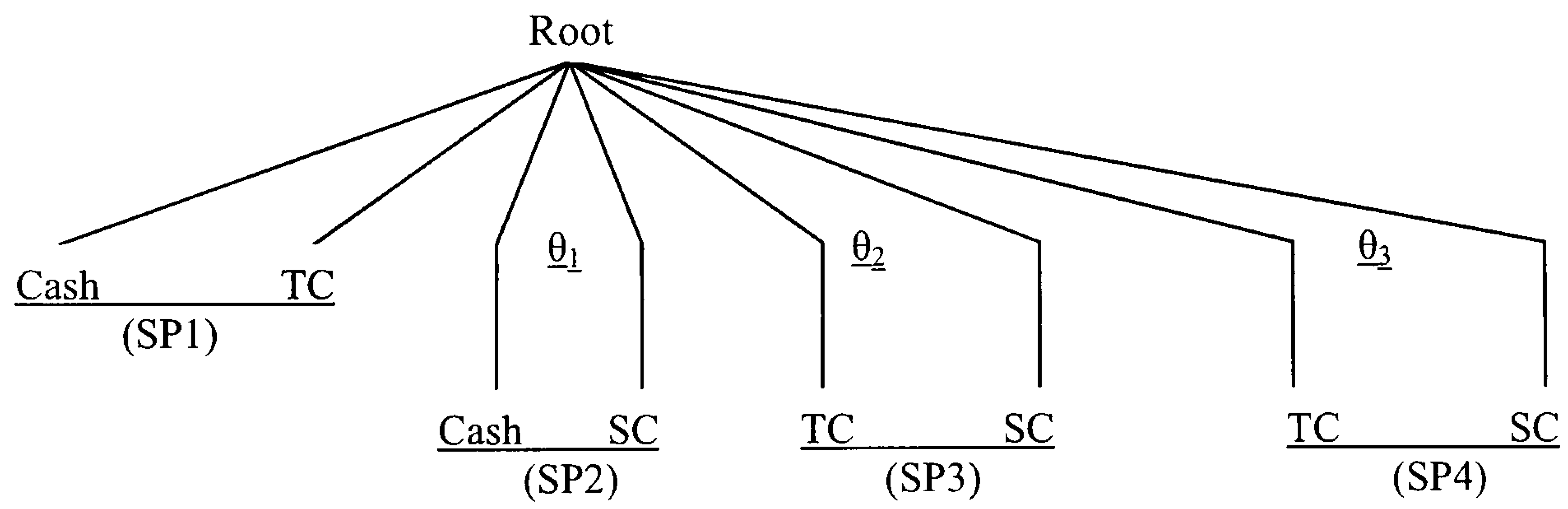


Figure 4.4 Hierarchical Structure of SP Data Combination

The hierarchical structure for combining four SP data sets is illustrated Figure 4.4. As can be seen in Figure 4.4, two-level of the hierarchical structure is used. Alternatives in SP-1 are put on the upper level, directly linking with the root. And the rest six alternatives for three different SP games (SP-2, 3 and 4) are allocated in the lower level, by using scale factors (θ_1, θ_2 and θ_3) to link with the upper level. But it is worth noting that although eight alternatives exist in this hierarchical structure, actually they still are based on three alternatives: cash, travel cards and smart cards. They are distinguished only due to the presence of alternatives in the SP survey. We present only one SP game with binary choice situations each time to each individual, therefore, the one we present to the respondent is available to him/her and the rest three games are regarded as unavailable.

4.4.3. Joint Analysis with RP and SP Data

Modelling choice behaviours with data from multiple sources (RP and SP, for example) has received attention in recent years as an alternative way to cope with weaknesses resulting from using a single data set. In this research, such data enrichment is conducted, because the RP data of fare payment choices only contains information about the equilibrium but new variables and alternatives can not be involved. However, the SP data can cover a wider range of attributes, but it is needed to rescale SP data into real world behaviour.

Introduction to Scale Factor

In MNL models, we assume that the random residual ε is distributed IID Gumbel, such that choice probability can be written as:

$$P_{ij} = \frac{\exp(\beta V_{ij})}{\sum \exp(\beta V_{mj})} \tag{4.9}$$

In Equation (4.9), β is related to the common standard deviation of the Gumbel variate by:

$$\beta^2 = \pi^2/6\sigma^2 \tag{4.10}$$

The 'true' estimate of the utility in the RP and SP data results can be linked to the current estimates of utilities (we assume that before the data combination, we do not know which data source is more reliable) by the parameter β. Although the value of β is taken as 1.0 in practice, for two different data sources (RP and SP), the difference on the single standard deviation σ would result in the difference of βs in these two data sources.

Equation (4.10) can also explain why it is not correct to postulate the same error distribution for estimating and forecasting for the combined data sources. This produces 'scale' differences on the parameters between different respondents if such equality is improperly assumed we might finish estimating pseudo utilities instead of 'true' utilities. To avoid this problem, we need to adjust the SP data to actual behaviour, exploiting the advantages of the RP data in this sense, and estimating the parameters jointly. In addition, through jointly analysing the RP and SP data, the validation of different data sources can also be examined. By introducing 'scale factor', how the RP and SP data are reliable can be examined and 'true' utilities can be estimated.

In order to combine the two data sources, each with independent choice outcomes, allowance must be made for their different scaling properties. The approach uses a full information maximum likelihood estimation procedure of the hierarchical logit form to obtain suitable scale factors to make one or more data sets comparable. The violation of the constant variance condition in the MNL model (alternatively referred to as the independence of irrelevant alternatives property) resulted in the development of the nested (of hierarchical) logit model, which permitted differential variance between levels and/or branches within a level of the nested structure but a common variance within a branch (Hensher 1986, 1991; Borsch-Supan 1986).

For example, the utility maximised by each respondent in the RP context is given by:

$$U_{i,n}^{RP} = \sum\nolimits_k \alpha_k x_{n,ik}^{RP} + \sum\nolimits_l \beta_l y_{n.il}^{RP} + \varepsilon_{n,i}^{RP} \tag{4.11}$$

The utility maximised by each respondent in the SP context is given by:

$$U_{i,n}^{SP} = \sum\nolimits_k \alpha_k x_{n,ik}^{SP} + \sum\nolimits_m \gamma_m z_{n.im}^{SP} + \varepsilon_{n,i}^{SP} \tag{4.12}$$

where i and n indicate the decision maker and alternative, $x_{n,ik}^{RP}$ and $x_{n,ik}^{SP}$ are generic variables for all payment methods, $y_{n,il}^{RP}$ and $y_{n,im}^{SP}$ are lth and mth alternative-specific terms, $\varepsilon_{n,i}^{RP}$ and $\varepsilon_{n,i}^{SP}$ are error terms, and α, β and γ are the parameters to be estimated.

In this case, RP data constitute the primary set, since these data capture the actual behaviour of the individuals, and SP data constitute the secondary set. A framework has been developed by Ben-Akiva and Morikawa (1990), which postulates the difference between the errors in RP and SP. The detailed function can be written as follows:

$$\sigma_{RP}^2 = \theta^2 \sigma_{SP}^2 \tag{4.13}$$

where σ_{RP}^2 and σ_{SP}^2 are the variances of error terms in RP and SP models; θ is an unknown scale coefficient. Therefore, according to the Equation (4.13), Equation (4.12) can be changed to Equation (4.15) for a certain alternative A_i. And then the RP and SP choice models can be altered as follows:

$$U_{i,n}^{RP} = \sum\nolimits_k \alpha_k x_{n,ik}^{RP} + \sum\nolimits_l \beta_l y_{n.il}^{RP} + \varepsilon_{n,i}^{RP} \tag{4.14}$$

$$\theta U_{i,n}^{SP} = \theta(\sum\nolimits_k \alpha_k x_{n,ik}^{SP} + \sum\nolimits_m \gamma_m z_{n.im}^{SP} + \varepsilon_{n,i}^{SP}) \tag{4.15}$$

The scaling of θα in Equation (4.15) is the essential link between the two data models. This estimation problem can be solved by two well founded estimation approaches, sequential estimation (Ben-Akiva and Morikawa 1990) and simultaneous estimation (Bradley and Daly, 1997). Both estimation approaches are suitable for computational packages to analyse discrete choice problem. In the following contents, two methods are discussed in detail, particularly about how the scale factor works in joint RP and SP data.

Sequential Estimation Approach

The sequential estimation approach to model RP and SP data jointly was firstly introduced by Ben-Akiva and Morikawa (1990). The algorithm is as follows (Ortuzar and Willumsen, 2001):

- Estimate the SP choice model in order to obtain the estimated parameters (α_{sp}). Then, define a new utility (NU) expression by using estimated parameters in the SP (α_{sp}) and RP variables (X_{rp}):

  $NU_i = \alpha_{sp1} X_{rp1} + \alpha_{sp2} X_{rp2} + \ldots + \alpha_{spm} X_{rpm}$ (4.16)

  In the new utility function, variables (X_{rpm}) are common variables in both RP and SP data sets.

- Estimate the following RP model with the new utility (NU_i), scale factor (θ) and specific parameters and variables ($\beta_{rpm} Y_{rpm}$) in the RP data:

  $U_i = \theta NU_i + \beta_{rp1} Y_{rp1} + \beta_{rp2} Y_{rp2} + \ldots + \beta_{rpn} Y_{rpn} + \varepsilon_i$ (4.17)

- Multiply the SP data by θ to obtain a modified SP data set. Pool the RP data and the modified SP data and then estimate the two models jointly.

Simultaneous Estimation Approach

Simultaneous approach estimates RP and SP combination model by an artificial tree structure (Bradley and Daly 1991). The artificial nest is constructed to have at least twice as many alternatives as are observed in reality. One subset is labelled as RP alternatives, the other subset as SP alternatives. The indirect utility functions in each case are defined by the V_{rp} and V_{sp} expressions, defined above without scale factor (θ). The RP alternatives are placed just below the "root" of the nest, whereas the SP alternatives are each placed in a single-alternative "nest". For the SP observations, the average indirect utility of each of the "dummy composite" alternatives uses the theoretical basis of the inclusive value concept associated with linking levels in a nested logit model (McFadden 1981) to define the logsum equation of the expected maximum utility (EMU) for SP data (the lower nest options) as:

$EMU_{sp} = \log_e (e^{V1} + e^{V2} + \ldots e^{Vn})$ (4.18)

And then $V^{comp} = \theta EMU_{SP} = \theta \log \sum_{N=1}^{n_{sp}} \exp(V_{N_{SP}})$ (4.19)

where the sum is taken over all of the alternatives in the nest corresponding to the composite alternative and

$V^{SP} = U^{SP} - \eta = \Sigma \alpha_i X^{SP} + \Sigma \beta_j Y^{SP}$ (4.20)

is simply the measured part of the SP utility (η is the error term in the SP utility model). Then, because each nest contains only one alternative in this specification (Equation 4.19), therefore *log* and *exp* cancel out and leave θV in the right side of the equation (4.19), and then we have exactly the form required as long as the values of θ is constrained to be the same for each of the dummy alternatives. This θ is called scale factor for the SP utility.

$V^{COMP} = \theta V^{SP} = \theta (\Sigma \alpha_i X^{SP} + \Sigma \beta_j Y^{SP})$ (4.21)

The hierarchical structure for this research is illustrated as Figure 4.5.

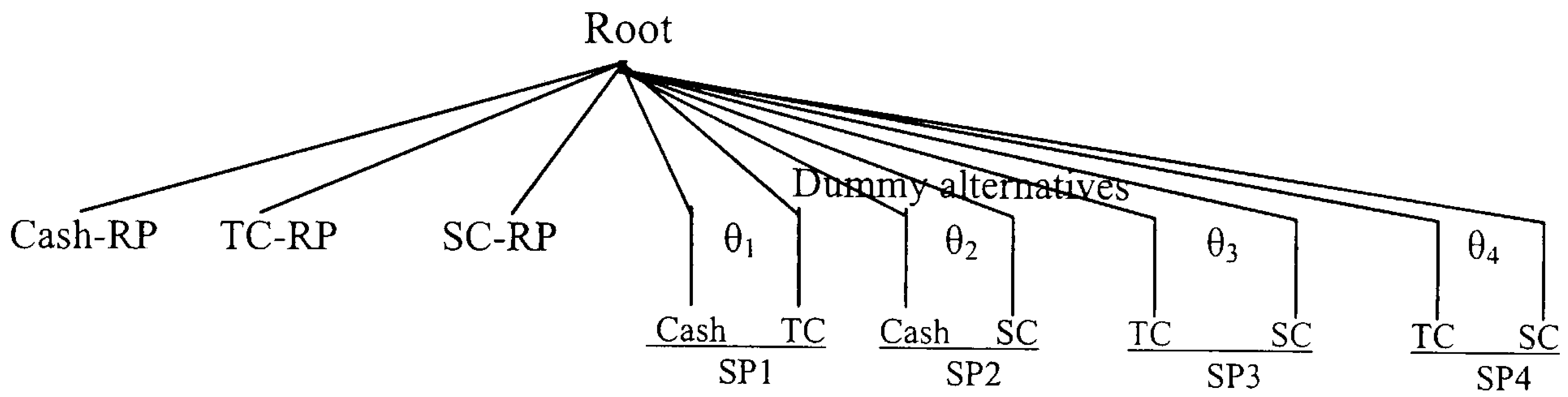


Figure 4.5 Artificial Tree Structure for Joint RP and SP Estimation

It can be seen in Figure 4.5, three RP alternatives: cash, travel cards and smart cards, are put in the upper level. Eight SP alternatives are set in the lower level. In the upper level where the RP alternatives are placed, eight SP dummy alternatives are used to link with eight real SP alternatives in the lower level. Considering four different SP experiments used in the SP survey, four different factors (θ_1, θ_2, θ_3, θ_4) are used to scale these four pair-wise choice experiments.

In the hierarchical structure, for an RP observation, the SP alternatives are set unavailable and the choice is modelled as in a standard logit model. For an SP observation, the RP alternatives are set unavailable and the choice is modelled by a nested logit structure. Since the dummy composite alternatives are placed just below the root of the tree, as are the RP alternatives, a standard estimation procedure will ensure that θ is estimated to obtain uniform variance at this level. In the meantime, this structure means everything being estimated in the SP data is based on units of RP data in the upper nest. In addition, because the individuals are not modelled as choosing from the whole choice set, this artificial construction does not require the usual consistency assumptions for nested logit models that θ should not exceed one, because the individuals are not modelled as choosing from the whole choice set.

According to the discussion above, pros and cons of the sequential estimation and simultaneous estimation can be summarised as follows:

The sequential method deals with different data sources separately, therefore, estimation results for the RP and SP data will not be influenced by each other. When two separate estimations for the RP and SP data are satisfied, the sequential estimation is suggested, because different data sources are not required to combine together when using the sequential estimation.

Compared with the sequential estimation, the main advantage of the simultaneous estimation is that the simultaneous approach estimates the coefficients and scale factors in one model, while the sequential approach needs to calculate the scale first and then multiply another data source by the scale to carry out forecasting. Secondly, results by the sequential estimation are relatively independent as discussed before, but in the simultaneous estimation, results can be viewed as average values, because RP and SP data are used together. Therefore, in case estimation results by one data source is not good enough and another one is satisfied, the simultaneous method can be used to achieve the scaled estimation for both different data sources.

4.4.4 Model Validation

Following the model estimation, the next task in evaluating the performance of an well-estimated model is to test its ability of making predictions. This can be easily carried out with the assistance of the ALOGIT software. To validate MNL models described in this chapter, they are applied to the validation sample of 87 in the RP data and 620 in the SP data, specifically retained for this use (about 10% of total data in the RP and SP data set), using the coefficient values produced during the estimation process. This enables us to compare the models' performance in terms of correctly predicting the observed choices and in terms of recovering the market shares for the PT fare payment method choices, using data that is *unknown* to the models.

For every respondent in the validation data set, a choice probability of these three alternatives in the RP survey and binary choice situations of the SP survey can be obtained through estimated parameters in MNL models. According to individual choice probabilities, the average probabilities of choosing fare payment methods in the validation sample can be calculated. These aggregated choice probabilities (market shares) in the validation sample will be used to compare with the predicted market shares by the main data set (the control sample) to check whether the validity of estimation results can be achieved.

4.5. Fuzzy Logic/Neural Network Techniques in Discrete Choice Analysis

4.5.1. Fuzzy Logic Methods

Introduction to Fuzzy Set and Fuzzy Logic Theory

Fuzzy Decision Framework

Fuzzy set theory was introduced by Zadeh (1965) as a general approach to express the different types of uncertainty inherent in human systems. Zadeh (1973) claims that our ability to make precise and yet significant statements about the behaviour of a system diminishes, as the complexity of this system increases. He proposed the use of fuzzy sets and approximate reasoning methods to model such systems. Fuzzy sets are a generalisation of crisp sets. Members belong to fuzzy sets with a degree of possibility or membership. The grade of membership takes values within the interval [0, 1], and represents the degree to which an element is similar or compatible to the concept represented by the fuzzy set. A fuzzy set A defined on a universe of discourse x can be represented by a set of ordered pairs as

$$A = \{(x, M_X^A(x)) | x \in X\} \quad (4.22)$$

where M_X^A denotes the membership of element x to the fuzzy set A.

Hence, by definition, fuzzy set boundaries are vague, and the transition from member to non-member is gradual rather than abrupt (Klir and Folger, 1988). Furthermore, fuzzy sets can overlap and therefore, an element can belong to a number of fuzzy sets with different degrees of membership.

The decision framework proposed by Lotan and Koutsopoulos is based on the concepts of fuzzy control (Pedrycz, 1989) which has been used successfully in many industrial applications (Sugeno, 1985). The main component of the decision making mechanism is the fuzzy rule base which contains rules of the form:

"**If** ... (*system perceptions*) ... **Then** ... (*preferences towards alternatives*) ..."

describing the preferences of the decision maker given possible perceptions of the system's attributes. The final choice results from the combination of various rules each of which is executed (fired) to a degree reflecting the similarity between the individual's perception and the rule's premise.

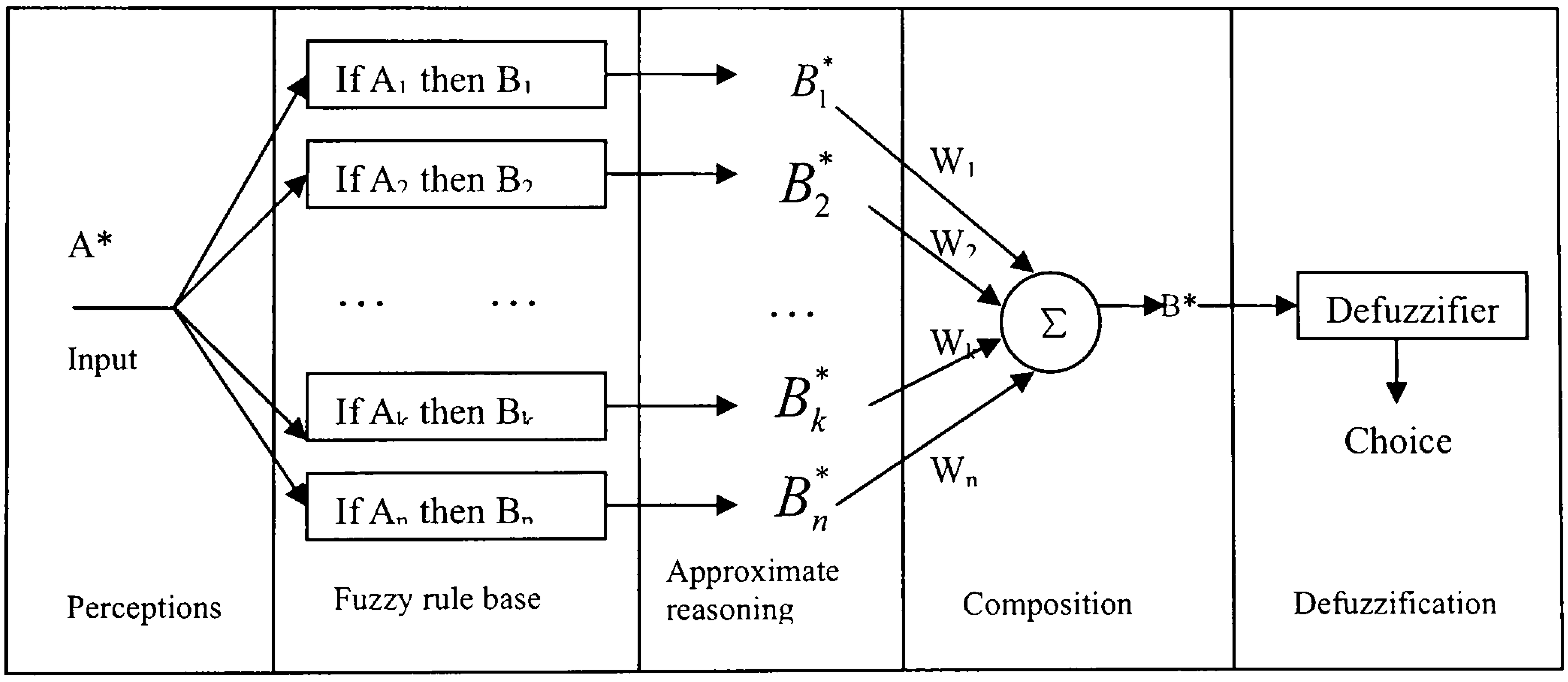


Figure 4.6 the Fuzzy Decision Making Framework

A representation of the general model of fuzzy decision making is illustrated in Figure 4.6. The model inputs, A*, representing the individual's perceptions, are matched with the premises of each rule, k. An inference scheme, called approximate reasoning, is then used to deduce the resulting implications, B_k^*, given the perceptions A*. The rules are processed simultaneously and a composition mechanism aggregates the implications B_k^* to a fuzzy preference B*, expressed in terms of its membership function. The final crisp choice results from the defuzzification of the preference B^*.

Rule Base

In the fuzzy logic system, it is assumed that travellers make their decisions based on simple rules, combined with attributes of alternatives, rather than trying to maximise a complicated utility function. Therefore rules are used to model the decision process and describe attitudes towards selecting an alternative given possible or vague perceptions of the system's attributes. A typical rule k is a statement such as

"**If** (x_1 is A_1...and x_n is A_n), **Then** (y_1 is B^1... and y_m is B^m)."

The variables x_i, i =1,2, ..., n, represent attributes of the alternatives considered in the choice set, y_j, j=1, 2, ..., m, represent the preference towards alternative j, the fuzzy sets A_i, i= 1, 2, ..., n, represent linguistic values of the ith system attribute (e.g. high, medium, very low of travel time) and the fuzzy sets B_j represent linguistic values of the attractiveness of alternative j, such as preferred, indifferent, probably not preferred, *etc*.

Approximate Reasoning

The attributes of an alternative, in a given situation, may not match exactly the premises of the rules in the rule base. The approximate reasoning mechanism allows for modifying the rule consequence according to the actual input:

$$\begin{array}{r} \textit{IF X is A THEN Y is B} \\ \underline{\textit{X is A*}} \\ \textit{Y is B*} \end{array}$$

The membership function or possibility distribution of the resulting output B* is related to the possibility distributions of the linguistic label B in the rule consequence, and the similarity between the input A* and the corresponding label A in the premise of the rule. In general, membership functions are based on the fuzzy set theory. A fuzzy set is a generalisation of a crisp set which allows each element, x, to belong to the set with a certain degree of membership g(x) ($0 \le g \le 1$), where higher g values represent higher degrees of set membership. The concept of membership function allows the definition of sets with vague boundaries where each set has a linguistic label such as: "*HIGH* travel cost", or "*LOW* boarding time". In general fuzzy sets enable to model human oriented systems more realistically by allowing the use of linguistic descriptors, phrases, hedges and modifiers (Zadeh 1973).

A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. In the FL model, an input variable can be separated by several MFs to indicate the different zones that an input value can belong in terms of probability (or called membership degree). Therefore, to any input attributes and outputs in a FL model, the membership functions can give a probability belonging membership degree for an input value or an output value, for example, if travel cost by cash is 60yuan per month, and three categorises are set to distinguish different perceptions of respondents to travel cost: *Low, Moderate* and *High*, under a membership function of travel cost, 60yuan may be *High* cost to the degree of 0.8, 0.2 belonging to *Moderate* category.

The degree of similarity between the input A* and the rule premise, often called the firing of the rule, is calculated using the max–min operator:

$$a = \max_{x \in X} \min(M_{A^*}(x), M_A(x)) \tag{4.23}$$

The membership function of the inferred output B* is derived using the correlation-product encoding scheme proposed by Kosko (1992), which preserves the shape of the membership function of set B (in the consequence of a rule):

$$M_{B^*}(y) = aM_B(y) \tag{4.24}$$

In the case of a rule k with multidimensional premises a_k, the firing strength of the rule k, is defined as:

$$a_k = \prod_{i=1,2,\dots n} a_{ki} \tag{4.25}$$

where a_{ki} is calculated by applying Equation 4.24 to the ith premise of the rule.

The use of the approximate reasoning mechanism reduces the number of rules that are required in the rule base since the premises are only representative labels of the attributes and hence do not have to represent all possible input values.

The Composition Mechanism

The input perceptions A* may overlap with the premise of many rules (since an element x may belong to more than one sets). Hence more than one rules may be fired (processed) simultaneously, each of them to a degree reflecting the similarity between the individual's perceptions and the rule premise. The firing of a rule k, results in the fuzzy preference B_k^{*k} with respect to alternative j. The composition mechanism combines the fuzzy preferences j from all rules k that are activated and calculates the overall fuzzy preference B_k^{*k} of alternative j using the following aggregation operator:

$$M_{B_j^*}(y) = \sum_k M_{B_j^{*k}}(y) = \sum_k \alpha_k M_{B_j^k}(y) \tag{4. 26}$$

Note that if there is no overlap between an input and the premise of a rule then the contribution of that rule to the final attractiveness is 0 since the value of α_k for that rule will be 0.

Defuzzification and Choice

Given the fuzzy set B^*_j that represents the overall preference towards alternative j, for a given set of attributes, the defuzzification mechanism is applied to derive a crisp action (choice). Lotan and Koutsopoulos (1993a, b) used the centre of gravity method to defuzzify the preference set. They assumed that the centre of gravity of the preference set, *centroid*$_j$, is representative of the attractiveness of alternative j. They suggest two methods in order to translate the attractiveness into choice. The deterministic choice rule assumes that the alternative with the highest attractiveness is chosen. The probabilistic rule assumes that the centre of gravity, $centroid_j$, represents the systematic part of the utility of alternative j. Hence the utility of alternative j for individual n is given by: $U_{jn} = centroid_{jn} + \varepsilon$, where ε is an error term. The interpretation of this model is that the centre of gravity captures the overall attractiveness of an alternative, while the random term captures noise in human behaviour, missing rules, *etc*.

Previous Studies of Modelling Discrete Choice Behaviour with FL Technique

In recent years, a great amount of literature have focused on the applications and studies of fuzzy logic methods in travel choice models, such as route choice and mode choice behaviour (Cantarella and Fedele, 2003; Hoogendoorn-Lanser and Hoogendoorn, 1998; Lee et al, 2001; Lotan and Koutshopoulos, 1993; Mizutani and Akiyama, 2000).

Choice Behavioural Analysis

Cantarella and Fedele (2003) introduced fuzzy utility theory to analyse discrete choice behaviour. The authors presented a Fuzzy Utility Theory useful to model user choice behaviour, assuming that the perceived utility for each alternative was modelled through a fuzzy number. First, the maximum perceived utility fuzzy distribution was defined from the perceived utility fuzzy distribution of each alternative; then, the possibility that the perceived utility of an alternative be equal to the maximum value is defined; finally, user choices were deducted from these choice possibilities, through the uncertainty invariance principle. A general framework easily comparable with the RUT was developed. The proposed approach turned out to be consistent with other approaches to simulate uncertainty in choice behaviour through fuzzy numbers which were based on ranking indices: given a set of fuzzy numbers, a crisp number (ranking index) is associated to each of them, so that the fuzzy numbers could linearly be ordered, the most effective indices seeming those proposed by Dubois and Prade (1983). Once the value of the ranking index had been computed for the fuzzy number describing the perceived utility of each alternative, an estimate of the choice share for each alternative was obtained from the values of the ranking index. However, in this research, all preference data were presented by choice-based data. Particularly for the SP survey designed by binary choice situations, it is unnecessary to use ranking data, because SP experiments only required respondents to trade off between two alternatives, therefore, the use of fuzzy logic technique in this research will only focus on modelling choice data, rather than ranking data.

Mizutani and Akiyama (2000a, 2000b and 2001) discussed a series of choice problems by using fuzzy logic theory. They proposed that stochastic models based on the random utility theory such as logit models and fuzzy reasoning models can be combined in order to create advanced models for the estimation of travel behaviour as hybrid models. In their research, logit models with fuzzy logic utility functions were developed to analyse the mode choice behaviour (i.e., car and public transport). The principle of this kind of hybrid model was that a systematic component of random utility was formulated by a fuzzy inference system to replace the conventional linear utility function in a systematic component. But in these literatures, the authors only addressed relatively simple situations with two factors: availability of car and travel time by different modes. Most rules in the rule base were single-factor dominant. The reaction between each attribute, which could present trade-off of respondents, was not discussed in the rule generation. In addition, for complicated scenarios, whether the generation of fuzzy rules for the model could reflect human learning processes is another issue needs to be specified in detail.

Lee *et al* (2001) used fuzzy factors analysis to extract latent factors that would influence choice behaviour. And the utility function of the hybrid discrete choice model was formulised by fuzzy latent factors. The strength of using fuzzy latent factors to model choice behaviour was that it considered individual subjective in their decision making. By the empirical study, it was proved that the hybrid discrete choice model could enhance explanatory power of mode choice behaviour models by effectively incorporating the uncertainty of fuzziness of travellers' subjective data. Also, the authors found travellers had distinct perception for their ordinary choices but indistinct perception for other alternative. So incorporating fuzzy latent factors in choice behaviour models can effectively describe the uncertainty of human behavioural process.

Route Choice

In addition to choice behavioural analysis, another application of fuzzy logic is to analyse route choice problem, which is also a kind of discrete choice problem. Hoogendoorn-Lanser and Hoogendoorn (1998) proposed a new model based on the fuzzy logic paradigm to study a route choice problem for public transport networks with sufficient predictive capabilities to be useful in planning and market assessment.

Among the causes of the poor performance of standard choice models were imprecision and the qualitative nature of travellers' appraisal of observable and unobservable trip attributes. Additionally, random utility models assumed explicit relations for both the systematic utility and the distribution function of the random utility component present. The systematic utility was determined from the attributes of the alternative routes based on the concept of trade-off (e.g., travel time, travel cost, *etc*). Under these circumstances, fuzzy set theory was used to model trip attributes and fuzzy utilities of alternatives, which were equivalent to utilities in traditional choice models, were obtained by using the centre of gravity of the fuzzy utility to defuzzify. The detailed function of the fuzzy utility is listed as follow:

$$U_j = \frac{\sum u_{Kj}(u)du}{\sum(u)du} \tag{4.27}$$

Uj: utility of alternative j

ukj: fuzzy utility of Kth fuzzy decision rule of alternative j

Subsequently respondents chose the alternative with the highest defuzzified utility, corresponding to the RUT for the logit models. Except the fuzzy utilities used in this paper, another aspect worth noting is that of the calibration method to better the model outputs. The application of genetic algorithm to calibrate fuzzy rules significantly increased the percentage of correctly predicted situations from 72%-75% to 84%. However, the calibration process in this paper only focused on the calibration of fuzzy rules, and how the membership functions can influence the FL model and forecasting ability has not been explained, because as an important component in FL models, different membership functions would potentially influence the model outputs.

Lotan and Koutshopoulos (1993) presented a modelling framework for route choice in the presence of information based on concepts from fuzzy set theory, approximate reasoning and fuzzy control. They used fuzzy set theory to model perceptions of network attributes, and traffic information provided by an information system. Rules of the form: "*If... Then...*" were used to model the decision process, and to describe attitudes towards taking a specific route given (possibly vague) perceptions on network attributes. One of advantages of the fuzzy model used is that latent attractiveness of each alternative can be evaluated. In addition, the use of fuzzy sets and approximate reasoning distinguishes the suggested approach from "classical" expert systems, by allowing the modelling of vague concepts and enabling flexible rule interpretation with rule adjustments.

The authors used five-scale to illustrate preferences. However, if data were presented by discrete binary choices (e.g., stated choice data), the generalisation ability of the model has not been properly explained. Secondly, the model only was examined by single data source: simulated stated preference data. Whether the model is applicable and valid for those survey data and the multiple data sources (i.e., RP and SP), the authors did not give a proper explanation.

To sum up, the implications from the FL model applications can be summarised as follows:

- Introduction of 'Fuzzy Utility', which can be viewed as a sort of choice probabilities, sharing the similar model expression with logit models. However, the principal difference between traditional utility expression based on the RUT and 'fuzzy utility' is that in FL model, the utility is based on fuzzy rules and the approximate reasoning mechanism (min-max algorithm), and the non-linear relationship between inputs and outputs can be presented by 'fuzzy utility' expression.

- Uncertainty and vagueness of decision making: through taking into account uncertainty and vagueness, which can also be viewed as non-linearity in FL models, the model performance and forecasting ability can be effectively improved.

How it can be linked with this research

Conventional models of discrete choice analysis (MNL models, for example) are based on the random utility framework. They assume that decision makers make rational decisions and choose the alternative which maximises their utility. Furthermore, they use an error term to capture the uncertainties and ambiguities inherent in the choice problem such as unobserved attributes, measurement errors, imperfect information, instrumental variables, and unobserved taste variations (Ben-Akiva and Lerman, 1985; Manski, 1973).

The advantages of using fuzzy sets and fuzzy logic theory as an alternative method to analyse discrete choice problems have been discussed above. These approaches model the decision makers' perceptions of the attributes of the various alternatives using fuzzy sets and linguistic variables, and the decision process itself, using concepts from approximate reasoning and fuzzy control. The underlying assumption is that decision makers use a few simple rules that relate their vague perceptions of the various attributes (particularly for attributes of boarding time savings, overall service quality, etc.) to their preferences towards the available alternatives.

When we discuss the connection of fuzzy logic methods with conventional choice models, we need to look at the characteristics of input and output data being used in this research firstly. In both the RP and SP survey, the input data are presented by attributes related to different fare payment alternatives. Regardless the RP data based on respondents' actual payment behaviours and the SP data based on hypothetical situations, respondents' perceptions about changes of different levels of attributes are vague and overlapped rather than crisp and independent (Lotan and Koutsopoulos, 1993a). As to the output data, respondents' choices toward different payment alternatives also reflect the decision making procedure of combining attributes related to alternatives. When evaluating benefits and effectiveness of public transport fare payment applications, based on various service quality attributes: travel cost, boarding time savings, accessibility, overall assessment of service quality etc., passengers can jointly trade off these attributes and only offer the linguistic information towards different payment applications, such as "Very High", "High", "Low", and "Very Low", etc. (e.g., boarding time savings can be stated by "*High*"). Compared with fuzzy set theory, using 'classical' mathematical techniques is frequently hard to quantify those linguistic variables. This linguistic information represents subjective knowledge; on the contrary, those formulae and equations are objective knowledge. Therefore, it would be of interest to consider subjective and objective knowledge. Fuzzy logic and fuzzy set theory is an extremely suitable concept with which to combine subjective knowledge and objective knowledge together (Teodorvic, 1999).

The link of fuzzy logic models with discrete choice models can be addressed from the utilities of alternatives. Fuzzy utility methods, which were proposed by Hoogendoorn et al (2000), Mizutani and Akiyama (2000, 2001 and 2003), can be used to measure and compare fuzzy utilities of different alternatives. Such fuzzy utility are of two types as follows:

- Fuzzy utility u_j and choice of the alternative j^* are directly connected with the fuzzy rule base (e.g., "**if** $A_1=a_1$ **AND** $A_2=a_2$, **then** utility=high"). Defuzzified utility U_j for each alternative by using the centre of gravity method is obtained and direct comparison of the fuzzy utilities u_j using, for instance, fuzzy ranking yielding the best alternative j^*.
- Another solution is that the systematic term (V) in the utility model is replaced by fuzzy logic term, which is based on fuzzy rule base. Compared with the first type, this solution only takes into account the systematic term in the utility model, not changing the expression of the error term, because the final choice probabilities are determined by the systematic term with fuzzy logic theory.

Expected Outputs

In this research, the expected outputs are values ranging between 0 and 1, which can be regarded individual choice probabilities of alternatives under the fuzzy inference system so that it is able to be compared with the conventional discrete choice model. In addition, through viewing the fuzzy inference rules and choice probabilities, it could explicitly provide us with an insight into the relation of cause and effect of fare payment behaviour.

As to the model performance, it can be assessed by using "percentage correctly predicted" as a measure of goodness of fit, compared with raw data. Also, proposed by Turksen and Wilson (1994), the term of share error was used to measure the prediction accuracy of market share by using fuzzy logic methods when studying consumer choice problem, particularly for stated preference data.

4.5.2. Neural Network Methods

Introduction to Artificial Neural Network Methods

In recent years, neural networks have been increasingly applied to a wide range of marketing research problems, including brand/mode choice problems (Bentz and Merunka, 2000; Nijkamp et al, 1997), market share forecasting (Agrawal and Schorling, 1996). Neural networks are particularly suitable for mode choice problems in transport domain as a comparative approach to conventional analysis methods (Vaughn et al. 1992), because the databases available usually contain a large number of respondents' responses to alternatives in the choice set, such as perceived quality and variations, regardless of revealed or stated preferences (or choices).

An ANN consists of a number of connected nodes (in the literature nodes are also referred to as neurons, units, or cells) each of which is capable of responding to input signals with an output signal in a predefined way. These nodes are ordered in layers. A network consists of one input layer, one output layer, and an arbitrary number of hidden layers in between. This number can be chosen by the user such that the network performs as desired. Typically one or sometimes two hidden layers are used. One reason for this is that one hidden layer is sufficient to approximate any continuous function to an arbitrary precision (Homik, Stinchcombe and White, 1989).

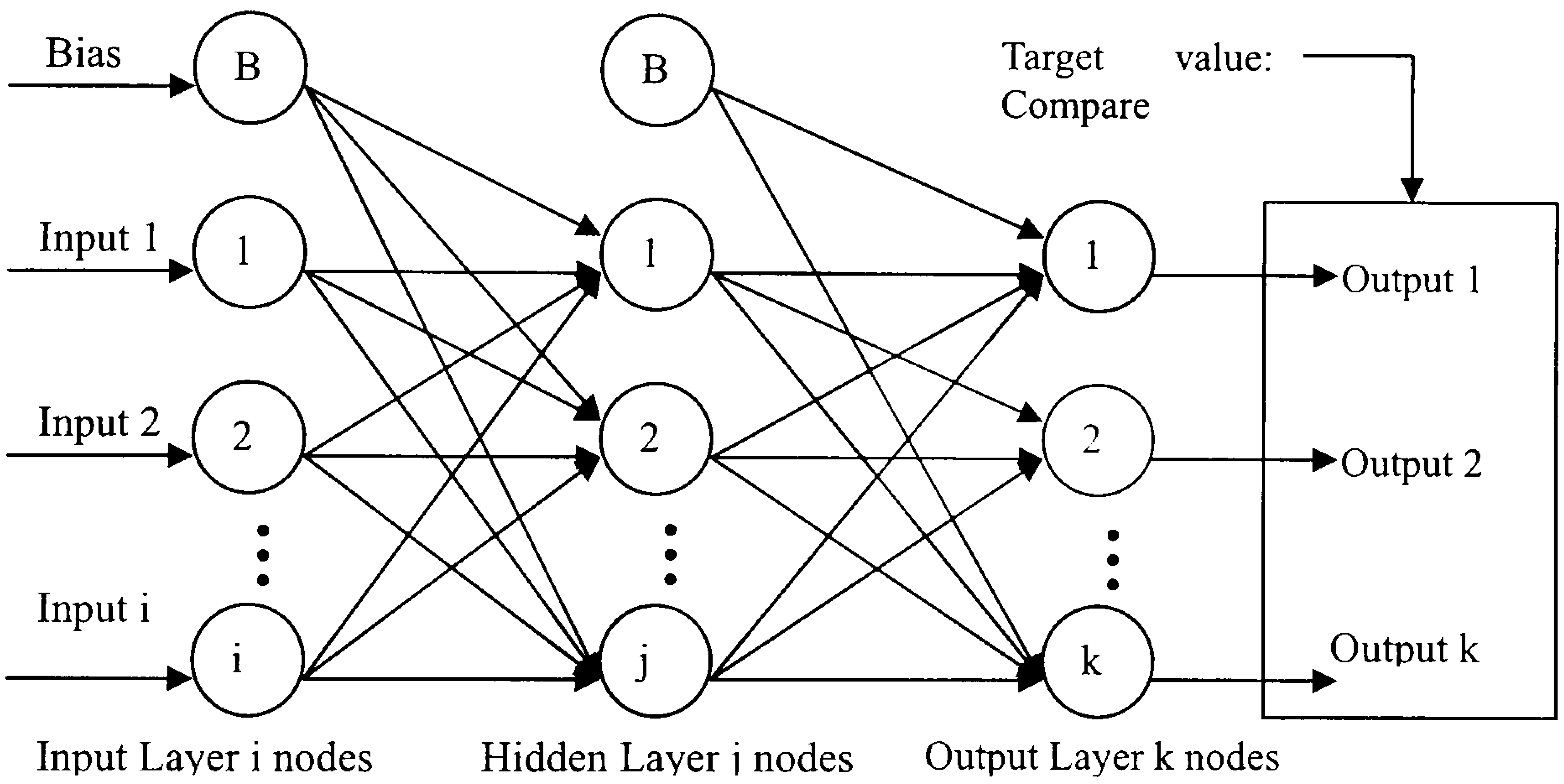


Figure 4.7 A Three Layer Artificial Neural Network with Biases

For an illustration consider the three-layer ANN in Figure 4.7. This ANN consists of three layers, the input layer (the leftmost), one hidden layer (in the middle), and the output layer (the rightmost). The nodes are connected such that each node is connected to all nodes of the previous and the successive layer (if such layers exist). The input layer is only connected forward to the first hidden layer and the output layer only backward to the last hidden layer. All connections are assigned a weight (a real number). Often an ANN also contains biases (denoted by node *b* in Figure 4.7). These are dummy nodes which always provide an output of +l. They are useful in translating the [0, 1] output from the logistic function.

Similar to estimation of logit models over an estimation period data, the ANN gets trained on a set of training data. ANN starts out by an initial set of weights chosen randomly, typically between (-1, 1). It then adapts the weights in such a way that given the input signals, the ANN's output signal(s) match the desired output signal(s) as closely as possible (the convergence limit is specified by the user).

We use a particularly popular algorithm called the backpropagation (BP) algorithm in this study. The basic algorithm works as follows. The input to a node is computed as the sum of the outputs of the preceding nodes multiplied by the weight of the connection. This is expressed as:

$$NET = \sum_{i=1}^{n} OUT_i w_i \tag{4.28}$$

Where: OUT_i = the output of node j in the previous layer,

w_i = the corresponding connection weight.

For the input layer OUT_i is simply the vector of input values. This sum is then transformed to a value between 0 and, 1 using the so called logistic or sigmoid function.

$$OUT = \frac{1}{(1 + e^{-NET})} \tag{4.29}$$

Starting with the first hidden layer this calculation is done from left to right until the output layer is reached. All training pairs are presented to the ANN and the sum of squared errors over the whole training set is computed. If the sum of squared error exceeds the specified error goal, the ANN adjusts the connection weights. This is called a training epoch. The ANN then begins another training epoch until either the maximum number of training epochs is reached or the sum of squared errors reaches the specified error goal. The training is said to be complete when either of this happens. One can think of this as moving on the (often multidimensional) error surface in the direction of the steepest descent. How well a network is trained is measured by the mean sum-squared error over the complete training dataset.

The connection weights are adjusted as follows. Starting with the weights connecting output layer and the last hidden layer the weight adjustments are propagated backwards using

$$\delta_{p,output} = OUT(1 - OUT)(TARGET - OUT) \tag{4.30}$$

Where $\delta_{p,output}$ is the delta value of node p in the output layer.

Based on this the weight change is calculated:

$$\Delta w_{pq,k} = \eta \delta_{qk} OUT_{pj} \tag{4.31}$$

Where $\Delta w_{pq,k}$ = weight change of connection from node p in layer k-1 to node q in layer k,

η = learning rate (which can be set by the user),

$\delta_{p,output}$ = delta value for the node q in layer k, and

OUT_{pj} = output of node p in layer j (same as k-1).

The new weight assigned to this connection is computed as:

$$w_{pq,k}(n + 1) = w_{pq,k}(n) + \Delta w_{pq,k} \tag{4.32}$$

where n denotes the current iteration (before weight adjustment) and n+1 the next iteration (after weight adjustment). This procedure is repeated for all nodes in the output layer.

Afterwards the incoming connections of the previous layer are updated.

For layers other than the output layer is used as follow:

$$\delta_{p,j} = OUT_{p,j}(1 - OUT_{p,j})(\sum_q \delta_{q,k} w_{pq,k}) \qquad (4.33)$$

Where: $\delta_{p,j}$ = delta value for the node p in layer j;

OUT_{pj} = output of node p in layer j;

$\delta_{q,k}$ = delta value for the node q in layer k and;

$w_{pq,k}$ = weight of connection from node p in layer k-1 (same as j) to node q in layer k.

The other steps remain the same. This procedure continues until a specified error is reached or a specified number of training epochs are over. All above is the whole procedure of the algorithm of ANN model. Next step in this section is to see how ANN can be linked with the survey data for this research.

Previous Studies of Discrete Choice Behaviour with ANN Technique

As we discussed in the previous section, the knowledge contained in fuzzy systems are transparent to the user but cannot be acquired directly from data. However, sometimes relationships between inputs and outputs are required to be identified (i.e., fuzzy rules); the disadvantage of FL models on it is shown out. Artificial neural networks (ANN's), on the other hand, have the ability to learn the knowledge from a set of data by the network itself without any *a priori* assumptions about the mapping relationship between inputs and outputs, but the knowledge gained is hidden from the user.

Artificial neural networks have been widely studied for information processing. But recently there has been an increasing interest in application of neural network techniques to transportation studies. In recent years, different transportation application problems analysed with neural networks have been reported including: classification and pattern recognition (Faghrin and Hua 1992), travel demand forecast (Yang et al. 1992; Shih-Miao Chin et al. 1992), freeway incident detection (Ritchie et al. 1992) and driver route choice analysis (Dougherty and Joint 1992). It is generally reported that the neural network has the ability to learn complicated problems without the requirement of giving explicit equations correlating input/output data, and can generate reasonable results efficiently. The neural network approach utilises an iterative data matching technique and is often confused with artificial intelligence (Berardinis, 1992).

In particular, this approach is being used as quick and efficient method to analyse discrete choice behaviour and as a comparative approach to conventional analysis methods (Vaughn et al, 1992). Meanwhile, more and more studies focused on discrete choice problems and market research (e.g., forecasting market share and segmentation analysis) by using ANN methods, especially about comparisons of advantages and disadvantages between logit models (e.g., MNL model) and ANN methods (Bentz and Merunka, 1998; Carvalho et al, 1998; Hruschka et al, 2002; Nijkamp et al, 1997), which are highly related to this research.

In this section, the literature review for ANN technique focuses on applications in travel demand forecasting. Meanwhile, through reviewing these previous studies, comparisons between ANN models with MNL model on forecasting ability and model performance of these two methods can be made.

Agrawal and Schorling (1996) empirically compared the forecasting ability of ANN with MNL in the context of brand share in the market place. Three-layer neural network structure was employed in this research. Except this study, some other literature also recommended this network structure (three-layer network structure) as the primary option (Yang, et al, 1993; Bentz and Merunka, 1998; Santamaria, 2003). In addition, trial and error method was used to optimise the number of nodes in the hidden layer and the number of epoch in the ANN model. The results indicated that the forecasting ability of the ANN model was better than that of the MNL model from five aspects the authors discussed (the complexity of the choice context, reorganization of input data pattern, alternative switching behaviour, household heterogeneity, and sensitivity to the number of observations). Meanwhile, the authors analysed sensitivity of the forecasting error to the length of the estimation (training) period for the MNL (ANN) model, and to the different schemes for classifying households into homogenous segments. The results were reasonably robust to the different clustering criteria and the length of estimation period. However, the ANN model in this paper only focused on the choice problem in the aggregate level and such comparison of the ANN model and MNL model also was based on aggregate results. According to the algorithm of ANN, individual choice probability may be obtained and measured. In addition, how those brand shares could be assigned on the individual level between the ANN and MNL model, which was not presented in this study, can give more detailed explanations of the forecasting ability of different models.

Bentz and Merunka (1998) studied brand choice decision-making with a hybrid approach combined MNL and ANN models. A feedforward neural network with Softmax (Bridle, 1990) output units and shared weights were employed and can be viewed as a generalisation of the Multinomial Logit model, providing a network has no hidden neurones. However, the complexity of the ANN model can be easily increased by changing the architecture of the network, enabling more complex relationships resulting from non-linear consumer preferences (threshold and interaction effects). So the main difference between the two approaches lies in the ability of neural networks to model non-linear preferences with few (if any) a priori assumptions about the nature of the underlying utility function, while the Multinomial Logit can suffer from a specification bias. The neural network is used as a diagnostic and specification tool for the Logit model, which will provide interpretable coefficients and significance statistics. However, poor interpretability of ANN models also was pointed out in this paper. Unlike estimated coefficients in MNL models, parameters in ANN models can not give information as much as in MNL models.

In this paper, the concept of Softmax was introduced, which was firstly used by Bridle (1990) to generalise MNL models with shared weights. The model details can be written as follow:

$$\frac{e^{\sum_k w_k x_{ijk}}}{\sum_{j=1}^{J} e^{\sum_k w_k x_{ijk}}} \tag{4.34}$$

W_k: Weight of input x

X_{ijk}: Input of trading-off occasion *i*, each alternative *j* is described by the same *k* attributes

Figure 4.8 below illustrates the network structure with the Softmax output. The network comprises *n* output neurones, *n* being the number of considered alternatives. For each trading-off occasion *i*, each alternative *j* is described by the same *k* attributes (x_{ij1} ,..., x_{ijk}). The neurone activation functions are all linear except the ones for the output neurones, which are normalised exponentials.

Because it is a continuous version of the 'all to the winner' activation function, for which the output of the neurone with the biggest input is 1, all the other outputs being 0, a Softmax output can therefore be viewed as a choice probability (Bishop, 1995). The principal advantage of Softmax output is that it can be viewed as the generalisation of MNL model, because in Figure 4.8 if no hidden layer is applied, those shared weights in Equation 3.3 can be regarded as the coefficients of alternative attributes.

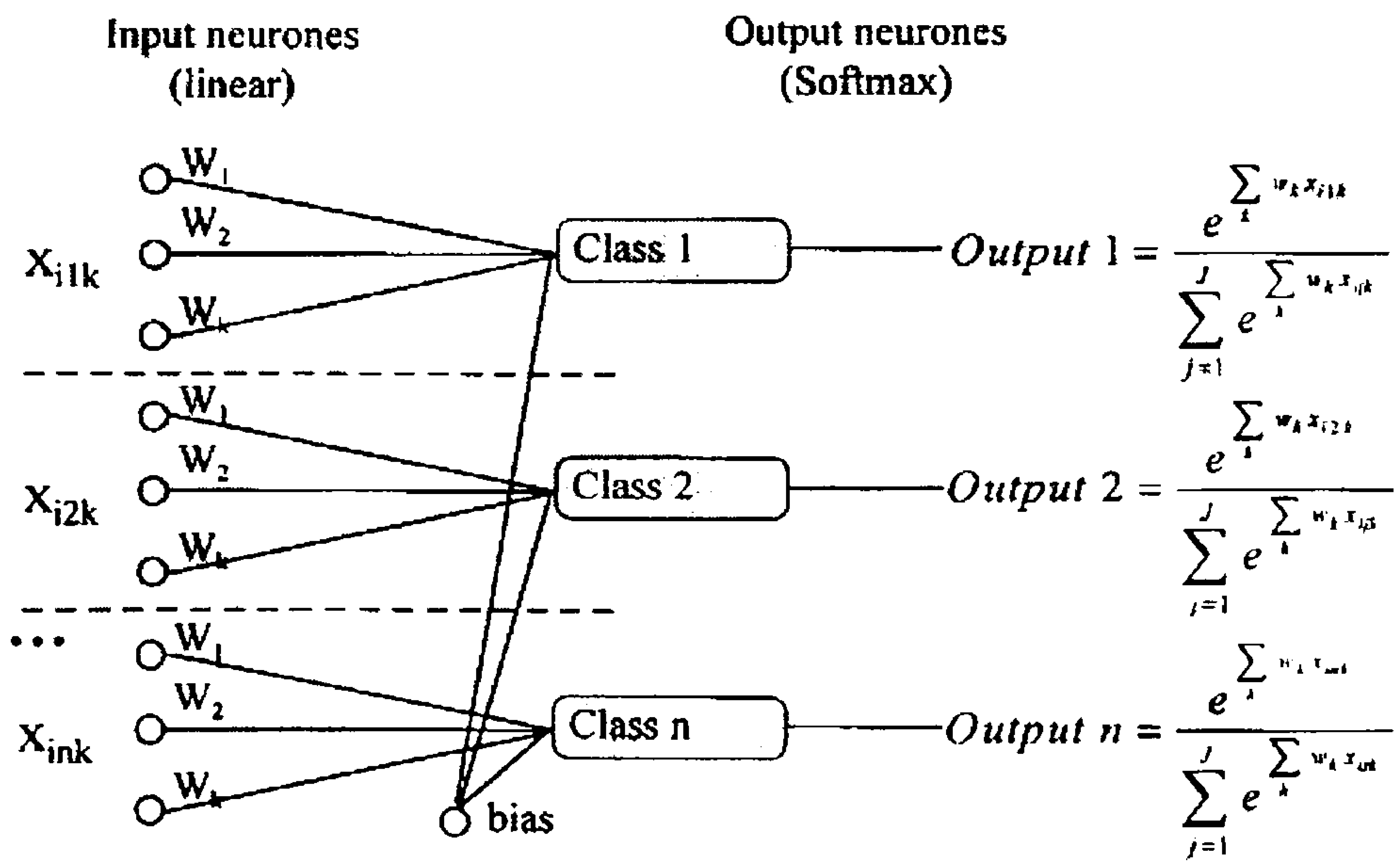


Figure 4.8Neural Network with Softmax Outputs

(partially connected and with shared weights)

Through comparing with the standard MNL model, the ANN model outperformed the MNL model and is relevant in cases where decision variables interact. Such interaction detection and modelling suggest an interesting research direction in the understanding of consumer choices at an individual level. On the other hand, the model can help explain household heterogeneity with socio-demographic variables and their interactions with alternative attributes.

Hruschka *et al* (2002) extended the work of Bentz and Merunka (1998), which firstly discussed ANN methods in brand choice problem and developed the combined models of ANN and MNL, in several respects. Firstly, the authors considered reference price as additional dynamic predictor besides brand loyalty. Secondly, they were also looking at the latent class extension of the (linear utility) MNL model developed by Kamakura and Russel (1989) which was abbreviated as LC-MNL model, whereas Bentz and Merunka only compared to the homogeneous linear utility MNL model. Latent class models conceive the population of households as finite mixture of classes or segments. They deal with heterogeneity of parameters w.r.t. households. The comparison of ANN-MNL models and LC-MNL models allows them to assess whether the latter models are necessary to account for heterogeneity across households. Thirdly, they estimated the loyalty smoothing parameter like the other parameters by maximising log likelihood. The methodology in this paper focused on two approaches: (1) combining neural network methodology with MNL choice models; (2) using latent class extension models, to allow for the discovery of nonlinear effects of marketing variables. According to the empirical studies, both the latent class and neural network methods clearly outperformed the homogeneous linear utility of MNL models. However, the latent class extension of the MNL model leaded to higher log likelihood values on estimation data. But the good performance of latent class models on estimation data did not carry over to test data. Neural network models achieved much better out-of-sample log likelihood values than their latent class rivals.

Nijkamp *et al* (1997) used the logit model as a benchmark for evaluating the results of ANN models, based on an empirical case study of mode choice (e.g., rail and road) from Italy. Two statistical models were discussed and compared: (1) a traditional logit model and (2) a new technique for information processing: the feedforward neural network model. In the study two different cases – corresponding to a different set of attributes – are investigated, namely by using only 'time' attributes and by using both 'time' and 'cost' attributes. Through comparisons, the feedforward neural network model seemed to provide reasonable predictions compared to those obtained by means of a logit model. Another aspect worth noting is that the same predictions in two different cases by MNL models were likely caused by the property of IIA underlying discrete choice models. On the contrary, the

ANN approach seemed to be more sensitive to changes in the input information. An important lesson in this paper, however, was to define proper neural network architecture and to train the network sufficiently during the learning phase. Finally, the sigmoid function used in an ANN analysis for training the network was essentially a logistic function related to a binary choice model. This may lead to another correspondence between ANN models and discrete choice models. So it was clear that the application of ANN analysis in transportation behaviour still deserved more careful methodological attention.

Carvalho *et al* (1998) used backpropagation (BP) artificial neural networks to forecast travel demand from disaggregate discrete choice data and compared them with logit models. The motivation of the methodology proposed by the authors is to overcome linearity in the standard utility model by introducing an underlying non-linear function of ANN models. The authors discussed three kinds of data in the same ANN and MNL models: synthetic data which fulfilled the underlying logit assumptions, synthetic data which breached the underlying logit assumptions and real data. The purpose of using simulated data in the paper was to be able to test the performance of both logit and ANN methods against a 'known' situation. Different from other studies on ANN models' forecasting ability, this paper compared ANN and MNL from error generated during the model estimation, such as RMSE (root mean squared error), MMS (mean of market share) and VSE (variance of square error). It was found that ANN without hidden layers exhibited almost identical performance with logit model in all three scenarios. For the synthetic data which breached the underlying logit assumptions and with real data, the BP network with a hidden layer can achieve a better model fit than logit models. However, as most papers pointed out, careful choice of the number of hidden units and training iterations was needed to avoid over-fitting and consequent degradation of performance.

Yang *et al* (1993) explored a route choice problem by using a three-layer neural network model. The results indicated that most subjects made route choice based mainly on respondents' recent experiences. It was also demonstrated that route choice behaviours are related to the personal characteristics as well as the characteristics of the respective routes. The model developed in this paper was for choice outputs consistently provided at a level of 75 percent accuracy, comparing with desired outputs in the data set. As to the pattern of the ANN model output, the authors used 0-1 value to represent the choice of different alternatives. During the training of the neural network, the desired output is set to be 1 if alternative A is chosen and 0 otherwise. During the testing or prediction, alternative A is estimated to be chosen if the output value is greater or equal to 0.5, and the side road is estimated to be chosen if the output value is less than 0.5. However, faced with 75% prediction accuracy, the authors did not give a further discussion about calibrating and improving the model performance and network generalisation, particularly for those raw outputs of 0.5, an insight into this phenomenon should be explained properly, because if we regard 0.5 as choice probability, that means respondents have the same preference toward either alternatives (if binary choice situations apply).

Combination of FL and ANN

Except the pure fuzzy logic methods used in analysing discrete choice problems with consideration of uncertainty and vagueness of perceptions, some literature tried to combine artificial neural network technique with fuzzy logic models. Fuzzy reasoning is a method to describe human approximation for decision making. On the contrary, artificial neural network (ANN) can realise a highly non-linear model with parameters given by a learning process based on an error minimisation principle. In addition, combined models (FL and ANN) can overcome arbitrariness of fuzzy rule selection (although fuzzy inference system is called a sort of expert system) by using the network learning process to generate fuzzy rules so as to improve the model performance.

Among those studies on analysing mode choice problem with fuzzy logic technique, Vythoulkas and Koutsopoulos (2003) extended the traditional fuzzy logic approach by incorporating rule weights, which captured the importance of a particular rule in the decision process. It also presented an approach for calibrating the weights using concepts from neural networks to simulate human learning process so as to find the best fit rules and the importance of rules in the rule base. In order to show such uncertainty in decision making, the authors treated rules in the rule base differently by introducing probabilistic rule choice and deterministic rule choice. Deterministic rules indicated all those rules have the same importance and were assigned the weight of one, while probabilistic rules can reflect uncertainty and randomness of human choice behaviour by redefining the utility model as follow:

$$U_{in} = \beta_0 + \beta_{centr\_i} centroid_i + \varepsilon_{in} \tag{4.35}$$

U_{in}: utility of alternative i of individual n.

β_0: constant

β_{centr_i}: weight of centroid i

ε: error term

By comparing different estimation results, the probabilistic rule choice showed a better output than deterministic rule choice.

However, there would be two potential issues of Vythoulkas and Koutsopoulos' model: (1) if a large number of parameters were considered in input data set, the model calibration would become relatively complicated in computation. The authors did not give a proper solution to deal with it in the neural network. (2) Furthermore, the lack of systematic approaches to perform evaluation and hypothesis testing is another issue which have not been tackled in the paper. It would not be solid that the authors simply used "% correctly predicted" and "error" to compare the outputs. To adjust network parameters and examine the performance of different neural network structures is a possible solution.

Compared with Vythoulkas and Koutsopoulos' work (neuro-fuzzy model), Akiyama *et al.* (1997) discussed fuzzy neural models to analyse the route choice problem. The difference between neuro-fuzzy and fuzzy neural models is that neuro-fuzzy models firstly used neural network technique to train each rule in the rule base and then fuzzy reasoning and proximate theory were eventually employed. That means the whole framework of neuro-fuzzy models is in fuzzy logic domain. But fuzzy neural models employ fuzzy logic theory prior to the data input to the neural network to determine the fuzzy inference mechanism and then the neural network is used to generate the human learning process and apply it to get outputs. In this paper, two kinds of models were examined: neuro-like fuzzy model and neural fuzzy model. But the authors only gave some brief positive comments about fuzzy neural (FN) model, which had not been examined by survey data. So the actual performance of FN model still cannot be proved sufficiently.

Sayed and Razavi (2000) put emphasis on dealing with non-linear relationships among different variables and increasing interpretability of the model, by combining the learning ability of neural networks and transparent nature of fuzzy logic technique. According to the classification of FL and ANN combined models, basically the model used in this paper was "Fuzzy-like neuro" model. The model used an ANN learning algorithm to determine its parameters (i.e., fuzzy sets and fuzzy rules) by processing data samples. Therefore, it can be trained to perform an input/output mapping, just as with an ANN, but with the additional benefit of being able to provide the set of rules on which the model was based. Another merit of Sayed and Razavi's model was that it could effectively identify and exclude those input variables that did not have a significant contribution to the whole estimation. Hence, only those significant variables were kept in the model such that the number of rules in the rule base can also be reduced effectively, which achieved the simplicity in the following calculation process. The transparency by introducing fuzzy logic theory enabled users to directly add their own expertise on the subject before or after the model was built.

To sum up, the implications from the ANN model applications can be summarised as follows:

- Non-linearity: because ANN models do not require any a priori assumption about mapping relationship between inputs and output, it is more suitable to model non-linear decision making process.
- Outputs of ANN models can be comparable with logit models, because individual choice

probabilities can be worked out in both models. However, weak interpretability of ANN models result in simplicity of model outputs, unlike logit models, directly outputting estimated coefficients, correlations of two variables, value of attributes, *etc.*

- Faced with a number of neural network algorithms, backpropagation (BP) has been widely used. Features of input (i.e., payment attributes) and output data (i.e., choices) in this research decide that BP algorithm can be employed to model discrete choice problem.
- Combination of FL and ANN techniques: the main contribution of combining FL and ANN techniques to this research is to allow the ANN model learning and capturing the best fit rule base so that the bias caused by the determination of the rule base could be reduced.

ANN models can also be linked with logit models, because it was found that if the hidden layer of network was removed, outputs of ANN model without hidden layer were almost the same as logit models, which is caused by the similarity between sigmoid transfer functions (e.g., logsig or tansig) in ANN models and the probability equation in logit models.

How it can be used in this research

The trading off process of human being faced with several alternatives can be the same under conventional discrete choice models (e.g., MNL models) with utility maximisation and ANN models with mapping-driven mechanism (Nijkamp et al, 1997). In this research, both MNL and ANN are based on attributes and levels of different fare payment alternatives as input data, and respondents' choices of different payment methods as outputs.

But the differences between conventional choice models (i.e., MNL models) and ANN models also should be stressed before using ANN models:

(1) Model parameters can be interpretable in discrete choice models (the relationship between alternative utilities and variables). However, parameters in ANN models lack of interpretability. Therefore, some outputs related with utility parameters in choice models cannot be directly obtained in ANN models, value of time, for example. But in ANN models, degrees of freedom are often large enough to allow the network to fit the same function with different combinations of parameters. This is probably the reason why neural networks have been called 'black boxes', capable of mimicking relationships between a set of variables but incapable of explaining the nature of these relationships.

(2) Compared with conventional choice models, ANN models are more flexible on the model design, because designers can choose the network structure (the number of layers and nodes), training function, transfer function, acceptable error goal, *etc.* However, in order to achieve the better network performance, trial and error, and comparisons

between simulated outputs and desired outputs are needed due to such flexibility.

The ANN method consists of following steps for analysing the preference data in this research:

- Model inputs and outputs: in ANN models, inputs are variables related with different fare payment alternatives; outputs (target values) are choices of alternatives.
- Selection of network structure, including the number of network layers, neurons and connection style between different layers and neurons, *etc*.
- Selection of training function and transfer function. This can influence the model estimation duration and acceptable error-approaching performance.

Before using ANN models, another necessary work is the data compilation for the data inputs suitable for the ANN models: some data need to be normalised to the range of [0 1]; qualitative data should be recoded as [0 1] input. The number of columns of each qualitative attribute depends on how many levels the attribute has.

The estimation procedure by ANN models is of complexity to programme the whole computation in some programming languages, such C++, VB, *etc*. However, some specific software packages are available for analysing ANN models (network design, applications, *etc*), such as Neural-Works Professional II Plus, Neural Network toolbox in MATLAB, facilitating our work for the complex computation. By considering the availability of software packages, in this research MATLAB is chosen for modelling RP and SP data with ANN methods. In addition, MATLAB also provides Fuzzy Logic toolbox for FL analysis as we are using in this research.

Expected Outputs

Expected outputs under ANN models are values between 0 and 1, because in the data sets, respondents' choices are coded with 0 and 1 (0 means choosing alternative *A*; '1' means choosing alternative *B*, if two alternatives *A* and *B* being traded off). After the ANN training process, the trained network could output any values ranging from 0 to 1 (probably they may not be exact values of 0 or 1), because the mapping mechanism of the network training process can only drive the best fit relationship between inputs (training values) and outputs (target values) by considering different actual mapping relationship between inputs and outputs. Thereafter, we may regard values between [0 1] as probabilities of choosing a certain alternative.

One of measurements to evaluate the ANN model's performance: "percentage of data-matching between outputs and targets", is to compare the simulated output values with the target values, and the higher matching percentage indicates the better network performance. Another aspect to identify the performance of the trained network is the network generalisation. The definition of the network generalisation is that by using different data sets from the training data, the network still can achieve satisfactory outputs (high data matching percentage with low error). In order to realise the network generalisation, "over-fitting", occurring during the network training, needs to be solved in the model calibration stage. "Over-fitting" phenomenon is that the error on the training set is driven to a very small value, but when new data is presented to the network the error is large. Moreover, the network has memorised the training examples, but it has not learned to generalise to new situations. Therefore, over-fitting problem could impact the performance of the network when using new data sources.

4.6. Model Application

4.6.1 Demand Forecasting and Evaluation

The evaluation of benefits of smart cards in the model application stage can be discussed by the following aspects: market share forecasts, valuations of attributes, travel cost elasticities. Market share forecasts can reveal changes of use of different fare payment options based on aggregated level when some payment attributes change. Valuations of attributes reflect users' willingness to pay towards different payment features and relevant variations. Travel cost elasticities examine users' demand changes with respect to changes of travel cost (own and competitive alternatives). In the mean time, in order to examine the homogeneity and heterogeneity of preference, the model application looks at the segmentation analysis. Through the segmentation analysis, benefits of smart cards to respondents with different socio-economic background can be identified (such as importance of attributes, willingness to pay, etc).

The first model application is market share forecasting through developing fare payment choice behavioural model with the RP and SP data.

The forecasting procedure is based on the sample enumeration approach. This is suitable for forecasting the effects of policies that impact differently on various groups of the public (Ben-Akiva and Lerman, 1985; Bradley and Kroes, 1992).

Choice probabilities of a certain alternative diverted to other payment methods can be obtained from the choice behaviour model (e.g., MNL models) for each respondent. These probabilities are influenced by the system characteristics and the personal characteristics. The predicted diversion for the whole sample is an average of the probabilities of the whole sample size. (The predicted diversion for each group of the public is an average of the probabilities of the people in the group.). Through forecasting market shares, the demand of new fare payment method, smart cards, can be particularly identified under a range of different situations, when respondents trade off between smart cards and other fare payment options.

Therefore, the evaluation study for fare payment methods (particularly for smart cards), include the following aspects:

- Analysing forecasted market share, meanwhile, comparing results from different data sources (i.e. pure RP, pure SP and joint RP and SP);
- Valuation of attribute: the value of time (VOT), as one of primary measurements to identify respondents' perceptions toward different transport alternatives, has been widely used to measure preference response. It indicates the amount of money travellers are willing to pay in return for savings in journey time (Pearmain *et al*, 1991). In this research, VOT is presented by travel cost and boarding time saving variables, called value of boarding time savings (VOBTS).

VOT can be measured by calculating the ratio of estimated boarding time savings and travel cost coefficients from a predefined utility function of mode choice model. For example, the utility function, U, may be defined as:

$$U = \beta_1 T_m + \beta_2 C_m + \varepsilon_m \tag{4.36}$$

where T_m and C_m respectively represent alternative *m'* s travel time and travel cost, β_1 and β_2 the parameters, and ε_m the error term. From Equation (4.36), the value of travel time can be calculated as:

$$VOT = \frac{\partial V / \partial T_m}{\partial V / \partial C_m} = \beta_1 / \beta_2 \tag{4.37}$$

According to Equation (4.37), the VOBTS for this research can be written as follows:

$$VOBTS = \frac{\alpha_{BTS}}{\alpha_{TC}} \tag{4.38}$$

α_{BTS}: Estimated coefficient of boarding time savings (Second);

α_{TC}: Estimated coefficient of travel cost (Yuan/month).

Valuations of other attributes mainly include those qualitative variables about fare payment methods in the survey. In this research the values of the attributes of fare payment alternatives are calculated by estimated coefficients of attributes and coefficient of travel cost. Because utility models we use are linear additive, the general expression is very similar with value of time in the previous section (See Equation 4.39):

$$VOA = \frac{\beta_i}{\beta_{cost}} \tag{4.39}$$

β_i: Estimated coefficient of attribute *i* to be measured by monetary value

β_{cost}: Estimated travel cost coefficient.

- Fare elasticities: An elasticity indicates sensitivity of demand to change in some variable when some other variables keep constant. Elasticities are defined as the proportionate change in demand after a proportionate change in some variable, therefore it can be written as follows (4.40):

$$\eta = \frac{\frac{\Delta V}{V}}{\frac{\Delta X}{X}} = \frac{\Delta V}{V} \cdot \frac{X}{\Delta X} \tag{4.40}$$

For a small change, the point elasticity is defined as:

$$\eta = \frac{\partial V}{\partial X}\frac{X}{V} \tag{4.41}$$

The purpose of analysing elasticities is to forecast payment demand when a proportion of change of payment attributes changes. Secondly, elasticities can reflect the competition degree between alternatives (e.g. the change of demand of a certain payment method when the change of travel cost by using the competitive payment alternative). In addition, elasticities can be used to assess the different discrete choice models that were employed for the RP and SP data analysis, such as pure RP, pure SP and joint RP and SP models.

In this thesis, two types of elasticities were looked at: own elasticity and cross elasticity. The 'own' elasticity measures the demand response to a change in 'own' payment service attribute-level (e.g. the percentage reduction in cash payment demand from a 10% increase in PT cash fare). The 'cross' elasticity measures the percentage response in travel cards or smart cards to a change in cash fare payment.

4.6.2 Segmentation Analysis

In the model applications, one of issues that must be addressed is that of market segments (e.g. high vs. low income; higher educational level vs. lower educational levels, *etc*). Market segmentation is the process in marketing of dividing a market into distinct subsets (segments) that behave in the same way or have similar needs, but different to other segments.

Market segments are important for several reasons:

- Segments often exhibit different preferences, so better descriptions of market behaviour can be obtained by taking them into account;
- Such differential preferences generally result in some groups being more interested in a given fare payment method than others;
- Market segments help to define sampling frames, sample sizes and sampling methods.

Why segment in this research?

Segmentation analysis may explore choice behaviour of different respondents' groups, and so that fare payment alternative demand can be specialised. Moreover, for discrete choice modelling reasons, the standard choice model treats all respondents as having the same weight for each attribute in the utility function. However, it is reasonable to expect that different groups of people may have different coefficients for some attributes. For example, coefficients of time and cost may vary across different income groups because of time and money constraints.

How to segment?

The requirements of successful market segmentation need to concern the following factors:

- Homogeneity within the segments;
- Heterogeneity between segments;
- Segments are measurable and identifiable;
- Segments are accessible and actionable;
- Segment is large enough to reliably estimate segment specific parameters in the utility models.

In this research, we assume a number of possible *a priori* segments based on the socio-economic variables designed in the RP and SP survey (we assume that such difference of choice related to different socio-economic segments exists in the data set), including, age, gender, employment status, household monthly income, educational level, availability of private transport and an attitudinal question: willingness to prepay PT fare. The whole sample is segmented by categories of socio-economic variables we designed, for example, by age factor respondents may be grouped by 16-25, 26-45, 46-60 and over 60 years old. However, if all these variables were used in the segmentation analysis, the model would be massive and complicated. Therefore, age, gender and household income are finally selected and used in the segmentation analysis, because compared with other factors, these three factors have been widely used as primary factors in previous studies when forecasting respondents' choice behaviour.

4.7. Summary

In this chapter, the methodology of this research is outlined as the research design. The whole evaluation of benefits of smart cards is as follows: First of all, the research focuses on evaluating benefits of smart cards in demand side (PT users); therefore, a preference survey based on individual PT users is used to collect preference data about different fare payment options. In order to achieve the data collection of different data sources, RP and SP surveys are used to collect choice behavioural data toward three payment means (cash, travel cards and smart cards). The RP data is based on people's actual choices, but the SP survey can capture choice behaviour under hypothetical situations (such as when some new payment features were introduced). The combination of the RP and SP data can make full use of advantages of two kinds of data sources to evaluate payment choice behaviours.

Because the benefit evaluation is finally explained by users' perceptions towards different payment attributes, the survey is designed according to features of fare payment options. In the RP survey, as existing payment applications, cash, travel cards and smart cards are used. In the SP survey, new attributes and new variations for the three fare payment options are introduced. Through these payment features in the RP and SP survey, value of boarding time savings and other attributes, the importance of attributes can be identified in the later data analysis.

Secondly, in the preference data modelling analysis, MNL models are firstly used, including MNL-RP, MNL-SP and MNL-RP&SP models. The straightforward outputs of MNL models include estimated coefficients of attributes, valuation of attributes and market share forecasts. The benefits of smart cards can be explained by these outputs, such as when some new features were introduced into smart cards, changes of the acceptance and use of smart cards can be seen through the demand forecast. In addition to using conventional choice models (i.e., MNL models), fuzzy logic (FL) and artificial neural network (ANN), as two alternative techniques to MNL models, are introduced in this research to analyse discrete choice data. This is an exploration for improving the forecasting ability by using FL and ANN techniques. The principal motivation of using FL and ANN techniques is to make use of the non-linearity, uncertainty and linguistic description for the decision making. Furthermore, whether the different modelling mechanisms in FL and ANN models can improve the model performance will be revealed. Meanwhile, in this research, a comparison between these models (MNL, FL and ANN) is carried out to identify advantages and disadvantages on modelling and forecasting.

Finally, model application and evaluation study are conducted, which involve demand forecasting, valuation of attributes and segmentation analysis, *etc*. The purpose of segmentation analysis is to examine the heterogeneity of preference under different socio-economic backgrounds. Also, the segmentation results may suggest the future developments aiming at different groups of people. Valuation of attributes can provide perception and willingness to pay of respondents for different services, furthermore, to identify the importance of attributes of payment methods. Fare elasticities, as another evaluation measurement, can have an insight into changes of demand with respect to changes of travel cost of own or cross alternatives.

In the following chapters, the detailed contents of the research methodology are presented respectively, including the survey design, data collection, and data analysis, model applications, *etc*.

Chapter 5
Survey Design

5.1. Introduction

Since the objective of this research focuses on user demand analysis to look at the benefits of smart cards, the relevant survey was designed based on this objective. Two different survey methods: revealed preference (RP) and stated preference (SP) were designed for this research. The purpose of using RP and SP survey is to collect different preference data from individual users: respondents' actual choices and stated choices. When carrying out the evaluation study from demand side, we can make full use of advantages of two kinds of data sources: high reliability and fact validity of RP data; coverage of wide range of attributes and levels of SP data. On the other hand, through the RP and SP survey, in which payment attributes and levels were introduced, benefits of smart cards to users (measured by the importance of attributes, market share demand, valuation of attributes, etc) can be obtained.

This chapter describes the survey design of revealed preference (RP) and stated preference (SP) prior to the data collection in Chapter 6. First of all, the survey population and location are determined in Section 5.2. Section 5.3 introduces the RP survey design, including an outline of the survey design, questions being asked, and demographic questions. Following the RP survey design, the SP survey design is considered in Section 5.4. In the SP survey design, the first task is to determine the SP games being used to collect preference data from respondents. Then the relevant attributes and levels are selected to generate choice profiles by using fractional factorial design technique (Pearmain *et al*, 1991). In order to test the survey design, a pilot survey was conducted necessarily before the main survey in Section 5.5, and findings and lessons from the pilot survey suggest some modifications for finalising the survey design. Finally, Section 5.6 summarises the survey design and states the following task in the chapter of data collection.

5.2. Definition of Population

The population in the data collection is defined as all public transport passengers in Dalian urban area, China, who have access to any of three fare payment methods: cash, travel cards, smart cards, or their combinations. Based on the definition of the survey population, the following RP and SP surveys are designed on the basis of different payment means to examine different users' preferences in the modelling stage.

The main reason to select Dalian as the survey location is that different fare payment methods, such as cash, travel cards and smart cards, are widely available and used in the city. In addition, whilst it is not a reason, Dalian is my hometown. It is easier for me to get cooperation and permission from public transport companies, smart card companies and local governmental departments to conduct the data collection than any other cities in China. Personally, I have used most of the three fare payment methods for a number of years. This experience helps me design the questionnaires especially for Chinese context. Moreover, not only the pilot survey, but also the main survey once was carried out with the assistance of some friends in the city. Hence, the survey costs can be minimised.

5.3. Revealed Preference (RP) Survey Design

The RP survey is to observe data in a 'real world'. In this research, the RP survey is to identify passengers' actual choices and payment behaviour on the existing public transport fare payment methods.

5.3.1. RP Questionnaire Types

First of all, when designing the RP survey, alternatives for respondents must be determined. In this research, according to the current fare payment applications in China, three fare payment means have been chosen as the alternatives in the RP survey: cash, travel cards and smart cards. Amongst them, cash and travel cards are traditional fare payment methods. Although the smart card ticketing is right at the early stage of application in China, this new payment method is increasingly becoming one of major fare payment options. Therefore, in order to investigate and model users' revealed preferences towards these three fare payment applications in the following chapters, three different RP questionnaire versions (i.e. for cash payment, for travel card payment and for smart card payment) are designed and presented to different types of respondents. However, in these three RP questionnaire versions, basically the questions (including attitudes, behaviours, *etc*) that would be asked are the same, except that the tense and ordering of questions vary according to whom will be surveyed (in Section 5.3.3, the details are discussed).

Because we distinguish the respondent types when designing different RP questionnaires, it is necessary to understand what kinds of people should be sent the right questionnaire. The concept of "*respondent type*" in this context is defined as "those who ***primarily*** used one kind of fare payment method in the last month". By doing this the RP questionnaire is split into three different versions for primary cash users, primary travel card users and primary smart card users. Therefore, the length of questionnaire becomes shorter compared with only single RP questionnaire version containing all questions about people's actual payment behaviour towards these three payment methods, hence the survey duration for individuals could be reduced relatively. Another advantage is that it can make the questionnaires much easier to follow through some suggestive instruction. Based on this prerequisite, before surveyors send different RP questionnaire versions to different payment users, they must determine which group the respondents are in through verbal communication, and then give the suitable RP questionnaire papers.

The structure of the RP questionnaire is that: in the first section, some questions about users' actual payment behaviour for the primary payment mean are presented to respondents; following the first section, some conditional questions about the other two payment methods are included in the second and third section, respectively; the last section is about respondents' demographic backgrounds (please refer to Appendix B).

5.3.2. Variables in the RP Survey

In the RP questionnaires, the following variables (attributes) are considered by presenting relevant questions to respondents:

Cash fare type/travel card type/smart card type in current use: the role of ticket (or card) type in a certain payment method is to reveal respondents' choice behaviours towards different options available.

Travel cost: when reviewing previous studies on evaluating fare payment choices, travel cost variable was selected as one factor to identify the influence to choice behaviour. Therefore, the travel cost on the basis of Yuan (Chinese currency, 15 Chinese Yuan=1 British Pound) is also included in this RP survey. In order to reduce fatigue of respondents and save the survey duration for each single respondent, all costs in the RP survey are allowed to be estimated by respondents rather than accurate values.

Boarding time: boarding time is another attribute which can directly reflect service levels of the three payment means. Considering the differences between cash payment (cash based), travel cards and smart cards (cashless), we require respondents to answer average boarding time difference when using travel cards/smart cards relative to cash (i.e., how much quicker the boarding time can be in seconds on average compared to cash).

More trips made by using travel cards/smart cards: this is a specific question for travel cards and smart cards users to compare any changes of travel demand when using travel cards or smart cards with when using cash. Because travel cards/smart cards bring convenience and travel cost-saving (e.g., discounted fare in smart cards, unlimited trips in travel cards), asking whether more trips did (or could) happen is helpful to understand respondents' travel behavioural changes.

Seat availability: due to quicker boarding time than cash, the possibility for travel card or smart card users to get a seat on board would be higher than cash users under normal circumstances (such as normal passenger volume, on board device for fare collection in good condition, *etc*). This question can also reveal whether cashless payment methods can take advantages over cash or not.

Other functions of smart cards can be used: this question is to investigate how many extra functions of smart cards have been used or to what degree people realise they could use some existing extra functions in smart cards.

Easiness of purchasing or topping up: purchasing or topping up travel cards/smart cards is regarded as one of aspects to assess the service quality of these two cashless payments. On the other hand, smart card applications in other cities also suggested the importance of purchasing/topping up options for card users, such as for the Hong Kong Octopus card, planners put emphasis on how to provide more convenient purchasing or topping up options to card users.

Overall assessment: finally in each section of payment questions, an overall assessment about a fare payment application is presented to respondents by using five categories. Such information not only can be used in utility models, but also is one of necessary inputs in fuzzy logic/neural network analysis in the consequent chapter.

As to the detailed questions of the RP survey, please refer the RP questionnaire in Appendix B (considering the length of the thesis, only one version of the RP questionnaire is listed). In the RP questionnaire, question 1-9 in Section A are related to smart card payment, including smart card type (Q1), travel cost by smart cards (Q2), how much quicker than cash by using smart cards (Q3), whether more trips were made by using smart cards than cash (Q4), Seat availability (Q5), purchase/top-up methods (Q6), Easiness of topping up smart cards (Q7), whether other functions were used (Q8) and overall assessment (Q9).

5.3.3. Conditional Questions

Except those questions related respondents' actual fare payment (e.g. fare type, travel cost, boarding time, seat availability, overall assessment, *etc*), which the respondents used primarily in the last month, some conditional questions are designed and presented in the RP survey to identify the availability of other fare payment methods. In each section to identify the availability of one fare payment mean, a screen question is asked to determine whether the respondent need to answer this section or not. For instance, prior to questions about smart cards in the questionnaire for cash fare users, we asked a cash user: "*Could you use smart cards in the last month*?". If smart cards are available to him/her, then he/she should answer this section about smart cards (although actually he/she did not use smart cards in the last month). So in the second and third sections of the RP survey whether questions should be answered by a respondent depends on the availability of the alternatives to the respondent, rather than whether these alternatives were actually used or not. That is one of major differences between RP and SP surveys. Another reason to introduce conditional questions in the RP survey is to make a revealed choice situation and collect such preference data through one questionnaire paper. That is to say only data observation which contains two or more than two alternatives' information can be modelled and measured in the data analysis. The detailed layout for these conditional questions in one RP questionnaire is illustrated below (See Table 5.1).

The respondents need to answer four sections of questions in different RP questionnaires. Among of them, Section A and D are compulsory to every respondent, because of questions in Section A about payment respondents actually and primarily used in the last month. Section D contains demographic questions which also should be answered by respondents fully. But Section B (in RP-1, 2 and 3) and section C (in RP-1) are optional by asking screen questions to identify the availability of payment methods, which were not actually used in the last month. Due to the availability of cash payment for everyone (here we assume that cash payment is available to every respondent all the time in all public transport services), in RP-2 and RP-3, travel card users and smart card users are required to answer questions about cash payment in section C in the relevant RP questionnaire papers (Please refer the Appendix B: RP questionnaire for smart card users).

Table 5.1 The Layout of RP Questionnaires for Each Fare Payment Methods

Section / Questionnaire type	Section A	Section B	Section C	Section D
RP-1 (for cash users)	Questions about cash payment in actual use	Conditional questions about travel card payment	Conditional questions about smart card payment	Demographic questions
RP-2 (for travel card users)	Questions about travel card payment in actual use	Conditional questions about smart card payment	Conditional questions about cash payment	Demographic questions
RP-3 (for smart card users)	Questions about smart card payment in actual use	Conditional questions about travel card payment	Conditional questions about cash payment	Demographic questions

5.3.4. Combinations of Different Payment Methods

According to current fare payment applications for public transport and passengers' actual payment behaviours in China, some combinations of existing three fare payment methods (i.e. cash, travel cards and smart cards) are also taken into account when designing the questionnaires. For example, a quite common payment application in most China's cities is that travel cards are implemented within a limited service route only. Travel card (primary mean) users could use cash or smart cards (secondary mean) to pay other public transport services, in which their travel cards cannot be used, if such travel demand happens.

Questions to investigate such combinations and the relevant additional travel costs are presented with screen questions firstly. For instance, we asked "*Except using travel cards in the last month, did you need to use cash as your secondary payment mean?*". If the respondent answers "*Yes*", he/she needs to estimate the travel cost about the secondary payment mean. Through surveying fare payment combinations, travel costs can be identified so that full information of cost attribute can be collected and modelled in the data analysis to measure respondents' perceptions precisely.

However, the payment combinations also should be based on people's logical payment behaviour in practice, for example it is unlikely that a cash (primary mean) user chooses travel cards as his/her secondary payment methods, because travel cards offer more convenience and unlimited number of trips, resulting in less cost than cash fare. In total, three illogical combinations are not taken into account in the survey design (see Table 5.2.). They are: (1) Cash (main) plus travel cards; (2) Cash (main) plus smart cards; and (3) smart cards (main) plus travel cards. For (1) and (2), the reason to exclude them is because those obvious disadvantages of cash (inconvenience, more travel cost than travel cards and smart cards). Secondly, because of unavailability of travel card/smart card payment in some PT services, sometimes cash needs to be used as a secondary payment method. On the contrary, cash is always available to any users, so there is no need to use travel cards/smart cards as secondary payment method of cash. As to combination (3), it can be explained from the current smart card applications why smart cards plus travel cards are illogical. The existing smart card type in most Chinese cities is a 'pay as you go' card, which means users need to pay for each single journey with discounted fare. However, travel cards can offer unlimited number of trips once users pay monthly (or quarterly). Compared with smart cards, travel cards can save more on travel cost for frequent PT users. Therefore, passengers would more prefer using travel cards as their primary payment option. Table 5.2 lists all combinations that are considered in the RP questionnaires.

Table 5.2 Combinations of Fare Payment Methods in the RP Survey

Secondary Means / Primary Means	Cash	Travel Cards	Smart Cards
Cash	N/A	×	×
Travel Cards	√	N/A	√
Smart Cards	√	×	N/A

Note: "×" means this combination is illogical and will not be presented. "√" means the combination is logic.

5.3.5. Demographic Questions

Following questions about fare payments themselves, demographic questions are presented to the respondents in the last section of the RP survey (please refer Appendix B: Section D of the RP questionnaire), to gather demographic and attitudinal information to segment the preference data in the later data analysis. In previous studies on evaluation fare payment methods, age, sex, educational background and household income were widely used as indicator to group and identify respondents' preferences, particularly household income, which is always highly related to people's choices (Chira-Chavala and Coifman, 1996; Paynter and Law, 2003; Cheung, 2004). Demographic questions in this research include age, sex, educational level, household income, employment status. In addition to socio-economic factors aforementioned, the availability of private transport also is asked in this section in order to examine the influence of private transport to public transport fare payment choices and the relevant payment behavioural changes. Finally, considering travel cards and smart cards are presented on pre-paid basis in the current use, one attitudinal question about the willingness to prepay public transport fare is asked to respondents to get socio-psychological attitudes about pre-paid fare, though such attitudinal information that respondents provide may vary from their actual payment habits.

Table 5.3 Socio-economic Factors and Segmentations

	Segmentation
1. Age	1: 16-25; 2: 26-35; 3: 36-45; 4: 46-60; 5: over 60
2. Gender	1: Male; 2: Female
3. Educational level	1: High school or less; 2: Undergraduate student; 3: College graduate; 4: Postgraduate or equivalent
4. Employment status	1: Employed full-time; 2: Employed part-time; 3: Unemployed; 3: Student, working full or part time; 4: Student, not working; 5: Homemaker; 6: Retired
5. Household income per month	1: <￥1500; 2: ￥1500-￥2999; 3: ￥3000-￥3999; 4: ￥4000-￥5999; 5: ≥￥6000
6. Availability of private transport	1: Always; 2: Most of the time; 3: Sometimes; 4: Rarely; 5: Never or no personal vehicle
7. Willingness to prepay fare	1: Yes, I would pre pay weekly; 2: Yes, I would pre pay monthly; 3: Yes, I would pre pay quarterly; 4: No, I would not.

N.B.: ¥15 = £1

Table 5.3 lists socio-economic factors and categories for each factor in the RP questionnaires. For age factor, five categories are used to collect different perceptions from different age groups. In order to guarantee the data quality, we choose people over or equal 16 years old. The educational levels range from high school or less to postgraduate or equivalent, covering all educational backgrounds in Chinese society. Employment status is classified into 6 categories, in which students with and without works are specifically illustrated, because their regular travel and payment behaviour would influence their choices. Household income per month is classified into five levels by referring the official census data (in 2005 annual average household income in Dalian city was roughly 20000 yuan, which is equivalent to 1667 yuan per month) from Dalian Statistic Bureau (2005). Regarding the willingness to prepay, the options are based on the current choices about prepaid fare (e.g., prepay per month, prepay per quarter, etc.).

Finally, in order to understand the whole RP design based on sections discussed above. Figure 5.1 shows the flow chart of the RP survey design at the end of Section 5.3. It should be noted that this flow chart is particularly for RP-1 questionnaire version (for cash users). For RP-2 (for travel card users) and 3 (for smart card users), in general, the structure of the design is similar with that of RP-1, except that there is a small difference about Section C from RP-1. Because Section C is about cash fare payment (conditional questions for travel card or smart card users particularly) in RP-2 and RP-3, respondents must answer questions in Section C based on the assumption of the availability of cash payment to any respondents.

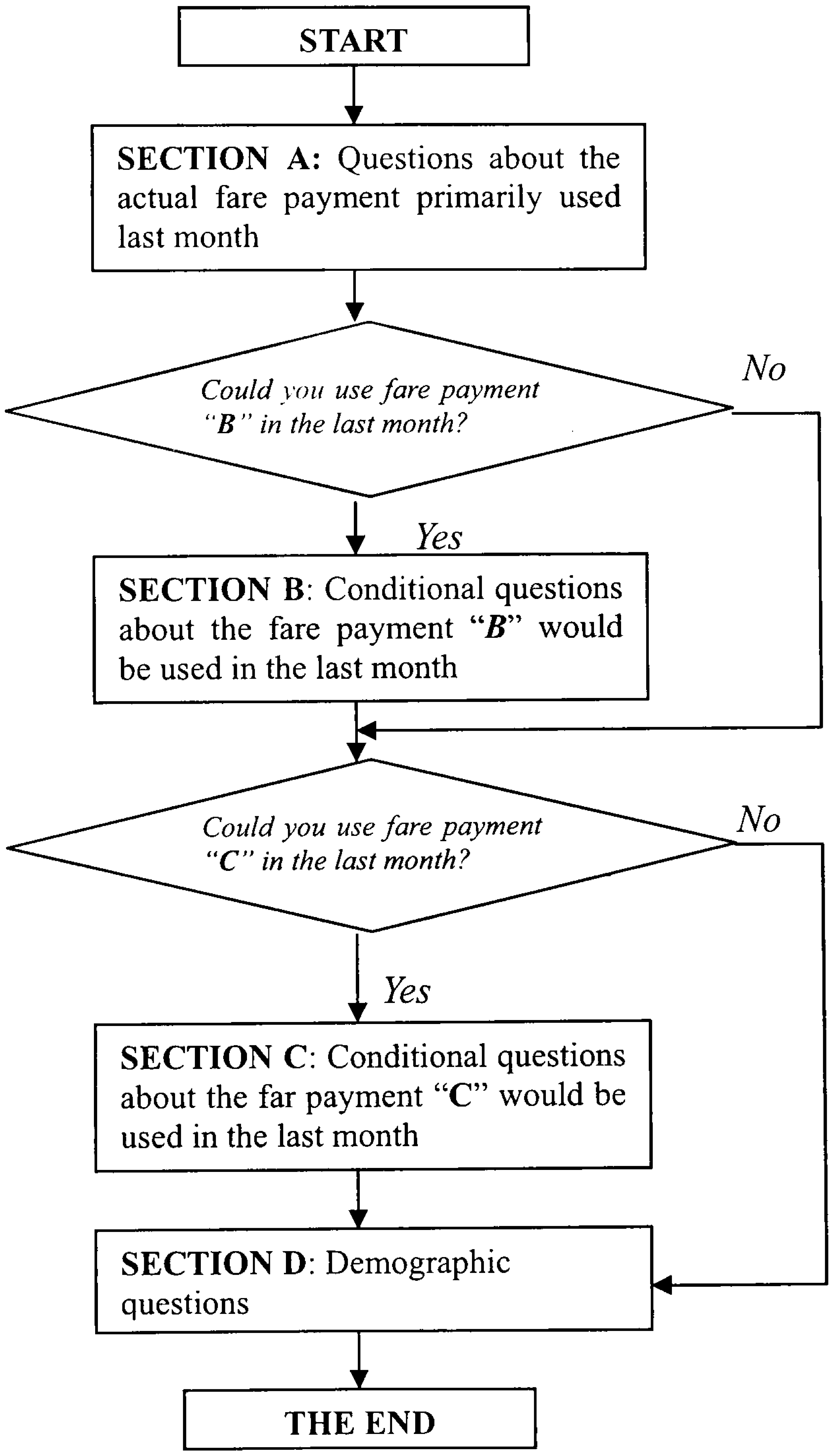


Figure 5.1 Flow Chart of the RP Survey Design

5.4. Stated Preference (SP) Survey Design

5.4.1. Determination of SP Games

The main reason for performing the SP experiments is the need to test some new applications and features on the existing fare payment means. In this SP survey, the SP experiments designed are presented in the form of trade-off games. Respondents are shown some hypothetical pair-wise situations about payment methods in one SP questionnaire and asked to select a preferred fare payment method in each situation. In this research, the SP choice context is still based on current three fare payment alternatives: cash, travel cards and smart cards, which are the same as the RP survey. But some new features and variations for existing attributes are introduced. Though three alternatives in one choice situation could be presented to respondents, in order to simplify choice tasks for respondents, the SP games are presented by binary choices.

In order to evaluate benefits and effectiveness of smart cards in this research, smart card alternative is divided into two sub-alternatives: pay as you go cards and pay monthly cards. First of all, pay as you go cards have been widely used in China, but some features can be added in the future development. Secondly, pay monthly (or weekly, quarterly) cards can be found in some other smart card applications in some countries (e.g. in South Korea, Singapore, the UK, *etc*). Moreover, considering current payment behaviours for prepaid fare in China, the main type of time-based payment is prepaid per month. Users have got more familiar with pay monthly type than weekly and quarterly type. So, monthly smart cards are considered in the SP survey.

After determining alternatives in the SP survey, the following task is to decide how many binary choice games will be used and what they are. Normally, cash, travel cards and smart cards (pay as you go and pay monthly cards) can generate six pair-wise situations in total. Among six trade-off situations, four SP games are determined to be used based on binary choices between three payment means, being sent to different fare payment users (See Table 5.4).

Table 5.4 SP Exercises and Suitability for Different Fare Payment Users

Exercise \ **Payment User Type**	**Cash users**	**Travel Card users**	**Smart Card users**
SP 1: Cash vs. Travel cards	√	√	×
SP 2: Cash vs. Smart cards (Pay as you go)	√	×	√
SP 3: Travel cards vs. Smart cards (Pay as you go)	×	√	√
SP 4: Travel cards vs. Smart cards (Pay monthly)	×	×	√

Note: "√" means the SP exercise can be presented to one kind of user type. "×" means the SP exercise is not applied to that user type.

The reasons to keep four SP games and exclude other two games in the SP survey are: (1) SP 1: cash and travel cards are widely used in our day-to-day life. The trade-off between them does exist among passengers, even before the smart card ticketing was introduced. However, smart cards (pay monthly) could achieve same function as travel cards for frequent users. Also some features and characteristics of pay monthly smart cards are similar with travel cards (such as pre-paid, cashless, fast boarding time, etc.). So finally we use cash vs. travel cards and exclude cash vs. pay monthly smart cards; (2) Even respondents would like to trade off travel cards and pay monthly smart cards, there is a direct way to compare them. That is the reason why SP 4: travel cards vs. pay monthly smart cards is kept in the SP survey; (3) as to pay as you go smart cards vs. pay monthly smart cards, the trade off can be found in SP 3: travel cards vs. pay as you go smart cards, because in SP 3 travel cards is prepaid on the monthly basis; (4) although pay as you go smart cards almost have the same payment characteristics as cash (e.g., pay for each single journey), smart cards also have some new features. Keeping SP 2 in the SP survey is helpful to identify the attractiveness of smart cards and intention of cash users to switch from cash to smart cards.

Among six pair-wise choice situations, cash vs. smart cards (pay monthly) and smart cards (pay as you go) vs. smart cards (pay monthly) are excluded. Compared with those four SP games used in the survey, reasons to exclude these two situations are:

(1) Trade-off between cash and smart cards (pay monthly) can be viewed as comparison between cash and prepayment on monthly basis. For this point, SP-1 has covered. Secondly, regarding trade-off between cash and smart cards (particularly for those specific features of smart cards), actually SP 2 has presented such situations to respondents. Since two aspects above can be covered by SP1 and SP 2, the situation of cash vs. smart cards (pay monthly) is not considered in the SP survey;

(2) Smart cards (pay as you go) and smart cards (pay monthly) share almost the same features of payment. The only differences between these two kinds of smart cards are: valid period of payment (pay monthly cards must be renewed in each month); and restriction of ridership (i.e. pay monthly cards can be used without limitation of ridership, but pay as you go card users are required to pay for each single trip). However, these two differences (features) are also considered in travel cards when trading off between travel cards and smart cards in SP 3 and SP 4. On the other hand, two different smart card types still belong to the smart card ticketing, therefore through comparing smart cards with traditional payment methods, it is more reasonable to reveal benefits of smart cards and forecast choice behavioural changes between different payment options.

Beyond determining SP games being used, another important task is to understand proper SP games should be sent to proper respondents in the sample. We expect that the respondents who get SP games have got familiar with two alternatives (or at least one alternative) in the trade off situation such that their trade-offs and stated choices could make sense and further data analysis can be reasonable. As can be seen in Table 5.4, SP 1 (cash vs. travel cards) can be presented to current cash or travel card users. In the same way, SP2, SP3 and SP 4 can only be shown to respondents who used or understand at least one alternative in two payment methods. To distinguish proper user types, prior to SP games, a screen question firstly is asked to identify this. Another reason to send SP experiments according to the existing user types is that it can reflect users' consumption psychology. For example, for current cash users, to ask them to answer the SP games about the trade-off between cash and other two payment means is feasible and reasonable, because they might more concern on how much better (or worse) the other two cashless payment methods could be when comparing with their current choice (i.e. cash fare payment).

5.4.2. Determination of Attributes

Each alternative in the choice set is characterised by a set of attributes. The definition of the variables of interests (attributes) is the second task in this SP survey design after determining alternatives being studied. It involves the selection of the attributes considered the most important ones (Bradley, 1988) for the decision making process and also some new ones that are related transport policy. When determining what kinds of attributes and how many attributes will be introduced in SP experiments, according to those previous experiences (Pearmian and Kroes, 1991), it is advisable to limit the number of attribute to avoid confusing respondents, although a particular SP design allows several attributes to be presented. So, in this research, in order to select proper attributes to respondents, several aspects should be taken into account:

a). the choice context must be taken into account. In this research, it is based on public transport fare payment method choices. All attributes about fare payments must be familiar and simple to any respondents (public transport users).

b). Secondly, one of important aspects for this research is the user demand forecast for different payment applications, particularly for smart cards, such a new application in public transport services. So proper attribute selection can be helpful to target, compare and predict the future development of fare payment applications.

c). all attributes being used in the SP survey design should be important (determinant) factors to affect passengers' choices toward public transport fare payment methods.

Table 5.5 lists the attribute selection of the SP survey. And Table 5.6 presents the details about attributes and levels in each SP game. Reasons and meaning about each attribute in the SP games can be explained as follows:

(1) *Travel cost*: travel cost is a very direct measurement to influence people's choices in transport studies (such as mode choice, route choice. *etc*) Without exception, this factor also is considered in this research, which is related to travel costs by using different payment methods. For cash payment, passengers need to pay for single trip each time. For travel cards, fare is paid monthly/quarterly for a given PT route service with unlimited number of trips during the period of time (one month or one quarter). For smart cards, discounted fare policy is applied (10%-20% off relative to standard cash fare). Another kind of smart card is pay monthly cards with the similar fare structure as travel cards.

(2) *Boarding time difference*: the motivation of selecting the boarding time difference as one of attributes stems from the difference between cash and cashless payment features. If we provided boarding time to passengers directly, it seems difficult to feel their average boarding time by different payment methods in hypothetical situations. However, if only giving them average boarding time difference by using cash and cashless payment (i.e., travel cards/smart cards), it is easier to understand and judge the performance of different payments. Moreover, previous studies also used the boarding time difference to measure smart cards (Chira-Chavala and Cofiman, 1996).

(3) *Public transport services covered by using certain fare payment method*: in China, most travel card applications can only be used on a single PT route rather than on the basis of zone as in the UK, for instance. Secondly, a similar problem of PT service routes covered also could happen on smart cards (e.g. the application in Hang Zhou, China). This would result in the increase of the travel cost for card users or affect passengers' choices of travel cards/smart cards.

(4) *Whether passengers can get change if they cannot pay the exact fare*: this is specific for cash fare, because cash fare is collected by fare boxes in Dalian, China, normally drivers are not allowed to handle cash in person for security reason. If passengers pay big value money instead of exact fare, it is highly possible that they cannot get change back. Some passengers complain about this policy, and PT operators also try to find some solution to make passengers satisfied.

(5) *Deposit*: since the smart card ticketing was firstly implemented in Dalian in 2001, deposit for initial smart card purchase has been required as compensation to smart card operators when cards were damaged. So this attribute is kept in the SP survey, but variations are given based on the existing application.

(6) *Overdraft*: small amount of overdraft is allowed to guarantee the last fare payment for passengers when credit in a card failed to pay a ticket. Although this facility has not been implemented in Dalian, we can find such application in Beijing (ITS China, 2004) and New York (Savage, 2000).

(7) *Multifunction*: This has been widely implemented in those successful smart card projects, such as the Hong Kong Octopus card (Chambers, 1998). From users' point of view, they also would like to see more extra services (e.g., shopping, banking, telecommunication, etc.) in one card to enhance the smart card use.

(8) *Geographic areas covered*: in most initial stages of smart card applications, smart cards can only be used in a given geographic area, resulting in inconvenience if card users travelling to areas smart cards cannot cover. Such interoperation of smart cards among some neighbouring cities is one of development directions for smart cards. In Shanghai and Suzhou, two neighbouring cities, the application has been under practice. In addition to smart cards, a similar situation can also be found in travel card applications.

(9) *Top-up/purchase methods*: Convenience for topping up or purchasing travel cards/smart cards has become another important aspect to measure the service quality of these payment methods. Previous experiences suggest that a variety of top-up/purchase methods could influence passengers to choose those payment means (Chambers, 1998; Paynter and Law, 2000).

Table 5.5 Attributes in the SP Experimental Design

Attributes	Alternatives	Variable Types
Travel cost	C, TC and SC	Generic term
Boarding time	C, TC and SC	Generic term
Public transport services covered by using certain fare payment method	TC and SC	Alternative specific term
Whether passengers can get change if they cannot pay exact money	C	Alternative specific term
Deposit	SC	Alternative specific term
Overdraft	SC	Alternative specific term
Multifunction	SC	Alternative specific term
Geographic areas covered	TC and SC	Alternative specific term
Top-up/purchase methods	TC and SC	Alternative specific term

Note: C—cash; TC—travel cards; SC—smart cards

As can be seen in Table 5.5, besides generic attributes (i.e. travel cost), some alternative specific attributes are selected for different payment methods, such as deposit, overdraft, multifunction, *etc* for the smart card ticketing, because some attributes are only applied in certain payment means rather than all alternatives in the SP survey. But as features different from other payments, these specific attributes are introduced and such information can be viewed as factors when respondents trade off between two alternatives. Among these SP attributes, travel cost, boarding time and deposit are quantitative. Others are qualitative variables that will be transferred to dummy variables when estimating the model. Also, it is worth noting that most SP attributes are derived from actual practice although the SP survey is based on hypothetical situations.

5.4.3. Determination of Attribute Levels

The next step of the SP survey design involves the establishment of the levels or variations of each attribute considered. These levels must include all situations that the respondent would be faced with during the experiment. On the other hand, the definitions for the attribute levels must be realistic and where possible related to the respondents' experience.

Generally, the attribute levels used in the experimental design are defined as variations relative to the attribute levels of an existing fare payment method, but for smart cards, because some new attributes are used, selection of relevant levels has to refer applications in other cities or countries, which have applied similar attributes (e.g. overdraft for smart cards, multifunction, geographic areas can be used, *etc*). The number of attributes considered, together with the various levels that each attribute will present, will define the size of the experiment. The experiments are normally designed as "orthogonal", that is, the attributes presented to the respondent will vary independently of each order (Pearmain and Kroes, 1990), avoiding multi-colinearity between attributes.

The detailed attribute levels for each SP exercise are presented in Table 5.6. The rules to determine these levels for each attribute can be summarised as follows:

(1) *Travel cost*: is based on the existing fare levels of three payment methods. Meanwhile, some variations are given. In this SP survey, travel costs for all payment methods are allocated four levels, because more variations of attributes can produce more trade-offs, so as to obtain better results of choice behaviour than three or two levels. As can be seen in Table 5.6-1, cash fare levels are 0.8yuan, 1yuan, 1.2yuan and 1.4yuan for a single ticket. 1yuan is the current single ticket cost. 0.2yuan variation between two nearby levels is given in order to make comparison with smart card fare. For example, 0.8yuan can be compared with discounted fare of smart cards in some SP games. On the other hand, increased cash fare can identify people's responses when some relevant services have been improved. For travel cards, 29yuan, 37yuan, 45yuan and 53yuan are given, in which 29yuan is the lowest pay monthly travel card cost for some bus routes in current use, thus is kept in the SP game. 8yuan difference is applied to cover a full range of monthly travel card costs in reality so that respondents can easily compare travel costs between travel cards and some other payment methods. For pay as you go smart cards, because discounted fare policy has been applied in practice, the maximum single fare should not be more than cash fare. So 1yuan is set as the top value for smart cards. The same variation (0.2yuan) as cash

fare is given. For pay monthly smart cards, because they can be used on any urban PT services, values in four levels are set higher than relevant travel card cost levels, which can be viewed as a compensation for a better service (better service could cost more).

(2) *Average boarding time difference*: is only used to compare cash with cashless payment (travel cards/smart cards). Chira-Chavala and Cofiman (1996) found the average boarding time difference by using cash and cashless payment ranged from 20seconds to 60 seconds roughly. So in this research, the average boarding time difference is given a slightly bigger range from 20s to 80s with four levels, 20s variation between two nearby levels.

(3) *Public transport services covered by using certain fare payment method*: because cash can be used anytime for any routes, we only consider services covered by travel cards and smart cards. Four levels also are set for travel cards and smart cards. The current travel cards in China can be used in limited PT routes, so this level is kept as the lowest level in the services covered by using travel cards. Unlimited the service policy has been widely implemented in some countries, such as the UK, Germany, *etc*, so with the increase of subsidy for public transport in China (Chen and Mao, 2003), the unlimited route policy would be implemented as an attraction of public transport services. Therefore, the other three levels of travel cards are set as unlimited routes. Meanwhile, the difference of extra cost is given to identify respondents' willingness to pay for such unlimited route services. For pay monthly smart cards, all four levels are presented by unlimited route services, but by using extra charges to distinguish these four levels.

(4) *Deposit*: according to the existing deposit standard for smart cards in Dalian, 50yuan is set as the top value. Also 30 yuan and 20 yuan for smart card deposit can be found in some other Chinese cities, such as Beijing, Shenyang. So 20yuan and 30yuan also are given as two levels. In recent years, in Beijing, Dalian, Shanghai, some smart card users questioned local PT operators about the purpose of deposit of smart cards. Recently the suggestion to cancel smart card deposit is becoming stronger and stronger. For this reason, 0yuan deposit is introduced as the lowest level in this research to see people's response.

(5) *Multifunction*: the content of multifunction in the SP survey is referred to those successful smart card applications, which have implemented multifunction in smart cards, particularly the Hong Kong Octopus card. Multifunction in this research covers shopping, telecommunication, entertainment, banking, parking fee and tolling payment.

Table 5.6-1 Exercise 1: Cash vs. Travel Cards

Levels / Attributes	Levels							
	Cash				**Travel Cards**			
	0	**1**	**2**	**3**	**0**	**1**	**2**	**3**
1. Travel costs	0.8yuan per ride	1yuan per ride	1.2yuan per ride	1.4yuan per ride	29yuan per month	37yuan per month	45yuan per month	53yuan per month
2. Boarding time	Average 20 second slower than travel cards	Average 40 seconds slower than travel cards	Average 60 seconds slower than travel cards	Average 80 seconds slower than travel cards	Straight getting on	Straight getting on	Straight getting on	Straight getting on
3. Public Transport services covered	--	--	--	--	Limited routes: Only one bus, or light rail route service in urban area	Unlimited routes without any extra charges	Unlimited routes with extra charge: 10% more than limited services	Unlimited routes with extra charge: 15% more than limited services
4. Whether passengers can get change if using fare boxes	No	Yes	--	--	--	--	--	--

Table 5.6-2: Exercise 2: Cash vs. Smart Cards ("Pay as you go" cards)

Levels / Attributes	Levels							
	Cash				Smart Cards (Pay as you go cards)			
	0	1	2	3	0	1	2	3
1. Travel costs	0.8yuan per ride	1yuan per ride	1.2yuan per ride	1.4yuan per ride	0.4yuan per ride	0.6yuan per ride	0.8yuan per ride	1yuan per ride
2. Boarding time	Average 20 second slower than smart cards	Average 40 seconds slower than smart cards	Average 60 seconds slower than smart cards	Average 80 seconds slower than smart cards	Straight getting on	Straight getting on	Straight getting on	Straight getting on
3. Whether passengers can get change	No	Yes	--	--	--	--	--	--
4. Deposit	--	--	--	--	0yuan	20yuan	30yuan	50yuan
5. Overdraft (Ex. 2-1)	--	--	--	--	No	Yes	--	--
6. Multifunction (Ex. 2-2)	--	--	--	--	No. only for public transport	Shopping, telephone, entertainment	Shopping, telephone, entertainment ,parking and tolling	Shopping, telephone, entertainment ,parking, tolling and banking
7. Geographic areas covered (Ex. 2-3)	--	--	--	--	Only in Dalian urban area	Dalian Urban and rural areas	Dalian and other nearby cities	Within one province
8. Top-up methods (Ex. 2-4)	--	--	--	--	Only at ticket offices	Ticket offices, banks, agencies	Ticket offices, banks, agencies, self-adding value machine	Ticket offices, banks, agencies, self-adding value machine, telephone and Internet

Table 5.6-3: Exercise 3: Travel Cards vs. Smart Cards

Levels / Attributes	Levels							
	Travel Cards				Smart Cards ("Pay as you go" cards)			
	0	1	2	3	0	1	2	3
1. Travel costs:	29yuan per month	37yuan per month	45yuan per month	53yuan per month	0.4yuan per ride	0.6yuan per ride	0.8yuan per ride	1yuan per ride
2. Public transport services covered	Limited routes: only used in one bus or light rail route in urban area	Unlimited routes without extra charge	Unlimited routes with extra charge: 10% more than limited services	Unlimited services with 15% extra charge	Any public transport services	Any public transport services	Any public transport services	Any public transport services
3. Deposit	--	--	--	--	0yuan	20yuan	30yuan	50yuan
4. Multifunction (Ex.3-1)	--	--	--	--	No. only for public transport	Shopping, telephone, entertainment	Shopping, telephone, entertainment , parking and tolling	Shopping, telephone, entertainment , parking, tolling and banking
5. Overdraft (Ex.3-2)	--	--	--	--	No	Yes	--	--
6. Top-up methods (Ex.3-3)	Ticket offices	Ticket offices and agencies	Ticket offices and agencies and banks	--	Ticket offices and agencies	Ticket offices, agencies and banks	Ticket offices, agencies, banks, telephone and online	--

Table 5.6-4: Exercise 4: Travel Cards vs. Smart Cards

Levels / Attributes	Levels							
	Travel Cards				Smart Cards (pay monthly cards)			
	0	1	2	3	0	1	2	3
1. Travel costs:	29yuan per month	37yuan per month	45yuan per month	53yuan per month	32yuan per month	40yuan per month	48yuan per month	56yuan per month
2. Public transport services covered	Limited routes: Only one bus, or light rail route service in urban area	Unlimited routes without any extra charges	Unlimited routes with extra charge: 10% more than limited services	Unlimited routes with extra charge: 15% more than limited services	Unlimited routes (any services in urban area) without any extra charges	Unlimited routes with extra charge: 10% more than limited services	Unlimited routes with extra charge: 15% more than limited services	Unlimited routes with extra charge: 20% more than limited services
3. Deposit	--	--	--	--	0yuan	20yuan	30yuan	50yuan
4. Multifunction (Ex.4-1)	--	--	--	--	No. only for public transport	Shopping, telephone, entertainment	Shopping, telephone, entertainment ,parking and tolling	Shopping, telephone, entertainment ,parking, tolling and banking
5. Geographic areas covered (Ex.4-2)	Only in Dalian urban area	Only in Dalian urban area	Only in Dalian urban area	Only in Dalian urban area	Only in Dalian urban area	Dalian Urban and rural areas	Dalian urban and rural areas and other nearby cities	Within one province

(6) *Geographic areas covered*: the current travel cards and smart cards can only cover very limited geographic areas. The variations of this attribute include: "only Dalian urban area", "Dalian urban and rural areas", "Dalian and nearby cities" and "the whole province", four levels. In order to distinguish the differences among these levels, geographic areas that can be covered become bigger and bigger from the lowest level to the highest one.

(7) *Top-up/purchase methods*: card users would like to see more options to top up or purchase their cards. Also the detailed applications for this attribute are based on the existing applications in other travel cards/smart cards projects, such as ticket offices, agencies, banks, mobile phone and online top-up. In order to distinguish the differences among these levels, those methods are added gradually from the lowest level to the highest one.

As to *Whether passengers can get change if they cannot pay the exact fare* and *overdraft*, two levels are applied (Yes/No). For the former attribute, the current application of farebox for cash fare requires that on board PT drivers do not pay back change if passengers cannot pay exact cash/coin. This has resulted in complains due to inconvenience for cash users. Therefore, binary option ('Yes/No') for this attribute is designed. *Overdraft* in this survey is included to guarantee the last trip payment for passengers. In the current application, overdraft has not been introduced in Dalian, so 'Yes/No' pattern of overdraft is presented instead of the detailed values about overdraft limit.

5.4.4. Fractional Factorial Design and Choice Set Generation

After determining attributes and attribute levels in this SP survey, profiles for choice-based experiments should be generated. A choice profile (or a choice situation) can be defined as a combination of attributes and levels included in the experiment in the way that they are completely uncorrelated between alternatives to permit rigorous testing of certain hypotheses of interest (Louviere, Hensher and Swait, 2000). In the SP survey, for example, SP-1 in Table 5.6-1, cash vs. travel cards, cash payment has four attributes: travel cost, boarding time difference, PT services covered and whether passengers can get change if they cannot pay exact fare, each attribute has been allocated relevant levels (or values), then one of cash payment situation could be described as follows by four attributes:

"(1) *0.8yuan per ride on **travel cost***;

(2) *average 20 seconds slower than cashless payment on **boarding time***;

(3) *it can be used on any **PT mode*** and;

(4) *passengers may **get change back** if they cannot pay exact cash fare*".

One of solutions to generate choice profiles is full factorial design. In the full factorial design, all possible combinations are included. However, respondents can only evaluate a fairly limited number of alternatives at a time, so a design incorporating all possible combinations of all levels of each attribute can only be used if there are very few attributes and levels. In this SP survey, for example, there are five attributes in SP 1 exercise (cash vs. travel cards). Among them, 4 attributes have 4 levels (travel cost of cash, travel cost of travel cards, boarding time, PT services covered) and one attribute with 2 levels (Yes/No for '*whether passengers can get change back*'). The full factorial design generates 512 choice situations (=4×4×4×4×2), apparently too many profiles to trade off for respondents

When a full factorial design generates too many choice profiles, the number can be reduced by adopting a "fractional factorial design" (Pearmain *et al*, 1991) so that only a selection of all possible combinations is presented to the respondents. The main characteristics of fractional factorial design technique are (1) all combinations generated are 'orthogonal'; (2) all selected combinations vary independently from one another. Based on the cook book (Kocur *et al*, 1982) of the fractional factorial SP design method, 16 profiles (pair-wise situations) are set up with respect to the combinations of different levels of attributes in the SP games. In order to simplify the SP design, firstly fractional factorial 'skeleton' designs for each SP game are generated (see Table 5.8), in which each attribute level is allocated to generate a choice profile. The second step for producing choice situations is to split these 16 pair-wise profiles into two sub-sets (known as 'block design') with the random selection procedure, containing 8 profiles in each set. However, it is worth noting that in order to combine the responses in difference sub-sets, we assume that the preferences across the samples of the respondents are sufficiently homogeneous.

Among the four SP games, SP1 (cash vs. travel cards) and SP3 (travel cards vs. pay as you go smart cards) are designed by absolute value of travel cost, because in these two games, units of travel cost are naturally different (Yuan per ride and Yuan per month), therefore original cost values for two different costs are kept in the survey design. However, in order to keep the number of replications/choices low, it is decided to go for the difference design in SP 2 (cash vs. pay as you go smart cards) and SP 4 (travel cards vs. pay monthly smart cards). Other two reasons for using the difference design for SP2 and SP4 are: partly because we do not know the current cost level (absolute value); and partly because in SP2 and SP4, the travel cost for two alternatives is designed as small number (Yuan per ride), which respondents have trouble with when trading off.

Table 5.7 Travel Cost Difference in SP 2 and SP 4

	Actual Levels	Cost Difference
SP 2	***Cash***: 0.8yuan; 1yuan; 1.2yuan; 1.4yuan ***Smart cards***: 0.4yuan; 0.6yuan; 0.8yuan; 1yuan	0yuan; 0.2yuna; 0.4yuan; 0.6yuan; 0.8yuan; 1yuan
SP4	***Travel cards***: 29yuan; 37yuan; 45yuan; 53yuan ***Smart cards***: 32yuan; 40yuan; 48yuan; 56yuan	-27yuan; -19yuan; -11yuan; -3yuan; +5yuan; +13yuan; +21yuan

In the attributes of SP 2 and SP 4, travel cost is designed as the difference between two alternatives, because different from SP1 and SP3, the cost units for two alternatives in SP2 and 4 are the same (i.e. Yuan per month). According to cost levels we set for each alternative in SP 2 and SP 4, possible levels of cost difference can be got as listed in Table 5.7. The cost difference between cash and smart cards in SP 2 ranges from 0yuan to 1yuan. In SP 4, the maximum absolute value of the cost difference is 27yuan and the minimum is 3yuan. For the travel cost difference, in order to make the values of different levels sensible, we keep four levels in the middle of these full differences. If the difference is too big, it would be easy for respondents to trade off on the basis of travel cost. So in SP 2, 0yuan, 0.2yuan, 0.4yuan and 0.6yuan are kept, and in SP 4, -19yuan, -11yuan, -3yuan and +5yuan are used. Especially for SP 4, in order to present smart card applications applying better services with higher costs, most differences are negative, which means smart cards could cost more than travel cards due to better services. Regarding the boarding time difference, we can regard the difference as how much quicker by using travel cards/smart cards than cash, because we assume cashless payment users can get on PT vehicles directly because of the convenient check-in process, while passengers would have to prepare for exact cash and insert cash/coin into fare boxes.

Also, as can be seen in Table 5.6, the number of attributes in some SP games is more than 5 (i.e. SP 2, 3 and 4). If all of these attributes were designed in one binary choice experiment, the number of choice profiles would become extremely large and it would also be difficult to trade off and manage for respondents and surveyors. For example, if 8 attributes (See Table 5.6-2) in SP 2 were used at the same time, more than 25 choice profiles could be generated. It is apparently too hard that respondents can trade off so many situations. Therefore, Exercise 2, 3 and 4 are split into some sub-exercises so that it may reduce the complexity of the survey design and fatigue of respondents when they trade off different alternatives. After splitting, attributes used for each sub-exercise can be seen from Table 5.6-2, 5.6-3 and 5.6-4. Some of attributes are regarded as common and principal attributes and are included in all separate exercises in the same SP game, to allow comparison of relative preferences over all attributes being investigated. The rest of attributes can be combined into these attributes as an additional one each time to generate different sub-exercises. For example, in total there are eight attributes in Exercise 2 (See Table 5.6-2). Because travel cost and boarding time can be used to measure value of boarding time and value of other attributes in the later model analysis, these two attributes are viewed as common attributes. Other two alternative specific attributes, *whether passengers can get change* for cash payment and *deposit* for smart cards payment, are also included in common attributes, which are related to monetary value. Finally four of them (travel cost, boarding time, whether passengers can get changes and deposit) are regarded as

Table 5.8-1 Skeleton Design of SP 1: Cash vs. Travel Cards

Alternatives	Cash			Travel Cards		
Attributes	**Travel cost**	**Boarding time**	**Change**	**Travel cost**	**Boarding time**	**Services**
1	0.8yuan	Average 20 second slower than travel cards	No	29yuan	Straight getting on	Limited route: Only one bus or light rail route covered in urban area
2	0.8yuan	Average 40 second slower than travel cards	No	37yuan	Straight getting on	Unlimited route with 10% extra charge more
3	0.8yuan	Average 60 second slower than travel cards	Yes	45yuan	Straight getting on	Unlimited route with 15% extra charge more
4	0.8yuan	Average 80 second slower than travel cards	Yes	53yuan	Straight getting on	Unlimited route: any public transport services without extra charges
5	1yuan	Average 40 second slower than travel cards	Yes	29yuan	Straight getting on	Unlimited route: any public transport services without extra charges
6	1yuan	Average 20 second slower than travel cards	Yes	37yuan	Straight getting on	Unlimited route with 15% extra charge more
7	1yuan	Average 80 second slower than travel cards	No	45yuan	Straight getting on	Unlimited route with 10% extra charge more
8	1yuan	Average 60 second slower than travel cards	No	53yuan	Straight getting on	Limited route: Only one bus or light rail route covered in urban area
9	1.2yuan	Average 60 second slower than travel cards	Yes	29yuan	Straight getting on	Unlimited route with 10% extra charge more
10	1.2yuan	Average 80 second slower than travel cards	Yes	37yuan	Straight getting on	Limited route: Only one bus or light rail route covered in urban area
11	1.2yuan	Average 20 second slower than travel cards	No	45yuan	Straight getting on	Unlimited route with 15% extra charge more
12	1.2yuan	Average 40 second slower than travel cards	No	53yuan	Straight getting on	Unlimited route: any public transport services without extra charges
13	1.4yuan	Average 80 second slower than travel cards	No	29yuan	Straight getting on	Unlimited route with 15% extra charge more
14	1.4yuan	Average 60 second slower than travel cards	No	37yuan	Straight getting on	Unlimited route: any public transport services without extra charges
15	1.4yuan	Average 40 second slower than travel cards	Yes	45yuan	Straight getting on	Limited route: Only one bus or light rail route covered in urban area
16	1.4yuan	Average 20 second slower than travel cards	Yes	53yuan	Straight getting on	Unlimited route with 10% extra charge more

Table 5.8-2: Skeleton Design of SP 2: Cash vs. Smart Cards (Pay as you go cards)

Ex. 2-1

Alternatives		Cash		Smart Cards		
Attributes	**Travel cost difference (Cash-Smart cards)**	**Boarding time**	**Change**	**Boarding time**	**Deposit**	**Overdraft**
1	0yuan	Average 20 second slower than travel cards	No	Straight getting on	0yuan	No
2	0yuan	Average 40 second slower than travel cards	Yes	Straight getting on	20yuan	No
3	0yuan	Average 60 second slower than travel cards	Yes	Straight getting on	30yuan	Yes
4	0yuan	Average 80 second slower than travel cards	No	Straight getting on	50yuan	Yes
5	0.2yuan	Average 20 second slower than travel cards	Yes	Straight getting on	20yuan	Yes
6	0.2yuan	Average 40 second slower than travel cards	No	Straight getting on	0yuan	Yes
7	0.2yuan	Average 60 second slower than travel cards	No	Straight getting on	50yuan	No
8	0.2yuan	Average 80 second slower than travel cards	Yes	Straight getting on	30yuan	No
9	0.4yuan	Average 20 second slower than travel cards	No	Straight getting on	30yuan	Yes
10	0.4yuan	Average 40 second slower than travel cards	Yes	Straight getting on	50yuan	Yes
11	0.4yuan	Average 60 second slower than travel cards	Yes	Straight getting on	0yuan	No
12	0.4yuan	Average 80 second slower than travel cards	No	Straight getting on	20yuan	No
13	0.6yuan	Average 20 second slower than travel cards	Yes	Straight getting on	50yuan	No
14	0.6yuan	Average 40 second slower than travel cards	No	Straight getting on	30yuan	No
15	0.6yuan	Average 60 second slower than travel cards	No	Straight getting on	20yuan	Yes
16	0.6yuan	Average 80 second slower than travel cards	Yes	Straight getting on	0yuan	Yes

Table 5.8-3: Skeleton Design of SP 2: Cash vs. Smart Cards (Pay as you go cards)

Ex. 2-2

Alternatives		**Cash**		**Smart Cards**		
Attributes	**Travel cost difference (Cash-SC)**	**Boarding time**	**Change**	**Boarding time**	**Deposit**	**Multifunction**
1	0yuan	Average 20 second slower than travel cards	No	Straight getting on	0yuan	No. only for public transport
2	0yuan	Average 40 second slower than travel cards	No	Straight getting on	20yuan	Shopping, telephone, entertainment ,parking and tolling
3	0yuan	Average 60 second slower than travel cards	Yes	Straight getting on	30yuan	Shopping, telephone, entertainment ,parking, tolling and banking
4	0yuan	Average 80 second slower than travel cards	Yes	Straight getting on	50yuan	Shopping, telephone, entertainment
5	0.2yuan	Average 20 second slower than travel cards	Yes	Straight getting on	20yuan	Shopping, telephone, entertainment
6	0.2yuan	Average 40 second slower than travel cards	Yes	Straight getting on	0yuan	Shopping, telephone, entertainment ,parking, tolling and banking
7	0.2yuan	Average 60 second slower than travel cards	No	Straight getting on	50yuan	Shopping, telephone, entertainment ,parking and tolling
8	0.2yuan	Average 80 second slower than travel cards	No	Straight getting on	30yuan	No. only for public transport
9	0.4yuan	Average 20 second slower than travel cards	Yes	Straight getting on	30yuan	Shopping, telephone, entertainment ,parking and tolling
10	0.4yuan	Average 40 second slower than travel cards	Yes	Straight getting on	50yuan	No. only for public transport
11	0.4yuan	Average 60 second slower than travel cards	No	Straight getting on	0yuan	Shopping, telephone, entertainment
12	0.4yuan	Average 80 second slower than travel cards	No	Straight getting on	20yuan	Shopping, telephone, entertainment ,parking, tolling and banking
13	0.6yuan	Average 20 second slower than travel cards	No	Straight getting on	50yuan	Shopping, telephone, entertainment ,parking, tolling and banking
14	0.6yuan	Average 40 second slower than travel cards	No	Straight getting on	30yuan	Shopping, telephone, entertainment
15	0.6yuan	Average 60 second slower than travel cards	Yes	Straight getting on	20yuan	No. only for public transport
16	0.6yuan	Average 80 second slower than travel cards	Yes	Straight getting on	0yuan	Shopping, telephone, entertainment ,parking and tolling

Note: The 'multifunction' column is changed to 'Geographic area covered' for Ex.2-3 and 'top-up methods' for Ex.2-4, following the same coding scheme

Ex.3-1

Table 5.8-4: Skeleton Design of SP 3: Travel Cards vs. Smart Cards ("Pay as you go" cards)

Alternatives	Travel Cards		Smart Cards		
Attributes	Travel cost	Services	Travel cost	Deposit	Multifunction
1	29yuan	Limited route: Only one bus or light rail route covered in urban area	0.4yuan	0yuan	No. only for public transport
2	29yuan	Unlimited route: any public transport services without extra charges	0.6yuan	30yuan	Shopping, telephone, entertainment ,parking, tolling and banking
3	29yuan	Unlimited route with 10% extra charge more	0.8yuan	50yuan	Shopping, telephone, entertainment
4	29yuan	Unlimited route with 15% extra charge more	1yuan	20yuan	Shopping, telephone, entertainment ,parking and tolling
5	37yuan	Unlimited route: any public transport services without extra charges	0.4yuan	20yuan	Shopping, telephone, entertainment
6	37yuan	Limited route: Only one bus or light rail route covered in urban area	0.6yuan	50yuan	Shopping, telephone, entertainment ,parking and tolling
7	37yuan	Unlimited route with 15% extra charge more	0.8yuan	30yuan	No. only for public transport
8	37yuan	Unlimited route with 10% extra charge more	1yuan	0yuan	Shopping, telephone, entertainment ,parking, tolling and banking
9	45yuan	Unlimited route with 10% extra charge more	0.4yuan	30yuan	Shopping, telephone, entertainment ,parking and tolling
10	45yuan	Unlimited route with 15% extra charge more	0.6yuan	0yuan	Shopping, telephone, entertainment
11	45yuan	Limited route: Only one bus or light rail route covered in urban area	0.8yuan	20yuan	Shopping, telephone, entertainment ,parking, tolling and banking
12	45yuan	Unlimited route: any public transport services without extra charges	1yuan	50yuan	No. only for public transport
13	53yuan	Unlimited route with 15% extra charge more	0.4yuan	50yuan	Shopping, telephone, entertainment ,parking, tolling and banking
14	53yuan	Unlimited route with 10% extra charge more	0.6yuan	20yuan	No. only for public transport
15	53yuan	Unlimited route: any public transport services without extra charges	0.8yuan	0yuan	Shopping, telephone, entertainment ,parking and tolling
16	53yuan	Limited route: Only one bus or light rail route covered in urban area	1yuan	30yuan	Shopping, telephone, entertainment

Table 5.8-5: Skeleton Design of SP 3: Travel Cards vs. Smart Cards ("Pay as you go" cards)

Ex. 3-2

Alternatives	Travel Cards		Smart Cards		
Attributes	**Travel cost**	**Services**	**Travel cost**	**Deposit**	**Overdraft**
1	29yuan	Limited route: Only one bus or light rail route covered in urban area	0.4yuan	0yuan	No
2	29yuan	Unlimited route: any public transport services without extra charges	0.6yuan	30yuan	No
3	29yuan	Unlimited route with 10% extra charge more	0.8yuan	50yuan	Yes
4	29yuan	Unlimited route with 15% extra charge more	1yuan	20yuan	Yes
5	37yuan	Unlimited route: any public transport services without extra charges	0.4yuan	20yuan	Yes
6	37yuan	Limited route: Only one bus or light rail route covered in urban area	0.6yuan	50yuan	Yes
7	37yuan	Unlimited route with 15% extra charge more	0.8yuan	30yuan	No
8	37yuan	Unlimited route with 10% extra charge more	1yuan	0yuan	No
9	45yuan	Unlimited route with 10% extra charge more	0.4yuan	30yuan	Yes
10	45yuan	Unlimited route with 15% extra charge more	0.6yuan	0yuan	Yes
11	45yuan	Limited route: Only one bus or light rail route covered in urban area	0.8yuan	20yuan	No
12	45yuan	Unlimited route: any public transport services without extra charges	1yuan	50yuan	No
13	53yuan	Unlimited route with 15% extra charge more	0.4yuan	50yuan	No
14	53yuan	Unlimited route with 10% extra charge more	0.6yuan	20yuan	No
15	53yuan	Unlimited route: any public transport services without extra charges	0.8yuan	0yuan	Yes
16	53yuan	Limited route: Only one bus or light rail route covered in urban area	1yuan	30yuan	Yes

Ex. 3-3

Table 5.8-6: Skeleton Design of SP 3: Travel Cards vs. Smart Cards ("Pay as you go" cards)

Alternatives	Travel Cards			Smart Cards	
Attributes	**Travel cost**	**Services**	**Top-up Methods**	**Travel cost**	**Top-up Methods**
1	29yuan	Limited route: Only one bus or light rail route covered in urban area	Ticket offices	0.4yuan	Ticket offices and agencies
2	29yuan	Unlimited route: any public transport services without extra charges	Ticket offices and agencies and banks	0.6yuan	Ticket offices, agencies and banks
3	29yuan	Unlimited route with 10% extra charge more	Ticket offices and agencies	0.8yuan	Ticket offices, agencies and banks
4	29yuan	Unlimited route with 15% extra charge more	Ticket offices and agencies	1yuan	Ticket offices, agencies, banks, telephone and online
5	37yuan	Unlimited route: any public transport services without extra charges	Ticket offices and agencies	0.4yuan	Ticket offices, agencies and banks
6	37yuan	Limited route: Only one bus or light rail route covered in urban area	Ticket offices and agencies	0.6yuan	Ticket offices, agencies, banks, telephone and online
7	37yuan	Unlimited route with 15% extra charge more	Ticket offices and agencies and banks	0.8yuan	Ticket offices and agencies
8	37yuan	Unlimited route with 10% extra charge more	Ticket offices	1yuan	Ticket offices, agencies and banks
9	45yuan	Unlimited route with 10% extra charge more	Ticket offices and agencies and banks	0.4yuan	Ticket offices, agencies, banks, telephone and online
10	45yuan	Unlimited route with 15% extra charge more	Ticket offices	0.6yuan	Ticket offices, agencies and banks
11	45yuan	Limited route: Only one bus or light rail route covered in urban area	Ticket offices and agencies	0.8yuan	Ticket offices, agencies and banks
12	45yuan	Unlimited route: any public transport services without extra charges	Ticket offices and agencies	1yuan	Ticket offices and agencies
13	53yuan	Unlimited route with 15% extra charge more	Ticket offices and agencies	0.4yuan	Ticket offices, agencies and banks
14	53yuan	Unlimited route with 10% extra charge more	Ticket offices and agencies	0.6yuan	Ticket offices and agencies
15	53yuan	Unlimited route: any public transport services without extra charges	Ticket offices	0.8yuan	Ticket offices, agencies, banks, telephone and online
16	53yuan	Limited route: Only one bus or light rail route covered in urban area	Ticket offices and agencies and banks	1yuan	Ticket offices, agencies and banks

Table 5.8-7: Skeleton Design of SP 4: Travel Cards vs. Smart Cards (Pay monthly cards)

Ex. 4-1 (Note: The 'multifunction' column is changed to 'Geographic area covered' for Ex. 4-2, following the same coding scheme)

	Travel Cards		Smart Cards		
	Cost difference (TC-SC)	**Services**	**Services**	**Deposit**	**Multifunction**
1	-19yuan	Limited route: Only one bus or light rail route covered in urban area	Unlimited routes in urban area without any extra charges	0yuan	No. only for public transport
2	-19yuan	Unlimited route: any public transport services without extra charges	Unlimited routes with extra charge: 10% more than limited services	30yuan	Shopping, telephone, entertainment ,parking, tolling and banking
3	-19yuan	Unlimited route with 10% extra charge more	Unlimited routes with extra charge: 15% more than limited services	50yuan	Shopping, telephone, entertainment
4	-19yuan	Unlimited route with 15% extra charge more	Unlimited routes with extra charge: 20% more than limited services	20yuan	Shopping, telephone, entertainment ,parking and tolling
5	-11yuan	Limited route: Only one bus or light rail route covered in urban area	Unlimited routes with extra charge: 10% more than limited services	20yuan	Shopping, telephone, entertainment
6	-11yuan	Unlimited route: any public transport services without extra charges	Unlimited routes in urban area without any extra charges	50yuan	Shopping, telephone, entertainment ,parking and tolling
7	-11yuan	Unlimited route with 10% extra charge more	Unlimited routes with extra charge: 20% more than limited services	30yuan	No. only for public transport
8	-11yuan	Unlimited route with 15% extra charge more	Unlimited routes with extra charge: 15% more than limited services	0yuan	Shopping, telephone, entertainment ,parking, tolling and banking
9	-3yuan	Limited route: Only one bus or light rail route covered in urban area	Unlimited routes with extra charge: 15% more than limited services	30yuan	Shopping, telephone, entertainment ,parking and tolling
10	-3yuan	Unlimited route: any public transport services without extra charges	Unlimited routes with extra charge: 20% more than limited services	0yuan	Shopping, telephone, entertainment
11	-3yuan	Unlimited route with 10% extra charge more	Unlimited routes in urban area without any extra charges	20yuan	Shopping, telephone, entertainment ,parking, tolling and banking
12	-3yuan	Unlimited route with 15% extra charge more	Unlimited routes with extra charge: 10% more than limited services	50yuan	No. only for public transport
13	+5yuan	Limited route: Only one bus or light rail route covered in urban area	Unlimited routes with extra charge: 20% more than limited services	50yuan	Shopping, telephone, entertainment ,parking, tolling and banking
14	+5yuan	Unlimited route: any public transport services without extra charges	Unlimited routes with extra charge: 15% more than limited services	20yuan	No. only for public transport
15	+5yuan	Unlimited route with 10% extra charge more	Unlimited routes with extra charge: 10% more than limited services	0yuan	Shopping, telephone, entertainment ,parking and tolling
16	+5yuan	Unlimited route with 15% extra charge more	Unlimited routes in urban area without any extra charges	30yuan	Shopping, telephone, entertainment

primary attributes, which are included in all sub-exercises. Overdraft, multifunction, geographic areas covered and top-up methods are used once each time based on the four primary attributes to generate four different sub-exercises. The same rule is applied in Exercise 3 and 4 to split the SP games into 3 and 2 sub-exercises respectively. Because Exercise 1 can generate a proper number of binary choice profiles, a single exercise is used for this exercise.

5.4.5. Presentation and Administration of the SP Survey

The SP games are presented as binary-choice experiments in the SP survey. Each respondent received only one SP questionnaire paper, which contains fare payment method(s) s/he used before and another one he/she needed to trade off, with these 8 pair-wise choice situations. The respondent is asked to choose one alternative he/she preferred in each situation. A detailed guide for answering the questionnaire is also presented in the first page of the questionnaire to assist the respondents to understand the choice context. Two examples selected from two SP questionnaire versions are illustrated as follows (See Table 5.9). The full SP questionnaires are attached in the Appendix B for reference.

Table 5.9 SP Choice Situation-Cash vs. Travel Cards

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	0.8yuan per ride	Average 20seconds slower than travel cards	Any	Yes	
Travel Cards	29yuan per month	Straight getting on	Limited route: Only one bus or light rail route service. But you still can pay by cash to take other services	N/A	

SP Choice Situation-Travel Cards vs. Smart Cards (Pay as you go)

Fare Payment Methods	Travel Cost	Public transport services Covered	Deposit	Multifunction	Choice
Travel Cards	29yuan per month	Unlimited routes with extra charge: 10% more than this value	N/A	None	
Smart Cards	0.8yuan per ride	Any public transport modes and routes	50yuan	Shopping, telephone, amusement	

The SP survey was administered by self-completion questionnaires that were collected back by surveyors after respondents completed. As with all self-completion questionnaires, there would be the problem of unknown bias introduced by questionnaire not being returned, particularly when using mail-back. So collecting completed questionnaires on the spot was used instead of mail-back method. Face-to-face interviews could guarantee high response rate and good data quality, but higher survey cost than other methods should be considered.

In addition, considering respondents' privacy, people prefer their backgrounds being recorded anonymously rather than face-to-face interviews.

At the beginning of an SP questionnaire, some screen questions were asked to determine the user types. Because self-completion survey was used, before respondents started the SP experiments, we must ensure that proper SP questionnaires had been given according to their existing payment behaviour. Moreover, asking respondents to report their user type is also important to identify the current market shares in the random sample and compare the prediction of market share in the later analysis.

5.4.6. Some RP Questions in the SP Survey

Prior to the SP binary choice experiments, some RP questions about the respondents' current payment behaviour were presented to identify respondents' actual payment behaviour, including payment methods used (ticket types, travel card types or smart card types), estimated number of trips by using different payment means. In these RP questions, it is worth noting that passengers should provide the total number of one-way trips as the estimated trips so that different units on travel cost (i.e. Yuan per trip and Yuan per month) can be converted to a consistent cost unit (Yuan per month) for the modelling analysis through referring the number of trips happened before.

5.4.7. Demographic Questions

In this research, the RP and SP are separated into two independent surveys, because (1) the RP and SP survey are used to collect different kinds of data, if two surveys were combined into one questionnaire paper, such long questionnaire would make respondents fatigue and the data quality would be impacted; (2) self-completion on board survey is mainly employed, so it could be quite difficult to target the same sample for the both RP and SP survey, compared with household survey. Therefore, two separate RP and SP surveys have to be used, taking no account of whether the samples are the same or not.

So, following the SP choice situations, the same demographic questions as the RP survey, including age, sex, educational level, employment status, household income per month, availability of private transport and the willingness-to-prepay for public transport fare, are presented in the last section of the SP questionnaire papers such that a segmentation analysis with the identical socio-economic variables could be conducted crossing the RP and SP data.

5.5. Pilot Surveys

5.5.1. Sampling for the Pilot Surveys

Before the main survey, a pilot survey within a small proportion of sample was conducted to test the draft of the RP and SP questionnaire papers under actual conditions. The aim of pilot survey is to ensure that:

(1) Its robustness prior to use as part of the actual survey process, in particular to test whether those new attribute levels can be actually understood be respondents under the self-completion SP survey.

(2) The feasibility and applicability of the proposed methods of surveying, such as on board survey.

(3) Understanding of how people vary their choices with respect to changes in the various attribute values, especially for the SP design.

RP Pilot Survey:

50 respondents were surveyed in the RP pilot survey during 10th-20th June 2005. 40 completed questionnaire papers were collected back with the response rate of 80%. It should be noted that this response rate could not be used as an indicator for the actual response rate in the main survey as the RP pilot was conducted with known contact mainly, such as contacting with friends and relatives by email or telephone before questionnaire papers were sent out. Meanwhile, a quick reminder also was given by these known contacts before completed questionnaire papers were collected. So under this circumstance, a higher response rate than the main survey could be obtained.

In the RP pilot survey, most respondents were from a group of people with a higher educational background (e.g. university students, research staff in some institutes, local government officers, senior staff in public transport companies, *etc*). The reason for surveying among these people is to ensure a higher response rate for the pilot survey and receive some suggestive and helpful feedbacks about the survey design. In addition, a small proportion of RP questionnaires were sent to people randomly selected on board, mainly to test the survey duration to complete one piece of questionnaire paper. The survey methods employed in the pilot survey was to randomly send printed RP questionnaire papers in those known contacts. Meanwhile, on board survey also was used. Finally, 10 papers were sent to a university, 10 to a research institute, 10 to a public transport company, and 20 to two different local government departments and passengers on board in two bus routes.

SP Pilot Survey:

It took about one week to invite and ask 48 respondents for the SP pilot survey in July 2005 before the main SP survey was carried out under way. Finally, 40 valid SP questionnaire papers were collected back with the response rate of 83.3%.

The similar sample type with the RP pilot survey was randomly selected among those people with a good educational background. However, 10 of 48 SP questionnaire papers were also sent to some passengers randomly on board. The survey method was mainly a self-administration survey, assisted by some interpretation about the SP survey from surveyors to avoid misunderstanding and confusion when the respondents were trying to trade off choice situations under this unfamiliar survey method (stated choice).

In the both RP and SP survey, each respondent's questionnaire was clearly labelled so that they would have a unique identification number for future analysis, which also identified which version of the questionnaire they had received. Such method to identify different respondents was also used in the main survey. During the SP pilot survey, considering the unfamiliarity of the SP survey to Chinese people, the survey team (myself and other two friends who had been trained) was also on the spot at all times to explain anything about the survey. Meanwhile, in order to guarantee the data quality (individual response was required) and minimise environmental noise which could impact the individual data quality, the respondents in the SP pilot survey were also repeatedly informed that their own opinions were valid and that they should not be influenced by other respondents when they were trading off different SP situations.

5.5.2. Findings and Lessons from the Pilot Survey

In general, the pilot survey questions were well understood and the whole RP/SP survey was relatively easy to conduct. On the other hand, the pilot study highlighted a number of important issues for conducting the main survey, including:

(1) The approach to respondents was important in getting them to listen and accept the survey. A proper way to interpret questionnaire papers would play an important role to motivate the respondents to agree to participate.

(2) As Dalian is a very popular holiday resort in China. Especially in summer, there would be a great number of tourists when the main survey was carried out during July-August 2005. So it could be difficult to collect full information when surveying a tourist visiting Dalian, who possibly does not know much about the local public transport fare payment system. This factor requires the surveyors need to filter tourists when they use questionnaire papers on board.

(3) Although the RP and SP survey were well understood by respondents in general, the survey duration for each single respondent may be an issue, because the length of the survey was longer than 15 minutes we scheduled. Even some SP surveys lasted for about 25 minutes due to respondents' unfamiliarity to the SP survey. Hence, an instruction by surveyors should be given necessarily to lead respondents to answer questionnaires, avoiding any confusion. For example, in the RP pilot survey, some respondents answered screen questions but wondered where they should go further after that, resulting in unnecessary time waste. So some assistant techniques could be used, such as using bold letters, signposts, *etc*. Meanwhile, in the main survey, constant

assistance from surveyors during respondents' answering questionnaire is also needed to solve any potential misunderstanding to the questionnaire, so as to reduce the survey duration.

(4) Some problems happened during on board survey which could result in a low response rate in the main survey potentially. For instance, crowded condition on board is viewed as the major obstacle to make the survey go through during peak-time. Finally it was found that surveying on-board from some bus/tram terminals was better than surveying during the journey, because on-board condition in terminals were better and questionnaires were sent to those who had got a seat.

(5) Four SP games and twenty different SP questionnaire versions were used, so it is necessary to use different colour papers to distinguish different questionnaire versions in the main survey. Meanwhile, to balance the number of different questionnaire versions sent and collected back is another issue to take into account for the main survey later on.

5.5.3. Modifications of the Survey Design

RP Survey:

Through the pilot survey, we found that the response to the RP survey was quite positive except some minor modifications on the questionnaires. Most feed back focused on the wording on the questionnaire paper. And most respondents need more detailed instruction to guide them, such as which parts they should answer and then where they should go, *etc*, so in the final RP questionnaires, these guide words were quite clearly visible with bold-faced typing.

(1) Some definitions were made clearly in the final version: the concept of trip here is one-way trip; the household income, which means the total income of all family members who work currently, were stressed particularly, because in the pilot survey some respondents viewed it as the income of the individual who was surveyed.

(2) Some screen questions were added before respondents answer questions about their secondary payment method(s): the aim of the screen questions is to distinguish whether the respondents actually used a payment method as their secondary mean. For example, after a travel card user answered questions about travel card payment, travel costs about cash and smart cards also were asked. But before he/she answered these additional questions, a screen question: "Did you use cash (or smart cards) in the last month as your supplementary payment method?", if he/she said "Yes", then continue to answer questions about his/her supplementary payment travel cost, otherwise move to other sections. Here, we must stress that the travel cost is only for the supplementary payment method, not the whole travel cost (including main and supplementary means totally).

SP Survey:

When viewing the pilot survey data, it should be noted that as one of problems we worried before: a respondent might choose the same alternative in all situations, did not occur in this pilot survey. That means in general the survey design was satisfactory and feasible, because what we expect about the respondents' responses should be presented by a variety of choices towards different situations, so that we can see that the variations of attributes make sense and the data can be modelled in the term of alternative utility. However, some minor modifications still were made for finalising the SP survey as follows:

(1) More detailed instructions about the SP choice experiments, for example: before starting the binary choice, some details about the alternatives being traded off was described for the respondents, such as what these two alternatives look like, some general features could be traded off. It is quite necessary to help those people, who do not know much about these payment methods, to understand the situations they could be in.

(2) Some confusion about boarding time difference between cash and other two cashless fare payment methods (i.e., travel cards and smart cards) might cause misunderstanding to trade off different situations. Therefore, for boarding time variable, it must be stressed in the SP survey that the time difference presented in SP profiles is only an estimated average value.

5.6. Summary

In this chapter, we discuss the RP and SP survey design, including the characterisation of decision situation, determination of attributes and levels, questionnaires types, SP games being used, and generation of choice profiles for the SP games. Finally, three RP questionnaire versions for three different primary users (cash, travel cards and smart cards) are designed. For the SP survey, four SP games are used, cash vs. travel cards; cash vs. pay as you go smart cards; travel cards vs. pay as you go smart cards; and travel cards vs. pay monthly travel cards.

In this chapter, through the SP survey design, we introduced new features and attributes of three fare payment options according to the literature review. The RP survey focuses on PT users' actual choice behaviour, while the SP survey investigates respondents' preference (intention) in different trade off situations. Four SP games cover different trade off situations between cash and travel cards, cash and smart cards, travel cards and smart cards. In the SP survey design, new attributes and levels related to different payment options were introduced, particularly for smart cards. These new features of fare payment applications can be evaluated according to respondents' trade-offs. Furthermore, demand changes of different payment options based on new situations can be obtained. Eventually, outcomes of the evaluation study are proposed to feed back to the policy making of authorities, to enhance the service quality of public transport.

Meanwhile, before the main survey, a pilot survey is discussed in this chapter to test the survey design. Through the pilot survey, some modifications have been made to finalise the RP and SP questionnaires. Furthermore, some issues and experiences, which were obtained in the pilot survey, are taken into account in the main survey so as to avoid any respondents' confusion and misunderstanding about the questionnaires, to guarantee a good data quality and high response rate.

In a word, this chapter presents two different surveys being used in the data collection. Meanwhile, the relevant variables (attributes) are determined, which are used in the later data analysis as indicators of the benefit evaluation. In the next chapter, the details about the main survey will be discussed.

Chapter 6
Data Collection

6.1. Introduction

Following the survey design and pilot survey in Chapter 5, the main data collection for this research was conducted in Dalian, China during July and August 2005. The aim of this survey was to collect the preference data (revealed preference and stated preference) from public transport users toward the existing and prospective public transport fare payment/collection methods (i.e., cash, travel cards and smart cards).

The data collection in this chapter begins with a framework of the data collection in Section 6.2, in which some basic concepts and preparation for the survey are introduced, such as survey permission, survey location, sampling, etc. Following the framework, the survey methods used in the data collection are discussed and described in detail in Section 6.3. In Sections 6.4 and 6.5, the main RP and SP survey are presented respectively, including the response rate, people's responses to different alternatives in the RP and SP, understanding of the SP survey, etc. In this chapter, preparation work for the data analysis of the next chapter is to analyse respondents' basic characteristics in Section 6.6, including distributions of socio-economic factors, reasons for such results in data characteristics, etc. Another purpose for analysing respondents' characteristics is to check the survey methods: whether the data collected were representative. Finally, Section 6.7 summarises issues and experiences in the data collection, which could not be identified in the pilot survey in the last chapter. Especially for the SP survey, although the SP approach has proven to be successful in the context of developed countries (Louviere *et al.*, 2000), it is still unfamiliar to the Chinese and very few studies have utilised it in China. Therefore, this data collection could be regarded as a trial to shed light on SP survey methods for future research in China.

6.2. Framework of Data Collection

6.2.1 Survey Permission

Survey permission is the prerequisite to make the data collection possible, because in the pilot survey, the on board survey method had been tested and it had shown a good performance. Therefore the same survey method was also used in the main survey. The survey locations and places determined where the permission should be obtained from. Prior to the main survey, the 1st Bus Company of Dalian, the 2nd Bus Company of Dalian, Dalian Modern Light Railway Co., Ltd. and Dalian Mingzhu Smart Cards Co., Ltd. had issued the authorisation to support the data collection on their service routes. Meanwhile, in order to send questionnaires in some clusters (we also tried this survey method in the pilot survey, e.g. companies, residential communities, schools, *etc.*), the relevant permission also had been obtained from schools, companies, residential communities, etc.

6.2.2 Survey Location

Due to on board surveys being employed and the anticipated difficulty in recruiting participants, the survey location or bus/tram routes needed to be planned and selected carefully. Therefore, in order to target the right participants, reduce the proportion of refusals to participate, and obtain a large number of multi-modal passengers in Dalian urban area, it was decided that the survey would be based on bus terminals and interchange points for different public transport modes in the city centre.

The survey sites and public transport routes chosen were required to provide:

- A good mix of users such as local buses, trams and light railway.
- A variation of travellers such as commuters, business travellers and travellers with different types of employment status and travel purposes.
- A large overall number of passengers, including peak and off-peak hours.

6.2.3 Sampling

Having decided on the design of the survey instrument and tested it in the previous chapter, the size of the sample of individuals must also be determined. The issues concerning sampling for SP surveys particularly are largely the same as for other market research surveys. A representative group of people in the area in which we are concerned should be obtained. On the other hand, we also need to identify suitable sub-groups (or 'segments') of the population of interest and obtain sufficient numbers in each group.

One important point to consider is the fact that although an SP survey produces multiple responses per individual (in this research, with 8 choice situations in each questionnaire, then one respondent can generate 8 responses to different binary choice situations), it does not mean that very small samples can be used to get ideal results. Normally, 30-40 individuals for an SP survey could be enough for data analysis (Pearmain, et al, 1991). In this research, we designed 4 SP games, in which there were 10 sub-exercises with 16 binary choice situations in each sub-exercise. Also, in order to reduce response fatigue, these 10 sub-exercises were split into 20 different SP questionnaires by using 'block design'. Therefore, considering the number of SP questionnaire versions, no less than 100 questionnaires should be sent out within these 10 SP sub-exercises respectively.

As to the RP survey, because we had separated three primary payment user types to send different RP questionnaire versions, the sample size for each user type was set the same quantity to guarantee the significance of data we collected. On the other hand, because evidence of the market share of three payment methods in use in Dalian, China, was not available before the main survey was carried out, the way to sample respondents by the actual market share of alternatives could not be used in this research.

6.2.4 Survey Presentation and Administration

In considering what kinds of ways should be used to present the questionnaires, the factors likely to influence surveyors' decision may be the following:

- The complexity of the survey (particularly for SP experiments) and the length of the questionnaires;
- The detail with which alternatives need to be described;
- The circumstances in which interviews would take place or questionnaires would be completed;
- The response relating to the administration of the survey;
- The survey cost (workload and time) should be controlled at a low level;
- It should be feasible and accepted by respondents.

In China, the mail-back method is very difficult to carry out because of lower cooperation from people and higher survey costs. In addition, compared with the mail-back method, the presence of surveyors when respondents answer questionnaires can help respondents understand questions and the data quality can be guaranteed. Therefore very few researchers use mail-back survey techniques in China. Although face-to-face interviews can increase the response rate, it is very easy to cause fatigue and confusion to respondents, if the survey questions and choices are too long and complex. The second concern about face-to-face interview method is the protection of the privacy of respondents. If respondents would not like to report their background data by this way (household income particularly), it would be very hard to obtain satisfactory response rates and data quality. In addition, the survey cost of interviews is quite considerable, including arranging survey time and interview place, training interviewers, etc. After discussing disadvantages of those survey methods unsuitable for this survey, self-administered questionnaires handed out and collected back by surveyors on the spot was finally selected, because it was much more appropriate for this research context. First of all, the survey cost can be controlled to a low level compared with mail-back survey and interview. Secondly, it requires the presence of surveyors on the spot to guide respondents to answer questionnaires if necessary, to remind respondents to submit their questionnaire papers. So the data quality can also be guaranteed. Thirdly, self-administered surveys can also be readily conducted, without some extra requirement, such as arranging time and place for face-to-face interview.

6.2.5 Survey Team Recruitment and Training

After determining the survey administration scheme, the following task is to recruit and train survey assistants. In Section 6.2.4, the self-completion survey is selected, which requires respondents to fill in the questionnaires by themselves. Although the self-administered survey is employed, a well-trained survey team should be on the spot to sample respondents, to distribute different questionnaire papers, and to offer information about the survey itself as much as they can to respondents if needed.

Survey team recruitment: the requirement to recruit survey assistants is that they must have got familiar with or recently used the current fare payment methods in Dalian public transport system. But it is also required that their individual perceptions and experience should not influence respondents' decision making during the survey. Finally two university students who studied in a local university were recruited, and including myself, the survey team consisted of three people.

Survey team training: The following task is to make all survey assistants understand the questionnaire. The best way to do this is to ask them to fill in some of questionnaire papers, including the RP and SP survey both. Only doing this can the surveyors deal with all questions from respondents when they answer the questionnaires, avoiding any misleading from the surveyors.

***Field test*:** finally, in order to check whether the survey team can work well and target potential problems during the main survey, a field test was conducted by using 15 questionnaire papers on three bus routes. As the field test was regarded as a practice for the survey team, the data collected in the field test was not included in the data set of the main survey.

6.3. Survey Methods

6.3.1 RP Survey Methods

In the RP survey, a peak and off peak time on-board survey was employed, because on-board surveys can easily target particular respondents (i.e., public transport users in this research) thus full information about PT fare payments could be obtained. Compared with the on board survey, it is not convenient for road side survey and household survey to determine public transport users, though people can be asked whether they are public transport users or not.

However, because in Dalian during peak time it is very common that more than 50-60 passengers are in one single-decker bus, it is quite difficult to send and collect questionnaire papers under such crowded conditions on board. This problem also happened on the pilot survey when on board surveys were carried out during the peak time. In order to survey during the peak time and obtain high response rates, the solution was to send questionnaire papers to those passengers who had already got seats on board at the starting point of one bus route, then the surveyor sat or stood by the egress door of the bus during the journey and collected completed papers when the respondents got off. The second concern about surveying passengers having seats on board is that it was very inconvenient and dangerous for passengers who were standing on board to finish questionnaire papers when the bus was moving. Although the questionnaire papers were only given to passengers who got seats, we can still view it as random sampling, because whether a passenger can get a seat on board is a random procedure, depending on the number of seats on board and number of passengers in the queue. In other words, those passengers who got seats had been selected randomly before we sent questionnaire papers.

As well as on board surveys during peak time, another solution was to survey during off peak time. First of all, surveying people who travelled during different time period can make the data more representative. Secondly, the conditions on board during the off peak time was much better than peak time, so the survey team could work under better conditions and higher response rate could be obtained.

6.3.2 SP Survey Methods

As stated in Section 5.7.4, Chapter 5, the whole survey was divided into two separate surveys: the RP and SP surveys, based on the concern of the long survey duration if the RP and SP were combined together. Although the RP and SP surveys were carried out separately by two stages, the same data collection method: on board survey was used in the SP survey. But in order to have the SP questions explained to the respondents by surveyors on board, in case of any misunderstanding or questions during the survey, those bus services which contained less passengers than busy lines during peak and off peak time were selected.

Considering the complexity of the SP survey to PT passengers relative to the RP survey, which required the respondents to report their payment behaviour directly rather than trade off each binary choice situation, another survey method, in-cluster survey, was employed to send and collect SP questionnaire papers, from companies, schools, communities, *etc.* To achieve this, first of all, the surveyors must ensure that those clusters have a great number of public transport users and some detailed instruction about how to answer questionnaires must be given in advance. Then on the following day those completed questionnaire papers were collected back by the second visit of the surveyors. The second method, in-cluster survey, can guarantee higher response rates, but the bias and error might be a potential problem, because we surveyed only in those selected clusters, ignoring the existence of others. Additionally, it was highly possible that the in-cluster survey could cause the simplicity of socio-economic backgrounds of respondents and a lack of representativeness of the SP data. So in order to keep respondents representative, reduce the effects of bias and error by using in-cluster survey, a great number of clusters (more than 20 companies, schools, residential communities and so on) were randomly selected and a very limited number of SP questionnaires were sent out in each cluster (only about 20-30 SP questionnaire papers).

6.4. Revealed Preference (RP) Survey

6.4.1 the RP Survey

The main RP survey lasted for 21 days from 8th July to 28th July 2005. In total, 1500 RP questionnaire papers were sent out (500 for cash, 500 for travel cards and 500 for smart cards, respectively) and owing to passengers' high interests in the survey, 1016 RP questionnaires were collected back. Among them, 869 valid RP questionnaires were picked up and could be used in the modelling analysis in Chapter 7 (251 for cash, 315 for travel cards and 303 for smart cards). The definition of 'valid questionnaire' in the RP survey is those questionnaires without any missing items about payment behaviour and with at least two alternatives choice situations made by the respondents. In addition, whether the respondents' background data was logical and consistent with their travel behaviour is another criterion to check the validity of the RP data.

The response rate was about 57.9% overall (See Table 6.1). Although it was much lower than the pilot survey, the response rate still reached a satisfactory level compared with 20-30% response rate in most public transport surveys (Ampt, 1990; Ortuzar, 2000). As can be seen in Table 6.1., 147 of 1016 returned questionnaires were excluded. Among those rejected data observations, non-response data and illogical data observations are the two main sorts of data rejected in RP-2 (for travel card users) and RP-3 (for smart card users). In Section 6.6, the details of illogical data are discussed. In RP-1 (for cash users), only those data with more than two (including two) alternatives were kept in the data set, because in the data analysis, the utility model required such choice situation to measure different utilities of alternatives to individuals. For those respondents who only had one option (cash payment in RP-1), the preferences to other choices cannot be obtained in the modelling stage. Therefore, in 311 returned RP-1 questionnaires, 52 such data are not considered in the modelling stage.

Table 6.1 The Response of the RP Survey

	Valid Q's	Returned Q's	Handed out Q's	Response rate
RP1: Cash Users	251	311	500	50.2%/ 62.2%
RP2: Travel Card Users	315	343	500	63%/ 68.65%
RP3: Smart Card Users	303	362	500	60.6%/ 72.4%
Total	869	1016	1500	57.93%/65.5%

The following aspects can be summarised for the main RP survey:

(1) Non-response: although such a problem was not a major issue in the RP survey, there were a small proportion of questionnaires missing some important information.

Particularly in RP-2 and RP-3, estimated cash payment costs per month (if they paid all PT fare by cash) were required to be answered by travel card and smart card users. However, some card users would think the cash cost was not relevant to themselves and difficult to estimate, so non-responses were given. The second cause of non-responses was the lack of time for some respondents to complete the RP questionnaires fully. They could only give information that can be answered without any effort, such as ticket types used, 'yes/no' questions, overall assessment, etc.

(2) Single choice without other alternatives: Although this sort of response did not happen on the pilot survey, it still can be easily understood if one respondent only uses single fare payment in reality. For RP-2 and RP-3, there is no data rejected due to the single choice problem, because for current travel card and smart card users, cash was thought available all the time for all PT services. So the respondents in RP-2 and RP-3 can offer us at least two alternatives in their responses. However, for cash users in RP-1, it is highly possible that travel cards and smart cards could not be used because of unavailability of these two alternatives. Under this situation, only cash can be used. In Table 6.2, responses to different alternatives are listed.

Table 6.2 Respondents' Actual and Possible Choices on Three Fare Payment Methods

	Only one payment method used, without alternatives	**Two alternatives**	**Three alternatives**	**Total**
RP-1	52 cash users (excluded in the modelling analysis)	156 (87 cash and travel cards; 69 cash and smart cards)	95	251/303
RP-2	0	123 (travel cards and cash)	192	315
RP-3	0	121 (smart cards and cash)	182	303

(3) Illogical responses: this problem mainly happened on respondents' payment behaviour relative to their socio-economic backgrounds. For example, a respondent with a full time job cannot use student travel cards. Therefore, such data should be excluded, because they cannot represent reasonable and actual situations among respondents. As to the details about the illogical data, Section 6.7 of this chapter discusses them.

(4) Understanding of the RP survey: In general, respondents' understanding of the RP questions was satisfactory, but some respondents still were not very clear about two sorts of questions on travel costs: 1). travel costs for those available payment alternatives, which were not actually used in the last month. Some passengers regarded those payment methods as their secondary options, rather than the main payment method if they would use it. Such passengers reported the travel costs in case their main payment methods were unavailable, not the full cost information if they would use them

as their main payment choices; 2). On the contrary, when we asked questions about people's supplementary payment methods in case their primary payment method cannot be used, some respondents gave their costs by using supplementary methods to cover all trips.

6.4.2 Basic Characteristics of the RP Data

Before modelling preference data in Chapter 7, to help understand the actual situation about fare payment preferences in the real market, statistical descriptions of attributes in the RP survey are summarised, including market shares of different payment options in the RP survey, travel cost, boarding time, and some variables relevant to users' assessment of the current payment applications.

(1) Primary Payment Choices

In Table 6.1, although we exclude 52 single cash payment users from the data set being used in the modelling analysis, they are still taken into account when basic statistical analysis is carried out in this section.

Cash ticket type:

Cash fare types in use were obtained in RP-1. Also, for RP-2 and 3, we assumed that cash was available all the time for all respondents, so current travel card users and smart card users can provide the relevant information about cash fare payment (Regarding the details of the RP questionnaires, please refer the Appendix B). As can be seen in Table 6.3, most respondents (including actual and potential cash fare users) would like to use flat fare, because flat fare can be found in all bus routes in the urban area (1yuan per single trip). Only mini bus, tram and light railway services implement zonal fare policy due to the longer service distance they offer.

Table 6.3 Ticket type used or could be used by the respondents

	Flat fare	Zonal fare	Total
RP1-Actual cash users' choices	202	101	303
RP2-Potential cash users' choices	232	83	315
RP3-Potential cash users' choices	211	92	303

Travel card type:

Table 6.4 lists the choices of travel card types, including actual and potential users. In RP-2, the number of users choosing monthly travel cards with limited bus route was overwhelmingly greater than the other three travel card types. The phenomenon also happened on RP-1 and 3, which was used to investigate the payment behaviour of potential users. Several reasons may cause this: (1) very regular travel purposes could drive respondents to choose travel cards with limited bus route services by reduced cost; (2)

considering the travel cost, quarterly travel cards require passengers to pay more than monthly cards. Monthly cards with unlimited services also cost more than cards with limited bus services.

Table 6.4 Travel card type used or could be used by the respondents

	A	B	C	D	Total
RP2-Actual travel card users' choices	252	29	21	13	315
RP1-Potential travel card users' choices	122	16	30	14	182
RP3-Potential travel card users' choices	131	20	13	18	182

Note: A: Monthly cards with limited bus route
B: Quarterly cards with limited bus route
C: Monthly cards with unlimited bus route
D: Quarterly cards with unlimited bus route

Smart card type:

Through Table 6.5, we can see that most existing smart card users chose 'pay as you go' cards in RP-3, because it was more flexible to use 'pay as you go' cards than other smart card types, due to it having no expiry date and being available on all PT services. Also, more than 75% of respondents in RP-1 and RP-2 could use 'pay as you go' smart cards. As a new option to substitute paper-based travel cards, pay monthly smart cards ('electronic travel cards') were at the trial stage and the number of cards was limited in use. Secondly, most respondents did not know much about this card type. All these could result in the proportion of pay monthly smart cards being very low. Moreover, most students were currently travel card users, so it was not a surprise that very few respondents used pay monthly smart cards.

Table 6.5 Smart card type used or could be used by the respondents

	A	B	C	D	E	Total
RP3-Actual smart card users' choices	251	33	9	10	0	303
RP1-Potential smart card users' choices	123	15	18	0	8	164
RP2-Potential smart card users' choices	158	19	5	1	9	192

Note: A: "Pay as you go" card
B: Electronic travel card (a minimum payment required per month)
C: Student smart card
D: Elder smart card
E: I do not know

(2) Travel Cost:

Actual users:

The travel cost variable is divided into two parts: one for actual users, another for possible users. Travel costs of actual users (cash, travel cards, and smart cards) are presented in Table 6.6.

Table 6.6 Actual User Travel Cost

	Cash	Travel Cards	Smart Cards
Actual Data Entry	303	315	303
Average Cost (yuan)	48.31	52.26	53.05
Max/Min	240/5	172/21	150/10
Percentile 10	10	30	20
Percentile 25	20	36	40
Percentile 50	40	48	50
Percentile 75	60	60	60
Percentile 90	100	80	100
S.D.	40.62	21.87	24.71

As can be seen from Table 6.6, the average values of costs of travel cards and smart cards are quite close, around 52-53yuan per month, but higher than cash. This could indicate that most occasional PT users with a lower travel demand preferred using cash. The range of travel costs of three payment methods are indicated by Max/Min values, but we can observe that max and min values of travel cost are somewhat extreme. In order to get a proper distribution of travel cost, in Table 6.6 percentile values are also used by five random ranks (i.e., 10th, 25th, 50th, 75th, and 90th). The percentile value of cash users reveals that most data (90%) are less than 100, although max/min values are 240/5. For travel cards and smart cards, 90% of cost data are less than 80yuan and 100yuan respectively. In addition, it can be seen that standard deviation values of travel cards and smart cards are lower than cash, so we can conclude that compared with cash payment, the cost with card payment is relatively stable.

Possible Users:

When investigating possible travel costs of payment methods that respondents did not actually use in the last month, we can see the reason why the respondents preferred the payment method they actually used rather than other payment methods, although they were available. First of all, for travel card and smart card users, if their trips were paid by cash, their costs were higher on average than paying by travel cards and smart cards (70.36yuan for actual travel card users; 68.65yuan for smart cards users in Table 6.7, while in Table 6.6, their actual costs ranged from 52-53yuan per month on average). Secondly, the average cost of travel cards (59.22yuan) and smart cards (52.43yuan) for existing cash users are higher than their actual costs on cash in one month (48.31yuan in Table 6.6), if cash users would use travel cards or smart cards for their all trips. That may be the reason why some passengers chose cash payment. Thirdly, the average possible travel cost of smart cards (that could be used by travel card users) and travel cost of travel cards (that could be used by smart card users) are very close (63-64yuan), which could indicate that on the travel cost attribute, there was no difference between smart cards and travel cards in card users' minds.

As shown in Table 6.7, percentile values are used to have an insight into the distribution of cost data as well, because max/min values in Table 6.7 are also shown to have a large range. In general, the distribution of cost value in Table 6.7 indicates that most data are less than 90-100yuan (at percentile 90th level).

Table 6.7 Possible User Travel cost

Fare Payment could be used	Cash		Travel Cards		Smart Cards	
RP questionnaire type	For TC users	For SC users	For cash users	For SC users	For cash users	For TC users
Actual Data entry	315	303	182	182	164	194
Average cost (yuan)	70.36	68.65	59.22	63.97	52.43	64.36
Max/Min	240/30	200/15	200/10	200/29	210/5	200/30
Percentile 10	50	45	30	40	20	49
Percentile 25	60	50	41	49	30	50
Percentile 50	60	60	54	60	50	60
Percentile 75	80	80	70	70	65	70
Percentile 90	100	100	90	90	90	87.5
S.D.	22.78	25.00	26.72	25.98	30.43	20.77

(3) Boarding time difference:

Actual users

Table 6.8 Boarding Time Difference of Actual Users (cash: baseline)

	Cash	Travel Cards	Smart Cards
Actual Data Entry		315	303
Average (second)		8.06 seconds quicker than cash	7.2 seconds quicker than cash
Max/Min		130/0	80/-3
Percentile 10		2	0
Percentile 25		2	1
Percentile 50		5	3
Percentile 75		10	5
Percentile 90		15	20
S.D.		12.75	12.97

As can be seen in Table 6.8, boarding time differences from cash, by travel cards and smart cards users, are very close, about 7-8 seconds quicker than cash, which means that these two cashless payment methods have the same advantage over cash on reducing boarding time. But the ranges of boarding time difference of two cashless payments are very different. Travel card users reported a bigger range than smart cards. On the other hand, some smart card users said boarding time by using smart cards was longer than by using cash (e.g. -3 seconds). The possible reason for this result would be the unreliability and malfunction of smart card reading devices on board, affecting card users' ability to pay. Therefore, their boarding time increased sometimes.

When observing percentile values, it can be found that at percentile 90th, boarding time differences from cash are 15-20 seconds. Therefore, we can conclude that although the range of boarding time differences is great, the majority of time values are less than 15-20 seconds.

Possible users

Compared with the similarity of boarding time differences (relative to cash) of travel cards and smart cards, the perception from potential travel card users on boarding time in Table 6.9 (9.24s and 9.81s) is greater than potential smart card users (6.44s and 6.27s), which means that travel cards could provide quicker boarding time than smart cards. Different ticket-checking patterns of travel cards and smart cards would be the main reason for such differences on the boarding time attribute. Travel card users only need to show valid cards to PT drivers when getting on, while smart cards need to be read by a device on board to pay the fare. In addition, the reliability of smart card reading devices might be another reason to cause such a potential difference on boarding time from travel cards.

Regarding percentile values, in each rank, generally travel cards are reported to have a quicker boarding time than smart cards as expected. Although the maximum value of boarding time difference seems very extreme for travel cards and smart cards, percentile value at 75th of travel cards shows that most data are less than 10s, and 5-6s for smart cards.

Table 6.9 Boarding Time Difference of Possible Users (cash: baseline)

Fare payment could be used	**Cash**	**Travel Cards**		**Smart Cards**	
RP questionnaire type		cash users	SC users	cash users	TC users
Actual Data entry		182	182	164	194
Average (second)		9.24s quicker than cash	9.81s quicker than cash	6.44s quicker than cash	6.27s quicker than cash
Max/Min		120/0	120/0	120/0	130/-3
Percentile 10		1	0	0	0
Percentile 25		2	2	1	1
Percentile 50		3	4	3	2
Percentile 75		10	10	5	6
Percentile 90		20	27	10	12
S.D.		18.04	16.8	15.07	12.98

(4) Seat Availability

Table 6.10 lists respondents' attitude toward seat availability by using travel cards and smart cards, relative to cash payment. As can be seen, most respondents (actual and possible card users) reported that compared with cash, there was no difference on seat availability after using card payment. So we can conclude that card payment types cannot bring better seat availability to users than cash payment.

Table 6.10 Seat Availability by using Travel Cards and Smart Cards

	Fare Payment Type	A	B	C	D
RP1	Travel cards	122	35	14	12
	Smart cards	126	31	4	2
RP2	Travel cards*	190	63	43	19
	Smart cards	134	37	19	2
RP3	Travel cards	215	66	21	1
	Smart cards*	122	33	14	13

Note: A. No difference; B. Slightly better; C. Better; D. Much better.
"*" means one kind of fare payment method is primarily used in reality. Others without "*" represent potential methods that could be primarily used

(5) More trips made using card payment than cash

Table 6.11 presents results about whether more trips were (or could be) made after using card payment methods than using cash. In general, most travel card users (including actual and possible) reported that they would like to travel more by using travel cards than cash, but the response of smart card users is very different. Most smart cards users could not make more trips even if they used smart cards. The reason for such differences between travel cards and smart cards is that: the travel card application does not have restrictions on the number of trips made, so it is not surprising that travel card users apparently could make more trips by using travel cards when they do not have to increase their travel cost Currently, the main application of smart cards is "pay as you go" cards, which means that users need to pay for each single trip. The more trips they make, the more travel costs users occur.

Table 6.11 Whether More Trips (were or could be) Made

	Fare Payment Type	Yes, I could	No, I could not
RP1: cash users	Travel cards	122	61
	Smart cards	54	109
RP2: travel card users	Travel cards*	155	160
	Smart cards	77	115
RP3: smart card users	Travel cards	112	70
	Smart cards*	129	174

Note: "*" means one kind of fare payment method is primarily used in reality. Others without "*" represent potential methods that could be primarily used

(6) Overall Assessment

Overall assessment results of respondents to the three payment methods are listed in Table 6.12. In general, most respondents reported their overall assessment of payment services as "neutral". However, the ratio of reporting 'totally unsatisfied' and 'unsatisfied' by using cash is more than that of travel cards and smart cards. More respondents said 'satisfied' with card payment than cash payment. However the overall benefit of card payment still does not seem very clear.

Table 6.12 Overall Assessment of the Three Payment Methods

	Fare Payment Type	A	B	C	D	E
RP1	Cash*	17	43	174	63	6
	Travel cards	3	7	106	56	11
	Smart cards	2	8	99	51	3
RP2	Cash	11	24	157	99	24
	Travel cards*	3	9	108	71	1
	Smart cards	12	83	181	35	4
RP3	Cash	12	72	162	54	3
	Travel cards	3	15	105	39	20
	Smart cards*	2	13	156	116	16

Note: A- Totally unsatisfied; B-Unsatisfied; C-Neutral; D-Satisfied; E-Totally satisfied.
"*" means one kind of fare payment method is primarily used in reality. Others without "*" represent those potential methods that could be primarily used.

(7) Top-up/purchase Type

Travel cards

Table 6.13 lists respondents' options to buy/renew travel cards. It can be seen that not only for actual travel card users, but also for potential users, buying/renewing their travel cards at agencies is more popular than at ticket offices. This is because the number of agencies is greater than that of ticket offices in the actual use, therefore it is more convenient for card users to buy/renew their cards through agencies. In addition, in Table 6.13, the response of using both buying/renewing methods is also considered, but compared with the single options, the number of respondents using (or potentially using) both options is very few.

Table 6.13 Choices on Methods to Buy/renew Travel Cards

	A	B	Both A&B
RP2-Actual travel card users' choices	118	196	1
RP1-Potential travel card users' choices	87	93	3
RP3-Potential travel card users' choices	71	107	4

Note: A: At ticket office; B: At agencies

Smart cards

Table 6.14 Choices on Methods to Buy/top up Smart Cards

	A	B	C	More than two methods
RP3-Actual smart card users' choices	55	191	53	A&B:3
RP1-Potential smart card users' choices	31	110	20	A&B:1 B&C:1 A&C:1
RP2-Potential smart card users' choices	27	132	32	B&C:1

Note: A: At ticket offices
B: By banks
C: By agencies

Regarding smart cards, users would prefer to buy/top up their smart cards through local banks rather than through ticket offices and agencies. The number of authorised local banks is more than that of ticket offices and agencies, so the convenience of buying/topping up is the main concern when smart card users chose their buying/topping up methods. In Table 6.14, options of choosing more than two options are considered as well, but the number is very small as in Table 6.13.

(8) Other Functions of Smart Cards

In the current smart card application, some extra functions have been implemented. Table 6.15 shows the use of these functions by actual and potential smart card users. Generally, paying for the PT fare seems to be the major use of smart cards in the RP survey, because most respondents chose "D. None. Only used for public transport fare payment". Concerns about the reliability of new functions for smart cards and less knowledge about these new functions are the two reasons for very few card users choosing these extra functions.

Table 6.15 Other Functions of Smart Cards (were or could be) Used

	A	B	C	D	More than two functions
RP3-Actual smart card users' choices	10	7	12	270	4
RP1-Potential smart card users' choices	9	6	15	133	0
RP2-Potential smart card users' choices	14	5	16	157	0

Note: A: Banking;
B: Parking/tolling fee payment
C: Shopping
D: None. Only used for public transport fare payment

6.5. Stated Preference (SP) Survey

6.5.1 the SP Survey

The main SP survey lasted for 21 days from 18th July to 17th August 2005. 1550 SP questionnaire papers were sent out. Among them, 1130 SP questionnaires were distributed on board by the survey team within these 21 days and the other 420 were given in 21 different clusters, including four schools, ten residential communities, four governmental departments and three companies. When surveying on board, the survey team tried to cover wider geographic areas and service routes in the Dalian urban area to increase the representativeness of the data and reduce the response bias.

Finally, 1184 papers were returned. Among them, 362 SP questionnaires were collected back from the in-cluster survey, with the response rate of 86.2%, and 72.7% response rate for the on board survey. It is easily understood that the response rate of in-cluster survey was higher than the on-board survey, because the respondents could take time to consider those choice situations and the second visit of the survey team also reminded the respondents in clusters to give their responses. However, as with the returned RP questionnaire papers, viewing and checking SP data to filter out invalid questionnaires is an essential preparation for the later data analysis. The criteria for checking returned SP questionnaires are as follows:

(1). Full information was required, including the RP questions prior to the SP games, full responses to every choice situation and socio-economic information. Particularly the number of trips in the RP section was used to convert the travel cost per trip (by cash or pay as you go smart cards) to cost per month. If respondents could not provide the trip information, it was impossible to measure utilities of alternatives with different cost units.

(2). Although in the SP pilot survey, the same preference in all choice situations for one individual did not happen, it was still likely to meet such responses in the large scale main survey. When viewing the returned SP questionnaires, it was found that a small quantity of this kind of response did exist. Such data can influence the accuracy of parameter estimation, because from such data the changes of respondents' perceptions towards different binary choice situations cannot be understood.

(3). Intensive checks about the rationality of individual responses: it would be common that if people want better services, they would like to pay more. However, if people simply choose an alternative with higher costs than another one, regardless of the changes in other attributes between these two alternatives, such decision would be irrational and should be rejected.

After viewing the data, 896 SP questionnaires could be used for the later data analysis. The response rate is about 57.8% on average. In this SP survey there were four games: Ex.1: cash vs. travel cards; Ex.2: cash vs. smart cards; Ex.3: travel cards vs. smart cards (pay as you go cards) and Ex.4: travel cards vs. smart cards (pay monthly cards). 120 SP questionnaires were valid for Ex.1, 311 for Ex.2, 264 for Ex.3 and 201 for Ex.4, respectively (See Table 6.16.). As can be seen in Table 6.10, the number of returned SP questionnaires is more than that of completed and valid questionnaires being used in the modelling analysis, because some of returned data were excluded according to the criteria discussed above.

Table 6.16 The response of the SP survey

	Completed and Valid Q's	Returned Q's	Handed out Q's	Response rate
SP1: cash vs. travel cards	120	162	250	48%/64.8%
SP2: cash vs. smart cards	311	399	500	62.2%/79.8%
SP3: travel cards vs. smart cards (pay as you go)	264	338	400	66%/84.5%
SP4: travel cards vs. smart cards (pay monthly)	201	285	400	50.2%/71.2%
Total	896	1184	1550	57.8%/76.3%

In general, the respondents understood the binary choice situations in the SP games.

However some of them queried the questions about stated choices, because they had very few chances to trade off their hypothetical choice situations in their real life, also the respondents sometimes asked surveyors whether those hypothetical situations (including some new variations of attributes) could be realised in reality and when. Under such circumstances, the survey team had to explain these questions and provide evidence of those successful SP applications on forecasting user demand in transportation studies.

Two kinds of survey methods were used, on board survey and in cluster survey. Generally, the performance of in cluster survey was better than on board survey in the response rate, owing to the ease of targeting respondents in clusters during sending and collecting questionnaires, and sufficient time for the respondents to consider choice situations. But one of drawbacks of the in-cluster survey was found when checking returned questionnaires from the same cluster: a respondent's choices could be affected by other members in the cluster, such as colleagues, friends or classmates, etc. although their socio-economic backgrounds may be different. Compared with the in-cluster survey, survey duration of the on-board survey was a cause of uncompleted questionnaires for some respondents, who got insufficient time to do the SP survey due to their short journey.

6.5.2 Basic Characteristics of the SP Data

(1) User Type in the SP Survey

The total number of valid SP questionnaires is 896. In the SP questionnaires designed, the first task is to determine user type of respondents (i.e., which kind of fare payment method they used in last month? cash, or travel cards, or smart cards) so as to send proper SP questionnaire papers to the respondents.

Table 6.17 Fare Payment User Type (Primary) in All SP Exercises

	Cash users	Travel card users	Smart card users	Total
SP1: cash vs. travel cards	46	74	0	120
SP2: cash vs. smart cards	133	26	152	311
SP3: travel cards vs. smart cards (Pay as you go)	0	165	99	264
SP4: travel cards vs. smart cards (pay monthly)	2	123	76	201

As can be seen in Table 6.17, most user types can meet the requirement for sending proper SP questionnaires in Chapter 5: Survey Design. In Chapter 5, it was required that respondents who took part in the SP game must understand or primarily use at least one of two alternatives in the SP game so that preferences between their current payment methods and hypothetical ones can be examined. For example, for SP 1, 120 respondents reported that they primarily used cash or travel cards in the last month as they needed to trade off cash and travel cards in SP 1. However, in SP 2 and SP4, a small proportion of users (26 travel card users in SP 2; 2 cash users in SP 4) did not belong to either user types being traded off in SP games but were still included in the data set because the data quality was acceptable after checking the data.

(2) Fare Payment Type Used

Cash:

The distribution of the respondents' cash fare types in the SP is consistent with the RP. Most passengers used flat fare. Such similarity could indicate the homogeneity of the RP sample and SP sample on fare payment choices. We can also find this phenomenon in Table 6.19 and 6.20 for travel card and smart card types respectively. The respondents' choices on card types (travel cards/smart cards) are similar to what we find in the RP survey.

Also combinations of different fare payment methods can be identified through choices of fare payment types in Table 6.18-6.20. For example, in Table 6.18, the total number of respondents who used cash fare ticket types in SP 2 is 261 (=67+66+65+63), more than the number of primary cash users in SP2 (133 individuals in Table 6.17). That means 128 of 261 respondents used the combination of cash and smart cards. Moreover, considering fare payment combination in the RP survey design in Chapter 5 was proved reasonable.

Table 6.18 Cash Fare Payment Choices in SP1 and SP2

	SP 1	SP 2-1	SP 2-2	SP 2-3	SP 2-4
Flat Fare	83	46	56	44	51
Zonal Fare	33	21	10	21	12
Total	116	67	66	65	63

Travel cards:

Table 6.19 Travel Card Type Choices in SP1, SP2 and SP4

	SP 1	SP 3-1	SP 3-2	SP 3-3	SP 4-1	SP 4-2
A	60	49	55	45	48	49
B	10	14	9	10	17	15
C	10	11	12	5	7	7
D	4	2	6	4	10	9
Total	84	76	82	64	83	80

Note: A: Monthly cards with limited bus route
B: Quarterly cards with limited bus route
C: Monthly cards with unlimited bus route
D: Quarterly cards with unlimited bus route

Smart cards:

Table 6.20 Smart Card Type Choices in SP2, SP3 and SP4

	Ex.2-1	Ex.2-2	Ex.2-3	Ex.2-4	Ex.3-1	Ex.3-2	Ex.3-3	Ex.4-1	Ex.4-2
A	54	46	46	48	54	71	59	69	72
B	0	2	0	0	7	0	2	1	0
C	0	0	0	0	2	0	0	1	2
D	0	1	0	0	0	0	0	0	0

Note: A: "Pay as you go" card
B: Electronic travel card (a minimum payment required per month)
C: Student smart card
D: Elder smart card

(3) Number of trips by using different payment methods

Table 6.21-6.23 list statistic characteristics on the number of trips by the three different fare payment methods. As can be seen from Table 6.22, the number of trips by travel cards is about 43-52 in one month on average, more than the other two payment methods. Because travel card users can use travel cards without any limit in one month and cash users and pay as you go smart card users have to pay for each single trip, it is a very common phenomenon that more trips could happen by using travel cards. Moreover, the percentile values can also reflect the difference of number of trips between three payment methods. In general, at percentile 75 and 90 level, the numbers of trips of travel cards are more than those of cash and smart cards.

Table 6.21 Statistic Characteristics on Number of Trips by Cash

	SP 1	SP 2-1	SP 2-2	SP 2-3	SP 2-4
Average	19. 51	20.65	22.42	23.63	19.25
S.D.	15.61	19.89	21.86	20.55	15.32
Max/Min	80/2	90/2	102/2	90/2	60/2
Percentile 10	6	4	5	5	5
Percentile 25	10	6	8	10	8
Percentile 50	15	20	15	18	15
Percentile 75	20	30	30	30	25
Percentile 90	80	90	102	90	60

Table 6.22 Statistic Characteristics on Number of Trips by Travel Cards

	SP 1	SP 3-1	SP 3-2	SP 3-3	SP 4-1	SP 4-2
Average	49	45.63	43.73	52.25	51.89	50.81
S.D.	24.67	21.09	18.96	23.23	28.17	24.73
Max/min	120/8	120/5	120/15	150/20	180/4	150/7
Percentile 10	20	20	20	25	20	20
Percentile 25	30	30	30	40	40	40
Percentile 50	48	47.5	44	50	50	50
Percentile 75	63.75	60	60	60	60	60
Percentile 90	80	66.5	60	75	78.4	79.8

Table 6.23 Statistic Characteristics on Number of Trips by Smart Cards

	Ex.2-1	Ex.2-2	Ex.2-3	Ex.2-4	Ex.3-1	Ex.3-2	Ex.3-3	Ex.4-1	Ex.4-2
Average	27.03	35.94	39.35	37.96	36.68	38.45	27.84	34.60	37.23
S.D.	19.84	22.32	24.48	23.38	20.44	18.55	16.21	17.74	24.95
Max/min	100/5	120/5	104/5	100/8	80/5	80/4	70/5	70/5	100/4
P10	5. 5	8	5. 7	10	10	10	10	10	10
P25	10	17.5	20	20	20	28	12.5	20	13.75
P50	20	35	40	30	30	40	28	30	30
P75	40	50	50	55.75	60	50	42.5	50	60
P90	55	60	60	71	60	60	50	60	70

(4) Market Share of Payment Type in SP Responses

Finally, through the SP data, we can obtain the observed responses in these SP experiments before the modeling analysis. Table 6.24 lists the number of choices of three payment methods. Market shares of the three payment methods are also presented in Table 6.24. By comparing with observed shares in the RP survey, we find that the shares of smart cards in two surveys are very different. The share of cash payment in the SP is much smaller than in the RP. The share of travel cards in the SP is very close in the RP survey. Therefore it can be implied that under hypothetical situations, some cash users may switch to other payment methods (e.g., smart cards) and the use of travel cards would remain the same level as the current application. A possible reason for such difference may be the different trade off situations between the RP and SP survey. Particularly for the smart card ticketing in the SP survey, some new features and variations of smart cards were introduced, therefore it is possible for some traditional payment users to switch to smart cards.

Table 6.24 Share of Payment Type in the SP Survey

	Cash	Travel Cards	Smart Cards	Total
No. of Responses	935	2219	3059	6213
Market shares in SP	15.05%	35.72%	49.23%	100%
Observed RP	33.3%	33.9%	32.8%	100%

6.6. Basic Characteristics of the Respondents

At the end of the RP and SP surveys, socio-economic questions were included to collect respondents' personal information, including six background questions (age, gender, education level, employment status, household income per month, availability of private transport) and one attitudinal question about willingness to prepay for PT fares. After the discussion of basic characteristics of the RP and SP data in Section 6.4 and 6.5, in this section, 1765 respondents' characteristics (869 in the RP survey, 896 in the SP survey) are described from those seven aspects (socio-economic variables) aforementioned.

6.6.1 Age

First of all, as can be seen in Figure 6.1 for the age distribution, among the five categories designed, the distributions of the respondents' age in the first four age intervals are very closely comparable, ranging from 22%-26% for each group. But the percentage of respondents over 60 years old is far less than other four age groups with a proportion of about 2%. Several reasons can be explained for the age distribution:

(1) The population we defined for this survey is public transport users in Dalian urban area and on board surveys were mainly used to send and collect questionnaire papers. However, because normally the elder group's travel demand is lower than the younger group due to their physical condition, travel frequency and travel purpose. Particularly during peak time, it would be common that the number of the elderly

people was far less than the other four younger groups when the survey was conducted.

(2) According to the census data in Dalian Statistical Yearbook (2005), the percentage of people over 65 years old was 10.9%. 77.18% people were 15-64 yeas in 2005. Although the percentage of people over 60 in this survey is about 2%, less than official statistics, the difference is still explainable and acceptable.

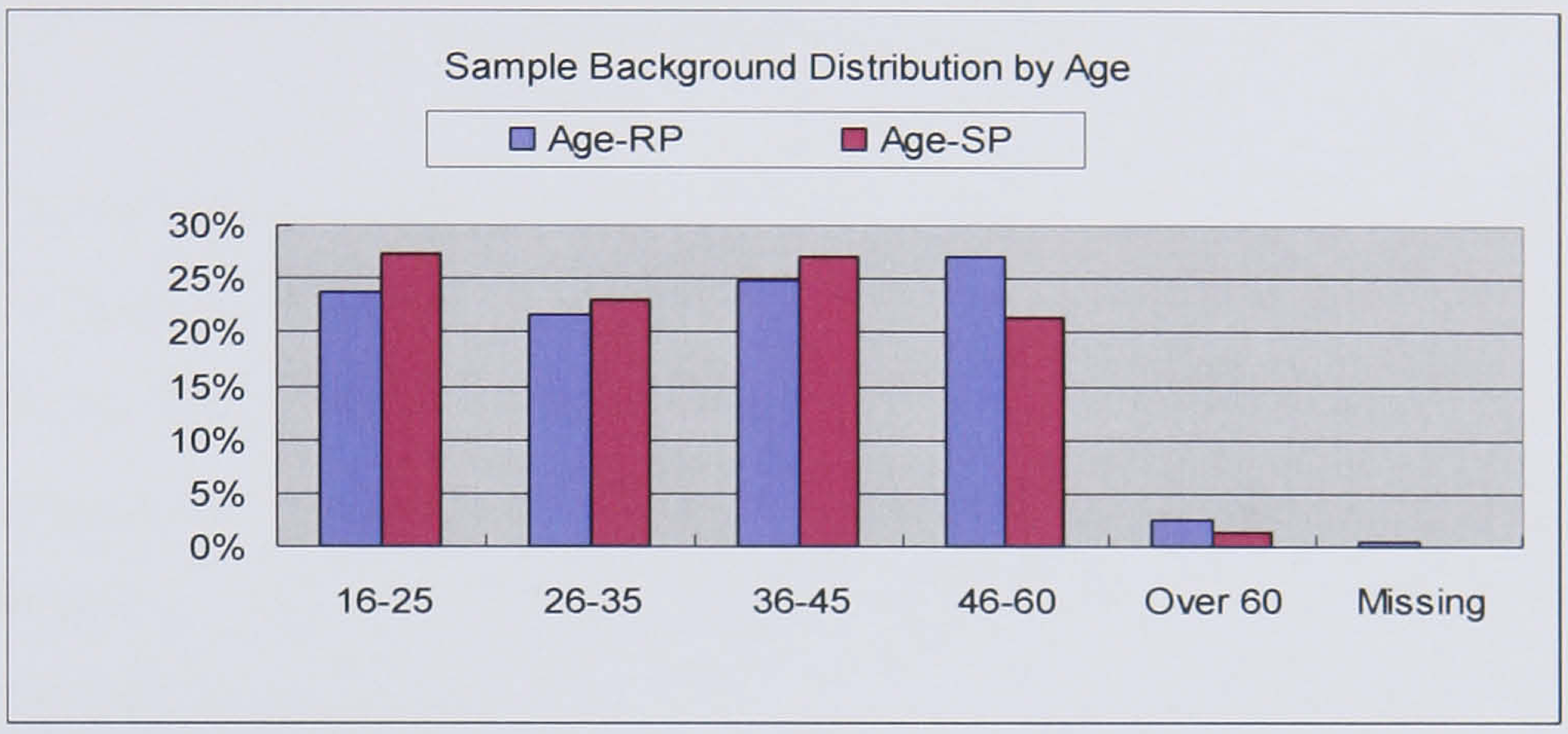


Figure 6.1 Sample Background Distribution by Age

6.6.2 Gender

The gender distribution in the both RP and SP surveys can reflect the reality of the society (See Figure 6.2), composed of 48.9% male and 51.1% female overall (data missing for 23 respondents), which also are close to the official percentages about sex in Dalian in 2005 (50.61% male and 49.39% female, Dalian Statistical Yearbook 2005), except that the number of females in the survey is slightly more than that of males. A possible reason for this result is a little bit more involvement of females than males in the survey. Female respondents seemed more patient when surveyors expressed the intention of carrying out the survey and were willing to take part in the survey.

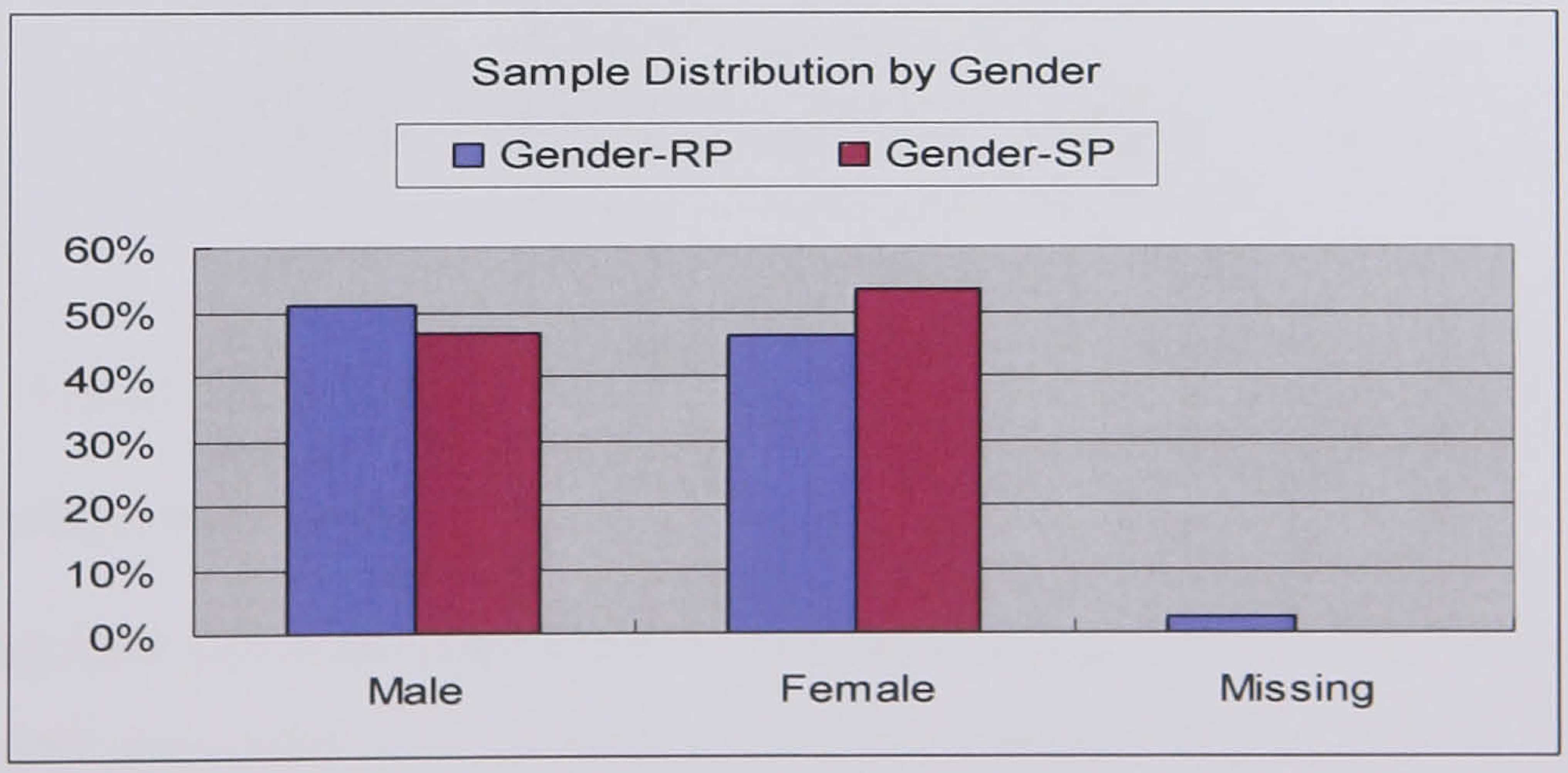


Figure 6.2 Sample Background Distribution by Gender

6.6.3 Educational Level

The distribution of educational level of respondents is presented in Figure 6.3. As can be seen from Figure 6.3, the percentage of respondents with educational level of high school or less is roughly 40% overall. About 60% of all respondents reported having some college education or more, in which about 30% were college graduates, about 25% undergraduates and about 5% postgraduate or equivalent. In general, the average educational attainment was higher than the official statistics from Dalian Statistical Yearbook (2005) (about 11.5% Dalian citizens having college degree or above, 88.5% having high school degree or less). There are several reasons why the average educational levels would be higher than the census data:

(1) The census surveyed educational levels for all people who were 6 years old or above. But in this survey, people ages less than 16 years were excluded so that it could make the group of people with higher educational backgrounds become relatively higher.

(2) Surveying in some clusters such as companies, universities, governmental departments, *etc*, which had a higher proportion of respondents with higher educational backgrounds?

(3) When determining the user type (i.e., cash user, or travel card user or smart card user) in the RP and SP survey, those better-educated respondents were more likely to report they were smart card users or they would use fare payment option with high technology. However in the survey, we needed to balance the number of different user types, so the possible fact that more smart card users having higher educational levels would influence the distribution of the overall educational level.

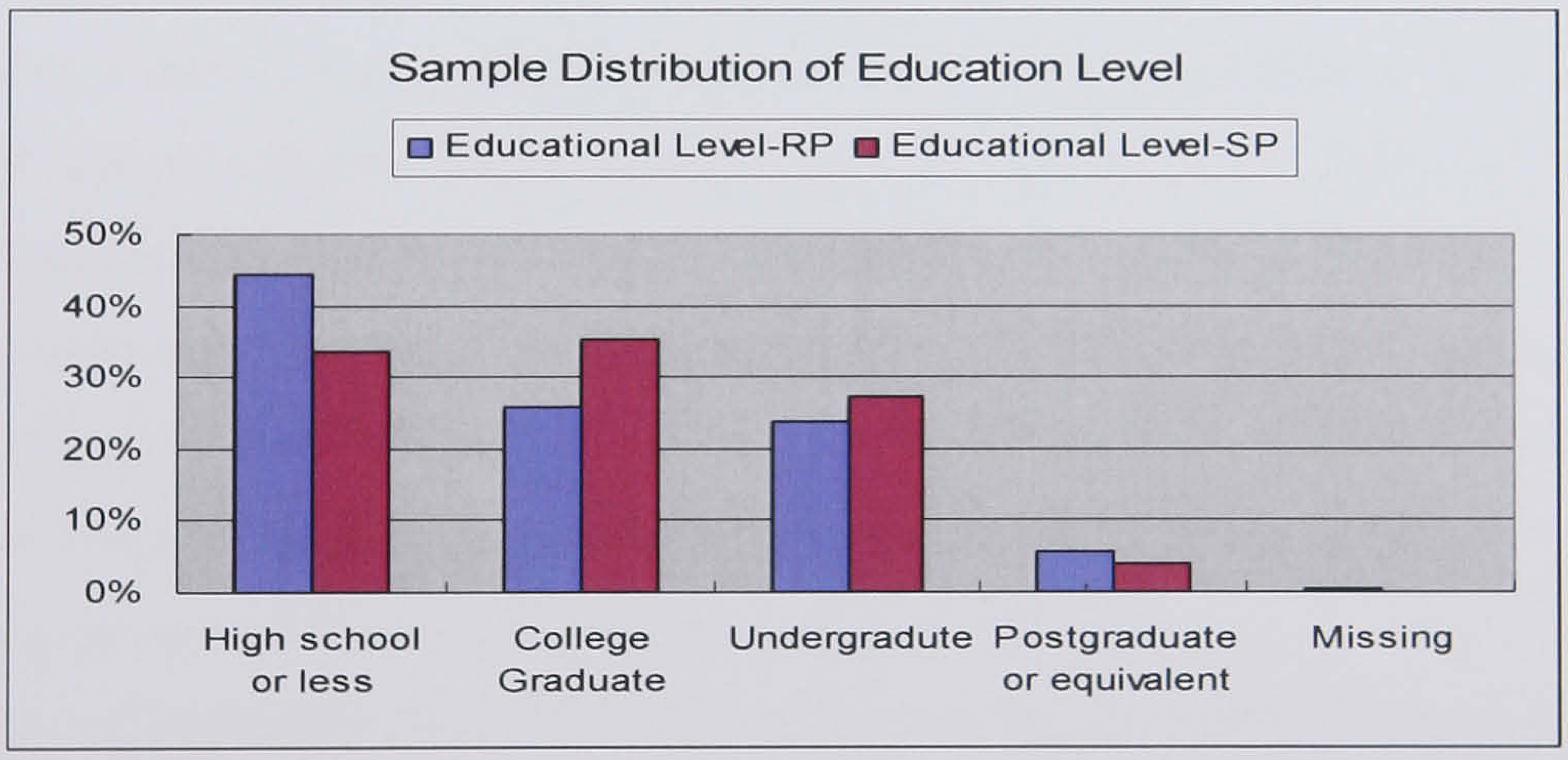


Figure 6.3 Sample Background Distribution by Educational Level

6.6.4 Employment Status

Employment status may be important to understand in order to explain differences in PT users' payment preferences. As can be seen in Figure 6.4, people who are engaged in a full time job occupy 66.7% and the percentage of students is 13.4%, including those with and without jobs. This distribution is consistent with the survey methods used, although the percentage of full time employee from the census data was about 36% in Dalian (Dalian Statistical Yearbook, 2005). During peak time, it is common that the proportion of people with full time jobs is much higher than passengers with other employment status. In addition, those clusters selected, such as companies and governmental departments, certainly had a greater number of full time employees. Another employment status that can be compared between the official census data and the survey data in this research is the proportion of unemployed people. About 8% of unemployed respondents are higher than the official unemployment rate of 3.4%. Surveying in some residential communities could cause this difference, because those people available for the survey during work days in residential communities were possibly unemployed.

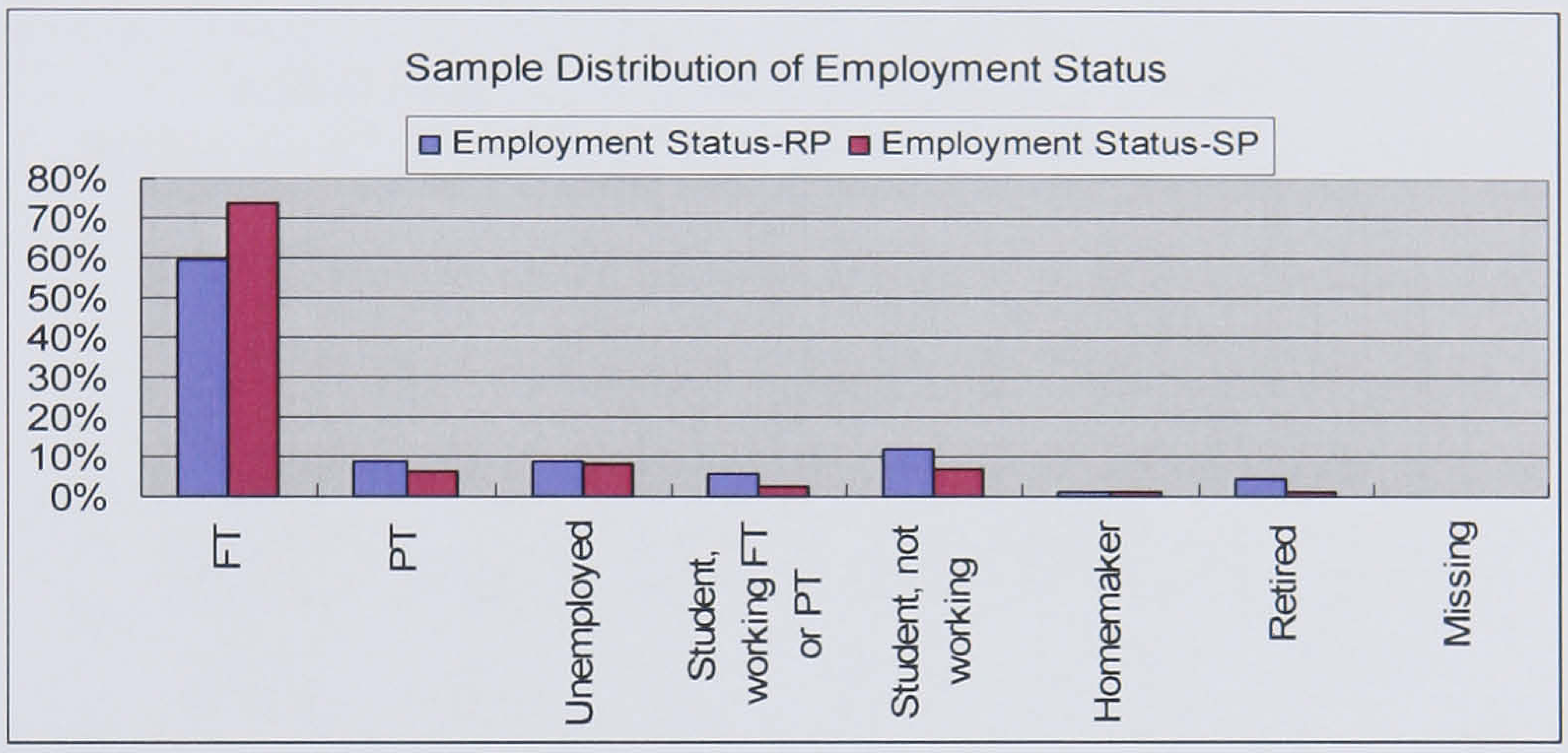


Figure 6.4 Sample Background Distribution by Employment Status

6.6.5 Household Income

Household income per month is one indicator of respondents' preferences to different fare payment. Figure 6.5 illustrates the distribution of household income per month in the RP and SP surveys. Nearly 45.5% household earn less than 1500 yuan per month, and 26.9% of household's income is between 1500-2999 yuan per month. Especially, in the RP survey, the percentage of household income less than 1500yuan per month is more than 50%. The proportion of household income more than 4000yuan per month only occupies about 10% of the sample. The census data shows that in 2005 the average household income per person per year was about 11994yuan (equivalent to 999yuan per person per month). The average size of household was 2.74 people in 2005 in Dalian (Dalian Statistical Yearbook, 2005). Therefore we can use 999yuan to multiply 2.74 to get the average household income per month (about 2738yuan per month, wherein the range of 1500-2999yuan for the majority). But the majority of household in this survey are in the range of 'less than 1500yuan'. Several reasons could cause this: (1) in this survey we did not identify marital status of respondents, so it was highly possible to survey a single person, especially aged between 16-25 years old. Therefore such people would only report his/her own income as household income; (2) we have identified employment status of respondents, about 34% of whom were low paid or did not have a stable income, such as part time workers, unemployed, students, retired, etc. Their report about household income also could influence the distribution of the overall household income.

Another aspect that can be seen in Figure 6.5 is that the household income in Dalian is relatively low, compared with other major cities in China (e.g., Beijing, Shanghai, Shenzhen, *etc.*). For example, in Beijing, the household income per person per month was more than 4000yuan in 2005 (Beijing Statistical Yearbook, 2005). The development of local economy, industrial structure, policies of the central government, etc., could bring such difference on household income.

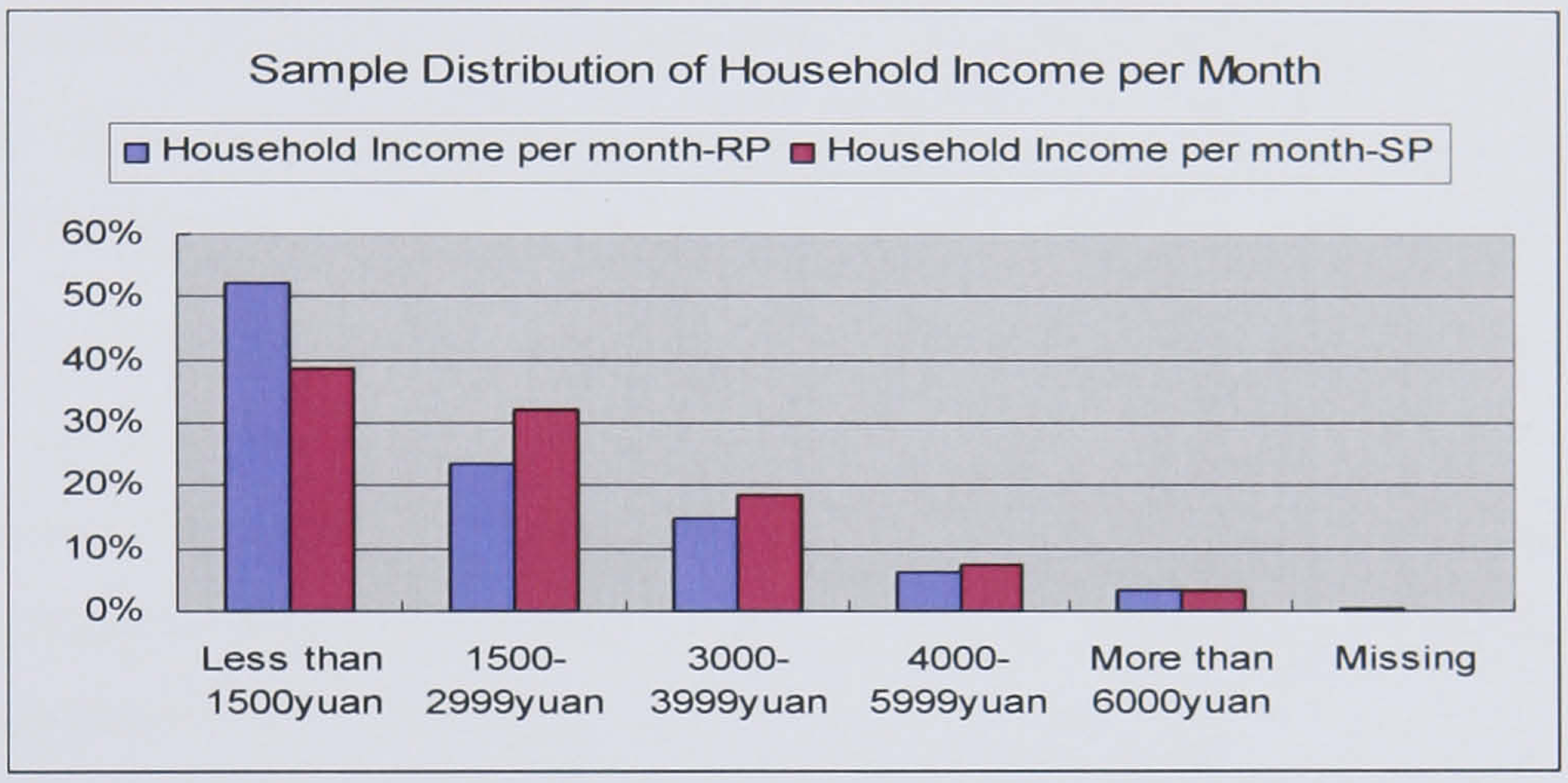


Figure 6.5 Sample Background Distribution by Household Income per Month

6.6.6 Availability of Private Transport

Figure 6.6 reveals that private transport availability is quite low in Dalian (only 8% of respondents always or most of the time used private transport).

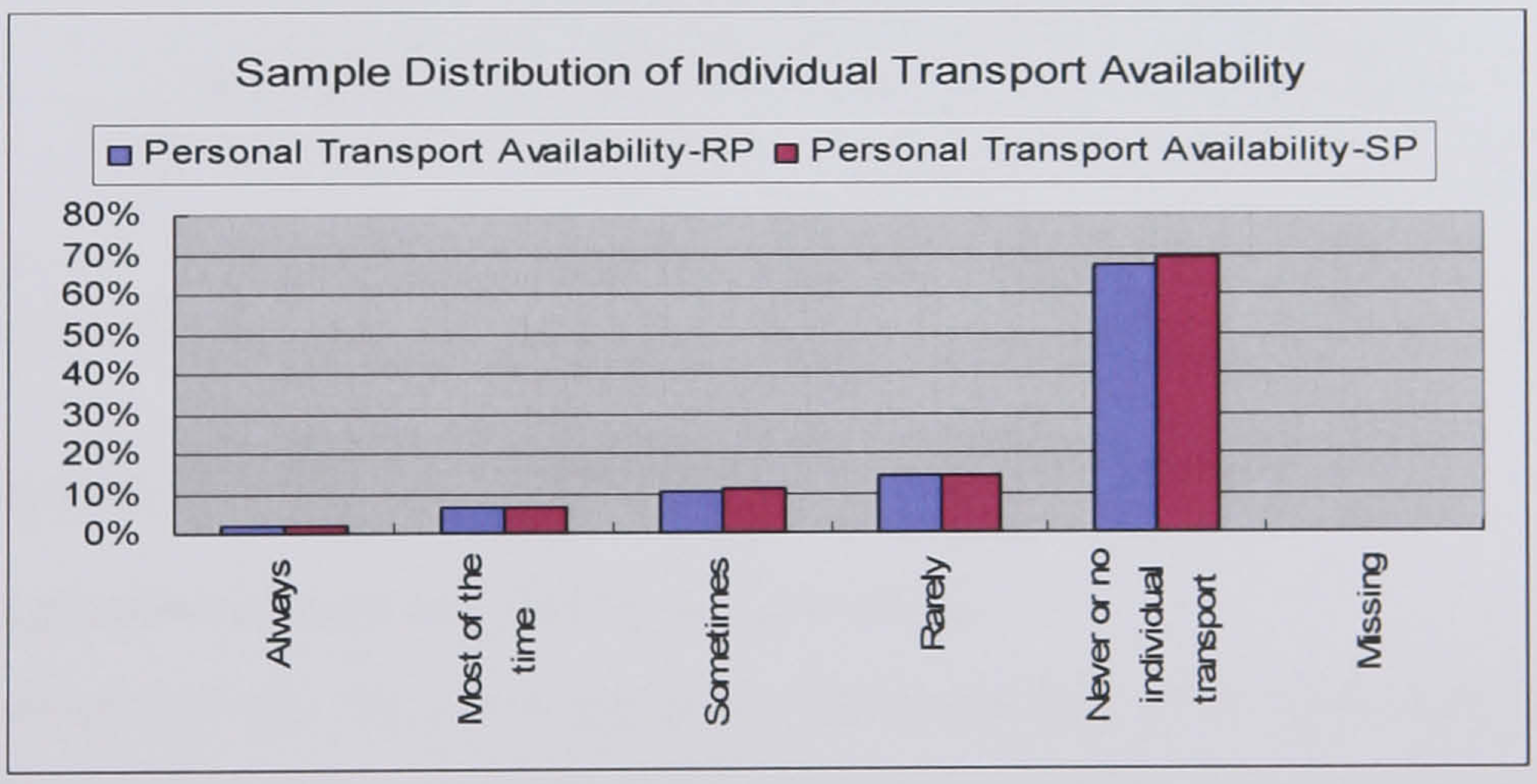


Figure 6.6 Sample Background Distribution by Individual Transport Availability

Although it was mainly public transport users who were surveyed, we find some respondents indicated that they not only used public transport, but also had private transport available to them, such as car sharing (we can assume that it is a kind of individual transport). Another group of people we may notice is 'always private transport users' (about 2%). They primarily used their private transport in the last month, but as to public transport fare payment, they still can provide some information. About 70% of respondents reported that they never use or did not have private transport. Private vehicle availability is an indicator of public transport dependency – those who do not have access to a vehicle are more likely to rely on PT for their mobility needs. For this point, it could be safely inferred that public transport is still the principal mode for the majority in Dalian.

6.6.7 Willingness to prepay PT Fare

As to the willingness-to-prepay for fare in Figure 6.7, more than 57.4% respondents would like to prepay fares, 34% of which chose pre pay per month. The reason for this result could be that at present most current prepaid fare users (say, travel card users) use monthly travel cards and to some extent their existing payment behaviours would influence their willingness-to-prepay attitude. In addition, 42.6% respondents reported that they would not like to prepay fares. That means they prefer paying fare each time for each single trip.

Another interesting thing that can be seen from Figure 6.7 is that in the SP survey, people's attitudes to prepaid fares (e.g. prepay per week, prepay per month, *etc*) are more positive than in the RP survey. A possible reason is that in SP games those new features for prepaid payment choices could influence the respondents' attitudes about prepaid fare policy when they traded off SP situations.

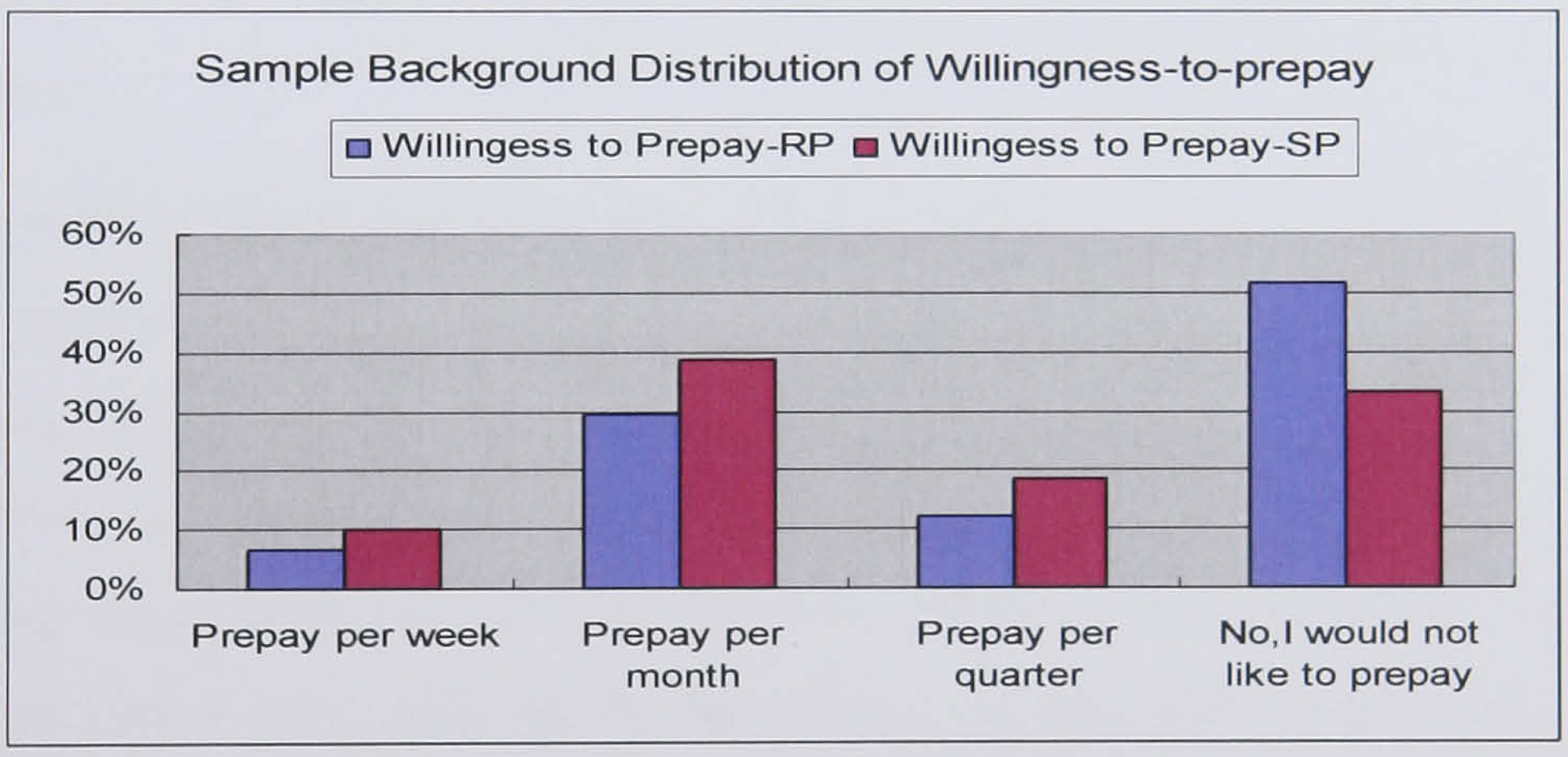


Figure 6.7 Sample Background Distribution by Willingness to Prepay PT Fare

6.7. Issues in the Data Collection

1).The respondents' understanding to questions

The major problem for the respondents' understanding of the questions is that when those conditional questions were presented to them in the RP survey, e.g. "Would you use smart card in the last month if they were available to you?", most respondents were not sure about the meaning of those conditional questions. It seemed that they were only concerned about their existing situation (in the both the pilot survey and main survey, surveyors met such problems). Under this circumstance, surveyors had to explain this to the respondents. The meaning of these conditional questions was to identify information about respondents' choices, including their actual choices, and alternatives not actually used, but they could use them if available.

Another reason for the misunderstanding in the questions is that in order to reduce the respondents' fatigue when they answer questions and to save pages of questionnaire, a brief instruction had been presented to the respondents. In fact, during practical surveys most respondents could not spare much time to read the instructions thoroughly, which meant that it was likely to result in misunderstanding.

2). The SP method to collect data in China

"Why do you design these hypothetical situations and ask me to trade off?" "What do those binary-choice situations stand for?". The respondents often asked surveyors above or similar questions when they filled in the SP questionnaire papers. Unfamiliarity with the SP method was the main issue when carrying out the SP survey in Dalian, China.

Once in the pilot survey, some brief explanation about the SP method had been given when respondents answered the questionnaires, but it looked very difficult for the respondents to catch the meaning of this question style. So in the main survey more details about the survey were presented (including features of alternatives needed to trade off) to them if needed. Through this data collection, we find some respondents took more time to think about the SP survey method itself before they commenced to answer the questionnaires.

3). Data quality: removing poor and illogical data

As discussed in Section 6.4 and 6.5, the number of questionnaire papers collected back was more than those papers with proper and logical data, which can be used in the future analysis. When all questionnaire papers had been returned, the next task is to check and remove those illogical and invalid data from the database. For those data with only single fare payment option (i.e., cash) in the RP survey, we still keep them in the data set, but exclude them in the discrete modelling stage, but recover them in the demand forecasting stage, because these data can only offer us certain information that the individual probability of choosing cash is 100%. Definition of invalid data and rules to filter good data being used in the future are as follows:

a). Illogical data:

Background data: through analysing the relationship among age, occupation, private transport availability and so forth, we can tell whether these data are logical or not. For instance, it is obviously impossible that a mature respondent, who has a full time job, uses student smart cards.

Travel behavioural data: we can tell from the relationship among ridership and the way used to pay fare, etc. For example, in the RP survey, first of all, a respondent indicated that he/she mainly used smart cards for travelling, while cash was viewed as his/her secondary fare payment mean, but when answering the question of travel cost in the last month, he/she showed that the cost of cash was far more than smart cards. The details about these situations in the RP survey are listed in Table 6.25. All these illogical data indicated that these respondents would have less cooperation and they did not take seriously to answer the questionnaires. Another reason could be that some respondents still did not fully understand the questions.

For the SP data particularly, some illogical data still existed in the relationship between travel cost and travel frequency, and some other fare payment attributes. When converting the travel cost of yuan per trip to yuan per month, we referred respondents' existing number of trips in one month by using different payment options. For example, for those occasional PT users, their total travel cost by cash or pay as you go smart cards would be lower than the cost of travel cards. But some respondents preferred one alternative with a much higher travel cost than another in a choice situation, instead of comparing other attributes synthetically. Such irrational choice data cannot show the influence of other attributes and levels to people payment choices. Also, it could impact on the significance of the model estimation in the next stage.

Table 6.25 Illogical Background Data in the RP Survey

	Age	**Employment**	**Income**	**Individual transport avail.**	**Ticket types**	**No. of trips in the last month**	**Travel cost**
1	Mature	FT	--	--	Student travel cards	--	--
2	--	FT	Low	Unavailable	--	Very few	High
3	--	--	--	--	--	Many	Low
4	--	--	--	--	--	Alternative "*A*" most of time; "*B*" is secondary	Cost of "***B***" is more than "***A***"

b). Incomplete questionnaire papers:

Some of the returned questionnaire papers were not completed, and some important information, such as travel cost, boarding time difference, etc., which will be used in the data analysis, were missing. Such data had been excluded in the data set to be used in the future analysis.

c). Incomplete questionnaire papers with missing those minor questions:

Those data, which missed some minor information, but all those important questions were fully answered by the respondents, have been kept in the data set. For example, the data from some respondents who overlooked/or refused to answer some socio-economic questions (e.g., age, sex, etc.).

4). Copied data:

Although some questionnaire papers were completed and proper information was provided by different respondents, most answers were obviously identical with each other, which were identified by the survey location they were surveyed, questionnaire labels and even their handwriting. Under this situation, we needed to pick up those same responses from the same survey location and keep one and delete the rest. One of the possible reasons for this problem is that respondents could influence others decision making, particularly in the same cluster, such as companies, schools, etc. Fortunately the total number of such questionnaire papers was only a small proportion (12 questionnaires in the RP; 34 in the SP).

Table 6.26 lists the outline of returned data quality. In general, compared with 24.3% rejection rate in the SP returned questionnaires, the data quality of the RP survey is better than that of the SP survey, with the rejection rate of 14.5% in the RP. As can be seen, single option response and incomplete questionnaires in the RP survey are the main issues in the invalid data. For single option responses, this is a fact that such passengers did exist in reality, for example, occasional PT users would only use cash without any other preferences. Insufficient survey time was the main reason to cause incomplete RP questionnaire papers. In the returned SP questionnaires, illogical response was the main reason to reject returned SP questionnaires. Incomplete questionnaires were another major reason to remove un-useful SP data.

Table 6.26 Summary of the Distribution of Different RP and SP Data

	Total Returned	Useful Data	Illogical	Incomplete	Copied Data	Single Option
RP	1016	869	34	49	12	52
SP	1184	896	108	91	34	55

6.8. Summary

In summary, this chapter discussed the data collection that was carried out in Dalian, China to obtain public transport passengers' preference (RP and SP) data about fare payment methods. With the cooperation of local authority and public transport operators, 869 RP questionnaires and 896 SP questionnaires could be used in the later data analysis, with the response rate of about 58% overall. Meanwhile, some basic preparation for the further data analysis have been done, including analysing basic characteristics of the RP and SP data, respondents' characteristics and checking the validity of the RP and SP data. Particularly through analysing types of users who took part in the survey, we can get some evidence of the existing market share of different payment methods in Dalian. Additionally, it is comparable with the forecast results of market share in the next chapter.

Through the data collection, some issues relevant to the RP and SP survey are discussed in this chapter. We find that in general RP questions can be well understood by respondents, except some conditional questions in the RP survey. For the SP survey, the main problem is the unfamiliarity of the SP survey method itself based on hypothetical situations. When checking the validity of the RP and SP data, we observe that illogical responses, uncompleted questionnaire papers and copied data are the three main reasons to reject some returned questionnaires.

From this chapter, findings can be summarised as follows:

1). About survey methods: in the data collection, two survey methods, RP and SP, were used. In general, respondents' understanding about the RP survey was better than the SP survey. This may be because the RP survey was based on respondents' actual behaviour, while the SP involved hypothetical trade off situation. Therefore, during the SP survey, more efforts to explain the whole picture of the survey were made by the survey team.

2). About alternatives in the RP and SP responses: in the SP survey, responses presented a variety of choices of different payment methods. However, single option of cash fare payment existed in the RP survey due to the unavailability of other payment alternatives (travel cards/smart cards) or lack of knowledge of other payment options. Such data in the RP were excluded in the following data analysis, but retained in the demand forecast analysis later on.

3). About characteristics of respondents: in general, it can be seen that the distribution of socio-economic data of respondents were satisfactory and reflected the reality of current Chinese society. But because this data collection focused on public transport users, the number of respondents who had less accessibility or weak mobility, such as the elderly, was relatively fewer than other categories. The aims of investigating socio-economic backgrounds of respondents are that: first of all, it can be used to examine the validity of data before the modelling stage through the statistical distribution. Secondly, the background data are to be employed in the segmentation analysis so as to identify the heterogeneity and homogeneity of responses.

After obtaining preference data in Chapter 6, the following task is to analyse individual preference data and get aggregated user demand (market shares) towards different payment options. Therefore, the relevant choice modelling analysis is carried out in Chapter 7 and 8. Further work is to model preference data in chapter 7 with Logit models and evaluate benefits of smart cards based on attributes and levels designed in the chapter 5. Meanwhile, some techniques as an alternative to logit models for improving estimate results of choice models (e.g., fuzzy logic and neural network technique) will be explored in Chapter 8.

Chapter 7
Data Analysis with Logit Models

7.1. Introduction

In the previous two chapters, we have discussed the survey design and data collection for this research. This chapter will model the preference data (RP and SP data) collected in Dalian, China. The purpose of this chapter is to have an insight into individual preference data through the logit model analysis. The benefits and effectiveness of smart cards are finally measured by fare payment demand forecast, importance of attributes, valuation of attributes and travel cost elasticities, etc.

In Chapter 5, the primary task was to determine the choice set being studied (cash, travel cards and smart cards). In the mean time, the survey design also determines attributes (features) and levels for the SP experiments, which are used as variables in the modelling analysis.

In Chapter 6: Data Collection, some basic preparations for modelling analysis of the RP and SP data have been done. Filtering and sorting out quality data is essential before the data analysis. Some criteria have been taken into account when doing this work:

- Data with missing important information should be excluded from the database being analysed;
- Illogical data, even with full information, also should be identified and removed from the data set being used for the data analysis;
- Finally, abnormal travel behavioural data, in which respondents' socio-economic background was not consistent with what they answered, also are not considered in the following data analysis.

After the data compilation in Chapter 6, the next task is to conduct data analysis in this chapter, which is organised as follows: Section 7.2 presents the modelling analysis of the RP data with MNL models. Following the RP data analysis, the SP data analysis with MNL models is discussed in Section 7.3. Besides the analysis for independent SP data sets, modelling the combined RP and SP data is also discussed by two different estimation approaches: sequential estimation and simultaneous estimation in Section 7.4. In Section 7.5, outputs for measuring the RP and SP data are presented by discussing valuation of attributes and fare elasticities. Following examining individual choice behaviour in previous sections, at the end of Section 7.6 different fare payment market shares are calculated on aggregate level. In Section 7.7, the model validation process is examined by using the RP and SP data retained before the MNL model estimation. These data, which did not involve the model estimation, can test the forecasting ability of the MNL we estimated. Finally, in order to identify the effect of socio-economic background of individuals on their choice behaviour, in Section 7.8 the segmentation analysis is carried out by using demographic data we obtained in the data collection.

7.2. Modelling RP Data

7.2.1. RP Input Data Preparation and Coding Scheme

Coding and Defining Dummy Variables

Prior to modelling the RP data, the first task is to prepare for the proper input data format, which is suitable for the model estimation in ALOGIT software. In the RP survey design, the following attributes for three fare payment methods are considered including: travel cost, boarding time, ticket type, overall assessment, whether more trips happened comparing with cash, seat availability, top-up/purchase methods, easiness of top-up/purchase and multifunction. Among these variables, travel cost and boarding time are quantitative and the software can recognise such data format, therefore, quantitative variables are kept the original data format in the data file. But for the rest variables, which are presented by qualitative categories, according to respondents' individual perceptions, need to be redefined and coded as dummy variables suitable for the estimation of ALOGIT.

As to the definition of dummy variables for the data analysis, the following rules are applied as discussed by Louviere, Hensher and Swait (2000):

If L is the number of levels the attribute has, the number of dummy variables for a single qualitative attribute can be defined as L-1 dummy variables, each of which represents one of the attribute's L-1 levels. Therefore, any arbitrary subset of L-1 of the L levels can be represented as follows:

- Create a dummy variable D_1, such that if the treatment contains the first level selected, D_1=1, otherwise D_1=0;
- Create a second dummy variable D_2, such that if the treatment contains the second level selected, D_2=1, otherwise D_2=0;
- Continue in this fashion until L-1 dummies are created, i.e., D_1, D_2... and D_{L-1}.

In categories of the qualitative attributes, the level is not coded into the dummy variables (because of L levels in the attribute and L-1 dummies being used) is regarded as the base, which means that all dummies are compared with the base on the sign (positive/negative) and size of the coefficient to indicate choosing a given dummy is better than the base or not. For example, variable of seat availability has four categories and three dummies are used. The base is defined as "No difference from Cash", and three dummies represent "Slightly better", "Better" and "Much better", respectively. If a respondent selects "better", then the dummy variable will be "1" for this level and others will be "0". In addition, estimated coefficients for dummies can be explained by comparing with the base. Because the worst level is set as the base for seat availability attribute, coefficients of dummies should present positive values to indicate respondents' perception for better service quality. The better the seat availability is, the greater the absolute value of the coefficient is. The individual utility for this alternative also is increased with the increase of coefficients of dummies.

The detailed variable coding scheme for each fare payment method is listed as follows (Table 7.1-7.3).

Table 7.1 Variables and Codes for Cash Fare Payment

Variables	Explanation
1). Travel cost (Yuan)	Cost per month
2). Boarding time (seconds)	It is set "0", because the time difference is used to show how much quicker by using travel cards or smart cards than cash. The detailed values of time difference is put in the boarding time variables of travel cards and smart cards
3). Ticket type *(Dummy variables)*	0: Flat fare 1: Zonal fare
4). Overall assessment *(Dummy variable)*	0: Totally unsatisfied 1: Satisfied 2: Neutral 3: Satisfied 4: Totally satisfied

Table 7.2 Variables and Codes for Travel Card Fare Payment

Variables	Explanation
1). Travel cost (Yuan)	Cost per month
2). Boarding time (Seconds)	It is set the time difference between cash to show how much quicker by using travel card than cash
3). Ticket type *(Dummy variable)*	0: Monthly cards with limited bus route 1: Quarterly cards with limited bus route 2: Monthly cards with unlimited bus route 3: Quarterly cards with unlimited bus route
4). Overall assessment *(Dummy variable)*	0: Totally unsatisfied 1: Satisfied 2: Neutral 3: Satisfied 4: Totally satisfied
5). Whether more trips happened by using travel cards *(Dummy variable)*	0: No 1: Yes
6). Seat availability by using travel cards (compared with cash) *(Dummy variable)*	0: No difference with cash 1: Slightly better 2: Better 3: Much better
7). Top-up or purchase methods *(Dummy variable)*	0: At ticket offices 1: Banks 2: Both above
8). Easiness of renewing or purchasing travel cards (Dummy variable)	0: Very difficult 1: Difficult 2: Neutral 3: Easy 4: Very easy

Table 7.3 Variables and Codes for Smart Card Fare Payment

Variables	Explanation
1). Travel cost (Yuan)	Cost per month
2). Boarding time (seconds)	It is set the time difference between cash to show how much quicker by using smart card than cash
3). Ticket type (*Dummy variable*)	0: "Pay as you go" card 1: Electronic travel card (a minimum money of payment required per month) 2: Student smart card 3: Elder smart card
4). Overall assessment *(Dummy variable*)	0: Totally unsatisfied 1: Satisfied 2: Neutral 3: Satisfied 4: Totally satisfied
5). Whether more trips happened by using smart cards (*Dummy variable*)	0: No 1: Yes
6). Seat availability by using smart cards (compared with cash) (*Dummy variable*)	0: No difference with cash 1: Slightly better 2: Better 3: Much better
7). Top-up or purchase methods (*Dummy variable*)	0: At ticket offices 1: Banks 2: Agencies 3: Two of above three 4: All three methods
8). Easiness of topping up or purchasing smart cards (*Dummy variable*)	0: Very difficult 1: Difficult 2: Neutral 3: Easy 4: Very easy
9) Multifunction (*Dummy variable*)	0: Banking 1: Parking fee payment 2: Shopping 3: More than two functions 4: Only for public transport fare payment

In Table 7.1-7.3, it should be noted that:

- Generic variables and alternative specific variables: in the RP survey, travel cost, boarding time and overall assessment are generic variables across these three fare payment methods. For cost and time variables, the reason for keeping them as generic terms is that money and time do not expect to vary due to different payment options being used. More alternative specific variables are allocated to the smart card payment, because of more new features applying in this payment method than other two traditional methods.
- Boarding time savings: this variable is only applied in terms of time difference (in seconds), when comparing cash with cashless payment methods (travel cards or smart cards), therefore, in the data set, the boarding time by using cash is set to a constant

("zero"), while the rest two payment methods' boarding time are presented by how much quicker the boarding time can be by using travel cards/smart cards than using cash.

- Levels on those qualitative variables: those categories of qualitative questions come from the existing fare payment applications. Therefore, although the same variables are used in both RP and SP survey (for example, multifunction of smart cards), they are still viewed as alternative specific variables when combining the RP and SP data, due to the different variations of the variables in different surveys.

Alternative Availability for Individual Respondents

In the RP survey, except questions about respondents' actual fare payment used, questions about respondents' payment alternatives that could be used, were also asked. Under such circumstances, it is highly possible that not all data observations can involve trade-off situations (i.e., two or more than two options) and can be used at the choice modelling stage. The detailed individual responses towards these three payment methods in the RP survey have been discussed in Table 6.2 of Chapter 6. It can be seen from Table 6.2 that the 'single option' does not exist in RP-2 and RP-3 (questionnaires for travel card users and for smart card users, respectively), because everyone can pay by cash. However, for primary cash fare payment users, travel cards or smart cards may not be available for them for some reasons (such as having no idea about it, smart card reading devices is not available, etc.). So, 52 data observations with single payment option all come from cash fare questionnaires.

Because single options without any alternatives cannot tell any trading off information between two or three fare payment options. In order to run the model estimation by ALOGIT package under choice situations, we need to take into account the alternative availability for each individual: some have three options; some have two alternatives, and some only have single option without any alternatives. Table 6.2 in Chapter 6 lists and discusses the availability and unavailability of the three payment methods for the sample. As long as two payment options can be made, even if the third one is unavailable, the modelling analyse can be carried out to identify respondents' trade-off behaviour.

7.2.2. RP Choice Model and Parameter Estimation

As discussed in Chapter 4: Research Methodology, MNL models are used to model discrete choice data based on the RUT. The utility function for the RP and SP data can be written as Equation (4.1). Three utility functions are presented in the RP model, for cash, travel cards and smart cards, respectively. In this section, the analysis involves determination of variables in the utility functions, estimation results and comparisons of MNL and HL models for the RP data analysis.

Determination of the Variables in the Models

In order to achieve statistical significance of the estimations, before we present the final MNL models with the RP data, some dummy variables in some attributes need to be combined. The reason for doing this is that some dummies were not statistically significant at 95% level in the original estimation. The presence of such coefficient cannot significantly influence the individual relative utility of an alternative, because the number of responses to some categories in an attribute is very few. On the other hand, some dummies in one attribute are very similar, such as 'totally satisfied' and 'satisfied' of overall assessment, therefore, combination of categories (some of them with insignificant estimation previously) can make the estimated parameters interpretable at 95% level. So the actual number of categories of dummy variables in the models is less than the number of levels of the RP survey designed. The combined dummy variables are listed in detail (in the bracket the relevant code of dummy variables in Table 7.1-7.3 has been indicated):

1) **Overall assessment**: initial dummy variables: *totally unsatisfied* ('0') and *unsatisfied* ('1'), are combined; dummies: *satisfied* ('3') and *totally satisfied* ('4'), are combined; Choice Base: *Neutral* ('2').
2) **Smart card type**: dummy variables (code '0', '1' and '2' in Table 7.3) are combined together. Base: *'pay as you go' cards* ('3').
3) **Seat availability**: dummy variables (code '1', '2' and '3' in Table 7.2 and 7.3): *slightly better*, *better* and *much better*, are combined. Choice Base: *No difference compared with cash* ('0').
4) **Top-up/purchase methods for travel cards:** in table 7.2, dummy: *at ticket offices* ('0') and *banks* ('1') are combined. Base: *'Both above'* ('2').
5) **Top-up/purchase methods for smart cards**: in Table 7.3, dummy: *two of three methods* ('3') is combined with: *all three methods* ('4') as the choice base
6) **Easiness of topping up**: in Table 7.3 dummy variables ('0' and '1'): *Very difficult*, *Difficult*, are combined as a new dummy variable; dummies ('3' and '4'): *Easy* and *Very easy*, are combined. Base: *Neutral* ('2')
7) **Multifunction:** in Table 7.3 all four dummy variables related to extra services of the smart card are combined together as a single dummy variable (code '0', '1', '2' and '3'). Choice Base is *'only for public transport fare payment'* ('4')

Estimation Results of the RP Models

The model coefficients are estimated by ALOGIT software and listed in Table 7.4 and 7.5. Two kinds of logit models are used in analysing the RP data, a standard multinomial logit model (MNL) in Table 7.4 and hierarchical logit (HL) model in Table 7.5. In the HL model estimation, two hierarchical structures as designed in Figure 4.2 and 4.3 in Chapter 4, are discussed and compared with each other.

In general, most estimated coefficients in both models have statistical significance at 95% level except boarding time and ASC of smart cards. As can be seen in Table 7.4, the travel cost coefficient presents negative value with highly statistical significance. That is to say, with the increase of travel cost, the relative utility of choosing this fare payment method will be reduced. The boarding time saving coefficient presents a positive sign, because in the RP survey, how much quicker the boarding time savings by using travel cards or smart cards than cash was asked, the greater the boarding time difference is, the higher the relative utility can be. However, the estimated value is not statistically significant, though the sign is correct.

Table 7.4 RP Estimation Results of Standard MNL Model

Variables	MNL Estimation
	Estimated Parameters (T-ratios)
1. Travel cost (Yuan)	-0.1080 (-13.7)
2. Boarding time (seconds)	0.00678 (0.9)
3. Overall assessment: **Dummy 1:** Totally unsatisfied & Unsatisfied **Dummy 2:** Satisfied & Totally satisfied **Base:** Neutral	 -0.4945 (-2.5) 0.3054 (1.9)
4. Cash ticket type: **Dummy variable:** Zonal fare **Base:** Flat fare	 0.4535 (2.1)
5. Seat Availability by using travel cards or smart cards, comparing with cash **Dummy variable:** Slightly better or Better or Much better **Base:** No difference	0.6906 (4.0)
6. Top-up/purchase methods of travel cards **Dummy variable:** ticket offices **or** banks **Base:** Both ticket offices and banks	-0.4456 (-2.1)
7. Top-up/purchase methods of smart cards **Dummy 1:** At ticket offices **Dummy 2:** Banks **Dummy 3:** Agencies **Base:** two or three top up methods used	 -1.753 (-5.8) -1.393 (-5.2) -0.9082 (-2.8)
8. Easiness of topping up/purchasing **Dummy 1:** Very difficult & Difficult **Dummy 2:** Easy & Very easy **Base:** Neutral	 -0.8876 (-2.9) 0.5864 (2.5)
ASC-travel cards:	0.09875 (0.5)
ASC-Smart cards:	0.6388 (2.6)
Log likelihood at zero: **Log likelihood:** **No. of Observations:** **Rho-squared value w.r.t constants:**	-793.319 -482.646 782 0.3891

In Table 7.4, those estimations for dummy variables can give reasonable explanations (correct sign: better level than the base with positive sign and worse level than the base with negative sign), corresponding to the definition of these dummy variables. Overall, the positive sign of the estimations indicates that the increase of the relevant variables or presence of variables (dummies) can make an individual's relative utility increase and more probability to choose this alternative, *vice versa*.

Dummy 1 for *overall assessment* is negative with statistical significance, which means that compared with the base ('neutral') the presence of 'totally unsatisfied or unsatisfied' may result in the decrease of alternative utility. On the contrary, the positive value of Dummy 2 for *overall assessment* indicates alternative utility is increased when respondents chose 'Satisfied or Totally satisfied' categories, relative to the base. The similar result also can be found in *easiness of topping-up/purchasing* of travel cards/smart cards, in which compared with the base ('neutral') the Dummy 1 ('very difficult' or 'difficult') is negative and Dummy 2 ('easy' or 'very easy') is positive. The same reason as *overall assessment* can be applied in *difficulty of topping-up/purchasing* to explain the effect of estimation results on the smart card utility.

Estimated parameter of *ticket type of cash fare* indicates that zonal fare is preferred when respondents chose cash fare payment, relative to the base ('flat fare'). It is easily understood that because zonal fare is charged by the travel distance (or zone), for those passengers only travelling with short distance or within a zone, zonal fare can save their travel cost compared with flat fare, which is charged a given value regardless how long passengers travel.

Compared with 'No difference' in the attribute of *seat availability*, the positive estimation result for selecting 'better, or slightly better, or much better' shows that due to the quicker boarding process than cash fare, these respondents agreed that the seat availability by using travel cards or smart cards was better than cash, and regarded travel cards or smart cards could bring them higher utility than cash.

For Top-up/purchase methods for travel cards, the estimated coefficient for a single top-up/purchase is negative. It is common that respondents who can realise and use two top-up/purchase methods for travel cards have relatively higher utility of travel cards than those who only use one top-up/purchase method, because various options for topping-up/purchasing travel cards can bring convenience to users. Moreover, it can be seen that the top-up attribute for smart cards and travel cards were used as two different variables in Table 7.4, because different categories (2 levels for travel cards; 4 levels for smart cards) of this attribute used for these two card payment options. So by using alternative specific top-up attributes, different effects of top-up options to two card payments choices can be identified.

In addition to sign of estimated coefficient, which is used to explain the plausibility of estimation of dummies, size (or magnitude) of estimated value also needs to be examined. In the mean time, through comparing the size of coefficients of dummy variables, effects of attributes on actual payment choices can be identified. For example, three dummies of the attribute of *Easiness of topping-up/purchasing* for travel cards/smart cards are negative compared with the base (using two or all three top-up/purchase methods). The negative sign indicates that the base can bring more convenience to card users than the dummy variables with negative sign, therefore, the presence of the dummies reduce the relative utility of smart cards.

All in all, it is found that the besides cost and time variable, the following variables would have the biggest effect on fare payment demand, including top-up/purchase methods of smart cards; easiness of topping-up/purchasing of card payment; and seat availability by using travel cards/smart cards.

Another output in the RP model is the Rho-squared value (with respect to constants), which is used to measure the goodness of fit of the model estimation. Values of ρ^2 between 0.2 and 0.4 are considered to be indicative of extremely good model fits, which is equivalent to R^2 ranging 0.7-0.9 for a linear regression model (Ben-Akiva and Lerman, 1985; Louviere, Hensher and Swait, 2000). The rho-squared value is about 0.3891 w.r.t. constant, indicating a good model fit in the RP model.

Comparison of MNL and HL Models

Table 7.5 presents the estimation results for two different HL models as discussed in Chapter 4 (See Figure 4.2 and 4.3). In order to compare the goodness of fit of the MNL and HL models, as suggested by Louivere, Hensher and Swait (2000), cost coefficient, the generic variable in the MNL model, is split into three coefficients for three alternatives in the HL models. It can be seen that three cost coefficients are very close to each other with significant difference from zero at 95% level. Moreover, these three estimated cost coefficients are also very similar to the estimated result in Table 7.4 by the MNL model.

Comparing estimated parameters between two HL models, we can see that results from these two models do not vary too much, except ASCs for travel cards and smart cards. Such similarity between two models suggests that there is no big difference between HL-1 (See Figure 4.2) and HL-2 (See Figure 4.3).

By comparisons between the standard MNL model and two hierarchical logit models, most estimation results in the MNL model and hierarchical logit models are very close. The scale factor (θ) in two HL models are 0.8603 and 1.071, very close to 1.0 (if the scale factor is 1.0, the HL model can be regarded as a standard MNL model). Because the scale factors of two HL models are too close to 1.0 and all three models can achieve good model fits, in order to simplify the model structure, the MNL model is used for the RP data set in the further analysis (e.g., data combination with the SP data, forecasting analysis, *etc*).

As can be seen in Table 7.4 and 7.5, ρ^2 in the MNL and HL-1 models are in the range of 0.2-0.4 (0.3891 and 0.3952), satisfying the acceptable goodness of fit level suggested by empirical evidences. Although ρ^2 in the HL-2 model is 0.411, it is still not far from the acceptable range. Therefore, the HL-2 model also has a good model fit.

Table 7.5 RP Estimation Results of Two HL Models

Variables	HL-1 Estimation (Figure 4.2) Estimated Parameters (T-ratios)	HL-2 Estimation (Figure 4.3) Estimated Parameter (T-ratios)
1. Travel cost-Cash (Yuan)	-0.1136 (-13.9)	-0.1138 (-10.3)
2. Travel cost-Travel Cards (Yuan)	-0.1190 (-6.7)	-0.0978 (-9.1)
3. Travel cost-Smart Cards (Yuan)	-0.1227 (-6.9)	-0.1006 (-9.1)
4. Boarding time (seconds)	0.01183 (1.4)	0.009951 (1.4)
5. Overall assessment:		
Dummy 1: Totally unsatisfied & Unsatisfied	-0.4771 (-2.4)	-0.4859 (-2.5)
Dummy 2: Satisfied & Totally satisfied	0.3263 (2.0)	0.2792 (1.9)
Base: Neutral		
6. Cash ticket type:		
Dummy variable: Zonal fare	0.5842 (2.7)	0.4907 (2.2)
Base: Flat fare		
7. Seat Availability by using travel cards or smart cards, comparing with cash		
Dummy variable: Slightly better or Better or Much better	0.7605 (3.9)	0.7473 (4.3)
Base: No difference		
8. Top-up/purchase methods of travel cards		
Dummy variable: ticket offices **or** banks	-0.4554 (-2.0)	-0.4698 (-2.2)
Base: Both ticket offices and banks		
9. Top-up/purchase methods of smart cards		
Dummy 1: At ticket offices	-1.844 (-5.7)	-1.559 (-4.5)
Dummy 2: Banks	-1.478 (-5.2)	-1.299 (-4.5)
Dummy 3: Agencies	-0.9832 (-2.8)	-0.8162 (-2.5)
Base: two or three top up methods used		
10. Easiness of topping up/purchasing		
Dummy 1: Very difficult & Difficult	-0.9586 (-2.9)	-0.8609 (-2.5)
Dummy 2: Easy & Very easy	0.5975 (2.4)	0.6727 (2.9)
Base: Neutral		
ASC-travel cards:	-0.7875 (-2.0)	-0.4948 (-1.2)
ASC-smart cards:	-0.1618 (-0.4)	0.06786 (0.2)
Theta (θ):	0.8603 (19.6)	1.071 (8.1)
Log likelihood at zero:	-793.319	-793.319
Log likelihood:	-464.7715	-465.3221
No. of Observations:	782	782
Rho-squared value w.r.t. constants:	0.3952	0.4110

7.3. Modelling SP Data

In this section, modelling SP data is discussed from the following aspects: Input data preparation; Jack-knife analysis and discussion of estimation results. Two kinds of logit models are used in this section: firstly, standard logit models analyse four different SP data sets; secondly, a hierarchical logit model for combining SP data sets is used and discussed.

This hierarchical structure has been discussed in Chapter 4 (Please refer to Figure4.4).

7.3.1. SP Input Data Preparation and Coding Scheme

SP Input Data Preparation

In order to estimate the SP models in ALOGIT, SP data sets need to be prepared for the layout suitable for the software input requirement. In the SP survey design, four SP experiments were used: SP-1: cash vs. travel cards; SP-2: cash vs. smart cards ('pay as you go' cards); travel cards vs. smart cards ('pay as you go' cards) and SP-4: travel cards vs. smart cards ('pay monthly' cards). So, initially four separate SP data sets for different SP experiments are prepared. The basic data layout in the data sets is as follows:

- Firstly, data about fare payment methods are put in by grouping choice alternatives (variables related to a payment method is set together).
- Then discrete choice data are arranged by given codes (e.g., 0 and 1) to indicate different payment choices.
- Followed respondents' choices, a column of respondents' ID number is used to label different people and this ID number also is used in the Jack-knife analysis later on.

Coding for Dummy Variables

In the SP survey, except travel cost, boarding time, deposit (smart cards), which were designed on the basis of quantitative value, all other attributes were qualitative, which need to be defined and coded as dummy variables in the modelling analysis. The same rule as the RP data set is used to define and code dummy variables in the SP data sets. The detailed variable definition (particularly the choice base) and dummy variable coding scheme for each SP experiment is listed in Table 7.6-7.9. As can be seen in Table 7.6-7.9, bases and dummy variables have been indicated for each qualitative attribute. In the ALOGIT control file, all dummy variable parameters are estimated and measured (i.e., the sign and size of estimated coefficient) by the base predefined in the dummy variables.

Table 7.6 Variables and Codes for SP-1: Cash vs. Travel Cards

Variables	Explanation
1). Travel cost (Yuan) *(Generic Variable)*	Cost per month
2). Boarding time (seconds) *(Generic Variable)*	Compared cash with cashless fare payment. Presented by how much slower by using cash than travel cards/smart cards (time difference)
3). Whether passenger can get change if they pay cash *(Dummy variables-Cash)*	0: No 1: Yes
4). PT service covered by using travel cards *(Dummy variable-Travel cards)*	0: Limited routes: Only one bus, or light rail route service in urban area (service1); 1: Unlimited routes without any extra charges **(Base)** 2: Unlimited routes with extra charge: 10% more than limited services (Service2) 3: Unlimited routes with extra charge: 15% more than limited services (Service3)

Table 7.7 Variables and Codes for SP-2: Cash vs. Smart Cards (Pay as you go)

Variables	Explanation
1). Travel cost (Yuan) (*Generic Variable*)	Cost per month
2). Boarding time (seconds) (*Generic Variable*)	Compared cash with cashless fare payment. Presented by how much slower by using cash than travel cards/smart cards (time difference)
3). Whether passenger can get change if they pay cash (*Dummy variables-Cash*)	0: No 1: Yes
4). Deposit (Yuan) (*Smart cards*)	Described by money value for the deposit of initial purchase of smart cards
5) Overdraft (*Dummy variable-SC*)	Whether the last trip can be guaranteed in case the credit remaining in a card is not sufficient to pay a ticket 0-No (**Base**) 1-Yes
6) Multifunction (*Dummy variable-SC*)	0-- No. only for public transport (**Base**) 1-- Shopping, telephone, entertainment (mf1) 2-- Shopping, telephone, entertainment, parking and tolling (mf2) 3-- Shopping, telephone, entertainment, parking, tolling and banking (mf3)
7) Geographic areas covered (*Dummy variable-SC*)	0— Only urban area (**Base**) 1— Urban and rural areas (Geo1) 2—In addition to urban and rural areas, other nearby cities included (Geo2) 3— Within one province (Geo3)
8) Top-up methods (*Dummy variable-SC*)	0—Only at ticket offices (**Base**) 1—Ticket offices, banks, agencies (topup1) 2—Ticket offices, banks, agencies, self-adding value machine (topup2) 3—Ticket offices, banks, agencies, self-adding value machine, telephone and on-line payment (topup3)

Table 7.8 Variables and Codes for SP-3: Travel Cards vs. Smart Cards (Pay as you go)

Variables	Explanation
1). Travel cost (Yuan) (*Generic Variable*)	Cost per month
2). PT service covered by using travel cards (*Dummy variables-TC*)	0—Limited routes: Only one bus, or light rail route service in urban area; 1—Unlimited routes without any extra charges (**Base**) 2—Unlimited routes with extra charge: 10% more than limited services 3—Unlimited routes with extra charge: 15% more than limited services
3). Top-up/purchase methods (*Dummy variables-TC*)	0— Ticket offices (**Base**) 1— Ticket offices and agencies (topuptc1) 2— Ticket offices and agencies and banks (topuptc2)
4) Deposit (Yuan) (*Dummy variable-SC*)	Described by money value for the deposit of initial purchase of smart cards
5) Multifunction (*Dummy variable-SC*)	0— No. only for public transport (**Base**) 1— Shopping, telephone, entertainment (mf1) 2— Shopping, telephone, entertainment, parking and tolling (mf2) 3— Shopping, telephone, entertainment, parking, tolling and banking (mf3)
6) Overdraft (*Dummy variable-SC*)	Whether the last trip can be guaranteed in case the credit remaining in a card is not sufficient to pay a ticket 0—No (**Base**) 1—Yes
7) Top-up methods (*Dummy variable-SC*)	0—Only at ticket offices (**Base**) 1—Ticket offices, agencies and banks (topupsc1) 2—Ticket offices, agencies, banks, telephone and online top-up (topupsc2)

Table 7.9 Variables and Codes for SP-4: Travel Cards vs. Smart Cards (Pay monthly)

Variables	Explanation
1). Travel cost (Yuan) *(Generic Variable)*	Cost per month
2). PT service covered by using travel cards *(Dummy variables-TC)*	0—Limited routes: Only one bus, or light rail route service in urban area; 1—Unlimited routes without any extra charges (**Base**) 2—Unlimited routes with extra charge: 10% more than limited services 3—Unlimited routes with extra charge: 15% more than limited services
3) PT service covered by using smart cards *(Dummy variables-SC)*	0—Unlimited routes without any extra charges (**Base**) 1—Unlimited routes with extra charge: 10% more than limited services (servicesc1) 2—Unlimited routes with extra charge: 15% more than limited services (servicesc2) 3—Unlimited routes with extra charge: 20% more than limited services (servicesc3)
4) Deposit (Yuan) *(Dummy variable-SC)*	Described by money value for the deposit of initial purchase of smart cards
5) Multifunction *(Dummy variable-SC)*	0— No. only for public transport (**Base**) 1— Shopping, telephone, entertainment (mf1) 2— Shopping, telephone, entertainment, parking and tolling (mf2) 3— Shopping, telephone, entertainment, parking, tolling and banking (mf3)
6) Geographic areas covered *(Dummy variable-SC)*	0— Only urban area (**Base**) 1— Urban and rural areas (Geo1) 2— In addition to urban and rural areas, other nearby cities included (Geo2) 3— Within one province (Geo3)

Among quantitative variables (i.e., cost and time variables), it should be noted that in the SP survey design we set that travel cards and smart cards are prepaid and users can straight get on the bus after checked by PT drivers and the boarding time difference is presented by how much slower by using cash than by using travel cards/smart cards, therefore, the boarding time for card payment are set to constant ('zero'). In Table 7.7 (SP3) and 7.8 (SP4), considering the similarity of on-board check-in procedure of travel card and smart card payment, there is no boarding time variable for situations of travel card vs. smart cards.

In the SP survey, two cost units were used: Yuan per month and Yuan per trip due to different payment types (e.g., cash payment is based on Yuan per trip, while the travel card/smart card payment is prepaid per month). In order to make the estimation results comparable, Yuan per trip for the travel cost variable was finally converted to Yuan per month according to respondents' reported number of trips by using a given payment method in their revealed travel behaviour.

7.3.2. SP Choice Models and Parameter Estimation

Modelling the SP data is firstly conducted by four different SP data sets separately, in which binary choices are included and a standard binomial logit model is used. Then four

SP data sets are combined to yield a hierarchical logit model. Before presenting the estimation results, in order to investigate the effects of repeated measurement problems, Jack-knife analysis needs to be discussed firstly.

Jack-Knife Analysis

Jackknife technique is used in the SP model estimation, therefore before the discussion of estimation results, the Jackknife technique, including the principle and application, is necessarily discussed in detail.

Jackknife Technique

Methods to analyse SP data require the assumption that each observation is independent. However, this assumption is not strictly valid when several repeated choices are made by each respondent, because an important feature of SP data is that multiple observations are obtained from each respondent. That is one of limitations of SP methods: "Repeated Measurement Problem" (Ortuzar and Willumsen, 2001).

As one of effective means to eliminate this problem, re-sampling technique has been used in many SP analyses. Cirillo *et al.* (1996) applied Jackknife and Bootstrap re-sampling techniques to correct the repeated measurement problems. The results of applying the Jackknife method confirmed that the estimated coefficient values remained unbiased, but the bias in the variance estimates were varied. They concluded that the repeated measurement problems were not serious in terms of size of the coefficients, and recommended Jackknife for practical work because it is easy to implement and produce smoother estimates at low re-sampling rates.

In the model estimation of this research, Jackknife technique is employed. The idea of Jackknife technique is to re-use the sample several times by dividing it into sub-samples and by recombining them to assemble an estimate of the unknown parameter which has good sampling properties and perhaps more importantly, to produce an estimate of the variance of this statistic.

The following formula is used to combine sub-sample estimates to get the Jackknife estimates:

$$\theta_{Jack} = r\theta_0 - (r-1)\bar{\theta} = \theta_0 + (r-1)(\theta_0 - \bar{\theta}) \qquad (7.1)$$

where $$\bar{\theta} = \frac{1}{r}\sum_{j=1}^{r}\theta_j \qquad (7.2)$$

θ_{Jack} : the final Jackknife estimate

θ_0 : the uncorrected estimate

$\bar{\theta}$: the mean of partial Jackknife estimates

θ_j : the jth partial Jackknife estimate

r: the number of sub-samples

The Jackknife variance estimation (σ^2_{Jack}) is:

$$\sigma^2_{Jack}(\theta) = \frac{n-1}{n}\sum_{j=1}^{n}(\theta_j - \theta)^2 \tag{7.3}$$

n: the sample size

Application of Jackknife

The Jackknife method has been implemented using a programme "JACKKNIFE" in ALOGIT. The programme, "JACKKNIFE", allows the choice of the number of sub-samples and modifies the control file of the estimation programme to skip certain observations. Then "ALOGIT" programme is used to estimate sub-models based on the each sub-sample. Finally, "JACKKNIFE" combines all the sub-models to produce final Jackknife estimates.

The number of sub-samples is important in Jackknife implementation because it improves the power of significance test and makes variance standard stable. The ideal number of sub-samples is the number of samples (i.e. $r = n$, n is the sample size). It was recommended to make the number of sub-samples, r as large as possible by Bissell and Ferguson (1975). It, however, was also suggested by Cirillo *et al.* (1996) for users to try different numbers of sub-samples and choose the lowest values of r where the estimates stabilise for the efficiency of the model estimates. In ALOGIT, the programme "JACKKNIFE" allows the number of sub-samples only between 2 and 99. In this analysis, first of all, different numbers of sub-samples were tested to determine where coefficient estimates and variances settle down. The sub-samples of '5', '10', '20', '30' '40' and '50' were tried to estimate four SP models. Finally, it can be found that coefficient estimates and variances in most SP models settled down around 20-30 sub-samples, which means that the effects of repeated measurement problem became stable when using sub-samples of 20-30. Therefore, in this study, 20 sub-samples are studied with the repeated measurements (i.e., the respondent's ID number that we used for each returned questionnaire paper).20 sub-samples also were suggested and commonly used in empirical studies (HCG, 2001; Cho and Kim, 2002).

Final Estimation Results by Jack-knife Method

Finally, estimated results of the four standard binary logit models are presented in Table 7.10.-7.13. Meanwhile, because a Jackknife (20 sub-samples) analysis is carried out to overcome "repeated measurement problem" in the SP estimation, all estimation results presented in four tables have already been jackknifed. The effect of the Jackknife analysis can be discussed from two aspects: coefficient estimates and t-ratio. Most Jack-knife estimates are slightly smaller than those of the standard logit models. It can be implied that logit models overestimated due to the repeated measurement problem. T-ratio in jack-knife is used to identify whether the significance of coefficients is overestimated or underestimated. When checking T-ratios between the standard logit models and jack-knifed models, we can see that some t-ratio values in the jack-knife analysis are lower than the standard logit models, indicating that the logit models overestimated the significance of the parameter.

In Table 7.10 and 7.12, it is worth noting that ASCs have not been included in the utility models for SP-1 and SP-3. It is found that alternative specific constant: ASCs in SP-1 (for travel cards alternative) and SP-3 (smart cards alternative) are highly correlated with cost coefficients (about 0.8-0.9). As the definition of ASC states, it represents the mean of the distribution of the unobserved effects in the random component ε, associated with alternative i (Louviere, Hensher and Swait, 2000). Such high correlation between the cost coefficient (deterministic attribute) and ASCs would result in a significant difference between cost coefficient estimations in SP-1 and SP-3. For example, before excluding ASCs in utility functions of SP 1 and 3, the cost coefficient estimates were -0.2102 and -0.1087 with statistical significance. Moreover, such difference between the common variable, though it was split by different cost units to examine the homogeneity of the travel cost variable, could make the combination of data of different SP experiments difficult and impossible. Therefore, in order to achieve the consistency of common attributes on estimation, ASCs in SP-1 and SP-3 are excluded in the utility models. However, in SP-2 and SP-4, the correlation between ASCs and cost coefficient is not very high (0.3 for SP 2 and 0.065 for SP 4), so ASCs in these two models are still included.

The travel cost is the common attribute in the SP survey. Regarding estimated cost coefficients for the separate models (SP-1, SP-2, SP-3 and SP-4), it can be seen that the estimation results are very close to each other, although in SP-1 and SP-3, split cost parameters are used for different travel cost units by using different payment methods (paid per ride and paid per month). Such similarity of common coefficient allows for the combination of different SP data sets together in the later stage.

In the SP data sets, most variables are coded as dummy variables, such as PT service route covered, multifunction, whether passengers can get change back if they cannot pay exact cash/coin, etc. Compared with the bases defined for these dummy variables, it is necessary to examine whether estimation results of dummy variables can be meaningful on sign and magnitude of estimated value, relative to other dummies in one attribute if the number of dummies of an attribute is more than two. First of all, it can be seen that all estimations of dummy variables in Table 7.10-7.13 have presented correct sign compared with the base. When comparing between dummies within one attribute, they also show the normal perceptions of respondents to those better service quality of fare payment. Meanwhile, it should be noted that very few estimates are statistically insignificant in these four models, therefore all estimated results are included in Table 7.10-7.13. The detailed discussion about the estimation results are as follows:

- Travel cost: the sign is negative in four models. It is very common that the increase of travel cost of payment alternative can result in the decrease of relative utility, and *vice versa*.
- Boarding time difference: because the boarding time is presented by how much slower by using cash than card payment, it is also not surprising that estimated coefficient present negative sign to indicate that the greater of the boarding time difference, the lower of the utility of cash payment.
- Whether getting change: this dummy variable is positive compared with the base (No, passenger cannot get change back). The positive sign shows that if passengers can get change back when they are unable to pay exact money, the relative utility may be increased, because it can bring convenience to cash users.
- PT services covered: because we set 'unlimited route can be used without any extra charges' as the base, it is reasonable that three dummies (i.e., limited use or unlimited use but extra charges required) are negative. The presence of any dummy will reduce the utility of card payment compared to the base. From these three dummies, the relativities between dummies also can be obtained: with the service of a dummy becoming worse, the estimate also relatively becomes more negative, which means the presence of this dummy can make the utility lower than those dummies with better services.
- Deposit: the negative estimation of the deposit variable also tells us that respondents do not like deposit at the initial purchase of the smart cards. It is easily understood that the deposit indirectly increases the travel cost, therefore the presence of this attribute can make the smart card utility decrease.
- Overdraft: the implementation of overdraft policy for smart cards can increase the convenience for card users, particularly when the credit remaining in a card is not sufficient to pay a ticket during the journey, therefore, when smart cards with overdraft function, respondents would like to use the smart card ticketing, compared with the base (smart cards without overdraft function)
- Multifunction: all three dummies (with some extra functions) are positive relative to the base ('only for PT fare payment, without any extra functions'). Moreover, with the functions becoming more, the size of the estimates of dummies also becomes greater than the previous level (i.e., Dummy 3>Dummy 2>Dummy 1). This can reflect that respondents more like using smart cards with multifunction.
- The similar explanation with multifunction can be applied in variables of '*Geographic*

area covered' and *'Top-up methods'*. The lowest (or the worst) category for these two variables is set as the base. The presence of dummies can increase the relative utility therefore the estimated coefficients are positive. The size of the estimation values also becomes greater when the dummies become better. In a word, respondents more would like to use smart cards with wider geographic areas that can be covered (or more top-up options can be available for smart cards).

Beyond the discussion of the effects on demand within an attribute, it can be found that by comparing across different payment attributes, the following payment attributes may have the biggest effects of payment variables on payment demand:

- In SP 1 model, PT service covered by travel cards has the biggest effect among variables except cost and time, because compared with 'whether passenger can get change back', on average three dummies in 'PT service covered by travel cards' would have more influences on individual utility of choosing travel cards, thus the relevant demand of travel cards also can be influenced more obviously.
- In SP 2 model, two smart card related variables: 'geographic areas covered' and 'top-up methods', have the biggest effects on the smart card payment demand. On the contrary, effects of 'multifunction', 'overdraft' and 'whether passengers can get change back' on individual utilities are not as high as the former two factors.
- In SP 3 model, 'top-up methods' and 'multifunction' are top two variables, having the biggest effects on the demand, compared with other variables in the model.
- In SP 4 model, 'multifunction' and 'geographic areas covered by smart cards' seem to have the biggest effects on trading off between travel cards and smart cards. From SP 3 and SP 4 model, it can be seen that respondents more focus on smart card related attributes to trade off between travel cards and smart cards as well as travel cost, service routes covered, etc.

The Rho-squared values in Table 7.10-7.13 are used to measure the goodness of fit of the logit models in the SP data analysis. As discussed in Section 7.2.2, the Rho-squared value with respect to constants between 0.2-0.4 indicates the extremely good model fits. We observe that two models with ASC (SP-2 and SP-4) can achieve relatively good model fits (0.2281 in SP-2 and 0.4072 in SP-4). Although the Rho-squared values in SP-1 and SP-3 are little far from the range of [0.2 0.4], the final step is to combine all SP data sets together, therefore we need to examine the goodness of fit of the model overall when jointly analysing the four SP models.

Table 7.10 Estimation Results of SP 1: Cash vs. Travel Cards

Variables		Description	Estimated Parameters	T-ratios
Travel Cost-1:	Alternative Specific Variable: Cash	**Yuan per ride** (converted to yuan per month)	-0.1308	-12.21
Travel Cost-2:	Alternative Specific Variable: Travel Cards	**Yuan per month**	-0.1288	-8.98
Boarding time:	Generic Variable	**Seconds**	-0.0108	-1.98
Change (dummy variable):	Alternative Specific Variable: Cash	*Passengers cannot get changes (**Base**)*	0.5325	2.51
PT Service Covered:	Alternative Specific Variable: Travel Cards			
Dummy 1:		**Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.0743	-0.35
Dummy 2:		**Unlimited routes with extra charge: 10% more than limited services*	-0.4798	-1.55
Dummy 3:		**Unlimited routes with extra charge: 15% more than limited services*	-0.8335	-2.34
		Choice Base: *Unlimited routes without any extra charges*		
Likelihood:	-221.2799			
Rho-Squared w.r.t. cons:	0.5521			
Rho-Squared w.r.t. zero:	0.5912			
No. of Observations:	703			

Table 7.11 Estimation Results of SP 2: Cash vs. Smart Cards (Pay as you go)

Variables		Description	Parameters	T-Ratios
Travel Cost:	Generic Variable	**Yuan per month**	-0.1227	-6.61
Boarding time:	Generic Variable	**Seconds**	-0.0152	-6.09
Changes (dummy variable):	Alternative Specific Variable: Cash	*Passengers cannot get changes (**Base**)*	0.1072	1.25
Deposit:	Alternative Specific Variable: Smart cards	**Yuan**	-0.0238	-8.40
Overdraft (dummy variable):	Alternative Specific Variable: Smart cards	*Smart cards have overdraft function (**Base**)*	0.6790	4.71
Multifunction	Alternative Specific Variable: Smart cards			
Dummy 1:		**Shopping, telephone, entertainment*	0.3647	2.45
Dummy 2:		**Shopping ,telephone ,entertainment, parking and tolling*	0.8510	2.79
Dummy 3:		**Shopping ,telephone ,entertainment, parking, tolling and banking*	1.5347	4.37
		Choice Base: None. only for public transport		
Geographic areas covered:	Alternative Specific Variable: Smart cards			
Dummy 1:		**Dalian Urban and rural areas*	0.7568	4.23
Dummy 2:		**Dalian plus other nearby cities*	1.0254	4.42
Dummy 3:		**Within one province*	1.2472	4.93
		Choice Base: Only at Dalian urban area		
Top-up methods	Alternative Specific Variable: Smart cards			
Dummy 1:		**Ticket offices, banks , agencies*	0.6634	2.92
Dummy 2:		**Ticket offices,banks,agencies,self-adding value machine*	0.7179	4.31
Dummy 3:		**Ticket offices,banks,agencies, self-adding value machine, telephone and Internet*	1.9249	3.94
		Choice Base: Only at ticket offices		
ASC-smart cards:	Smart cards		-0.7560	-4.41
Likelihood:	-1065.2622			
Rho-squared w.r.t. cons.:	0.2281			
Rho-squared w.r.t. zero:	0.3224			
No. of observations:	2041			

Table 7.12 Estimation Results of SP 3: Travel Cards vs. Smart Cards (Pay as you go)

Variables		Description	Parameters	T-Ratios
Travel Cost-1- TC:	Alternative Specific	**Yuan per month**	-0.1767	-11.73
Travel Cost-2- SC:	Alternative Specific	**Yuan per ride** (converted to Yuan per month)	-0.1761	-11.09
Deposit:	Alternative Specific Variable: Smart cards	**Yuan**	-0.0454	-7.07
Overdraft (dummy variable):	Alternative Specific Variable: Smart cards	*Smart cards have overdraft function* ***(Base)***	0.3103	1.58
Multifunction	Alternative Specific Variable: Smart cards			
Dummy 1:		**Shopping, telephone, entertainment*	0.3462	0.78
Dummy 2:		**Shopping ,telephone ,entertainment, parking and tolling*	1.1003	3.16
Dummy 3:		**Shopping ,telephone ,entertainment, parking, tolling and banking*	1.1365	3.74
		Choice Base: None. only for public transport		
PT services covered:	Alternative Specific Variable: Travel cards			
Dummy 1:		**Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.7796	-2.70
Dummy 2:		**Unlimited routes with extra charge: 10% more than limited services*	-0.8159	-2.88
Dummy 3:		**Unlimited routes with extra charge: 15% more than limited services*	-0.8611	-4.06
		Choice Base: Unlimited routes without any extra charges		
Top-up methods-TC	Alternative Specific Variable: Travel cards			
Dummy 1:		**Ticket offices and agencies*	0.0626	0.23
Dummy 2:		**Ticket offices and agencies and banks*	1.1693	3.09
		Choice Base: Only at Ticket offices		
Top-up methods-SC	Alternative Specific Variable: Smart cards			
Dummy 1:		**Ticket offices, agencies and banks*	0.0380	0.12
Dummy 2:		**Ticket offices, agencies, banks ,automatic adding value machine, telephone and online*	1.5077	4.03
		Choice Base: *Ticket offices and agencies*		
Likelihood:	-404.5521			
Rho-squared w.r.t. cons.:	0.6562			
Rho-squared w.r.t. zero:	0.6565			
No. of observations:	1529			

Table 7.13 Estimation Results of SP 4: Travel Cards vs. Smart Cards (Pay monthly)

Variables		Description	Parameters	T-Ratios
Travel Cost:	Generic Variable	**Yuan per month**	-0.2116	-13.29
Deposit:	Alternative Specific Variable: Smart cards	**Yuan**	-0.0125	-2.29
PT services covered-TC:	Alternative Specific Variable: Travel cards			
Dummy 1:		**Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.8788	-2.60
Dummy 2:		**Unlimited routes with extra charge: 10% more than limited services*	-0.9836	-3.21
Dummy 3:		**Unlimited routes with extra charge: 15% more than limited services*	-1.0764	-4.89
		Choice Base: Unlimited routes without any extra charges		
PT services covered-SC:	Alternative Specific Variable: Smart cards			
Dummy 1:		** Unlimited routes with extra charge: 10% more than limited services.*	-0.1634	-0.52
Dummy 2:		**Unlimited routes with extra charge: 15% more than limited services*	-0.2677	-0.76
Dummy 3:		**Unlimited routes with extra charge: 20% more than limited services*	-0.7101	-2.66
		Choice Base: Unlimited routes without any extra charges		
Multifunction	Alternative Specific Variable: Smart cards			
Dummy 1:		**Shopping, telephone, entertainment*	0.4193	1.11
Dummy 2:		**Shopping ,telephone ,entertainment, parking and tolling*	1.0289	3.06
Dummy 3:		**Shopping ,telephone ,entertainment, parking, tolling and banking*	1.4212	3.79
		Choice Base: None. only for public transport		
Geographic areas covered	Alternative Specific Variable: Smart cards			
Dummy 1:		**Dalian Urban and rural areas*	0.5178	1.38
Dummy 2:		**Dalian plus other nearby cities*	0.8245	2.38
Dummy 3:		**Within one province*	1.4064	3.65
		Choice Base: Only at Dalian urban area		
ASC-smart cards	Smart cards		-0.4747	-1.0
Likelihood:	-584.2743			
Rho-squared w.r.t. cons.:	0.4072			
Rho-squared w.r.t. zero:	0.4246			
No. of observations:	1320			

7.3.3 Combining SP Data Sets

In Section 7.3.2, the separate estimation for the four SP data sets have been discussed. The following task is to combine these four SP data sets by using a hierarchical structure. The combination of the SP data is the preparation for the joint analysis of the RP and SP data, because eventually the RP and SP data need to be pooled together as a whole for ALOGIT estimation. When checking estimation results of common variables in these four SP experiments, we can see that the similarity of the cost and boarding time coefficients (common attributes) in four SP experiments (See Table 7.10-7.13) allows for a hierarchical structure with different scales to combine these four SP data sets. The detailed hierarchical structure has been illustrated in Figure 4.4, Chapter 4.

Hierarchical Structure

The hierarchical logit model for the SP data combination is presented by eight utility terms, which come from previous four independent SP experiments. Figure 4.4 in Chapter 4 shows that two alternatives in SP 1 are arbitrarily set in the upper level, and the rest three SP games with two alternatives in each game are put on the lower level. Three scale factors are used for SP 2, 3 and 4. It should be noted that the assumed hierarchical structure may vary by comparing the scale estimates with 1.0. The scale of 1.0 is regarded as the direct link of alternatives with the root in the hierarchical structure. In this case, if some scale is close to 1.0, then we can assume that the relevant alternatives in the lower level can be upgraded to the upper level, where alternatives in SP 1 are allocated, so as to pool different SP models together.

In addition, because when designing the SP survey, two kinds of cost units are used: Yuan per ride and Yuan per month (e.g., cash payment is paid per trip, while travel card payment is paid per month), in order to examine the effect of different cost units on the model estimation, two situations are considered: hierarchical logit models with single cost coefficient and with two cost coefficients (yuan per ride and yuan per trip). The reason for doing this is that we assume passengers' attitude to Yuan per ride and Yuan per month could be different, although for the travel cost with Yuan per ride unit, we have converted it to Yuan per month so that the cost parameters estimated can be comparable. Splitting cost coefficients would help examine whether the payment period (paid on the basis of per trip and per month in advance) could influence respondents' choice utilities.

Estimation Results

In total 5593 data observations (90% of total data, the rest 10% is used to test the model validation in the later stage) are modelled in the hierarchical logit models. Before the final estimation results are presented, it is necessary to decide whether the single cost coefficient or two cost coefficients (for two units: Yuan per month and Yuan per trip) should be finally used in the hierarchical models. The estimation results of cost coefficients in two different situations are -0.1297 for the single cost coefficient model; -0.1282 (Yuan per month) and -0.1901 (Yuan per trip) for the two cost coefficients model. All estimates are statistically significant at 95 per cent level. But in order to simplify the model combination process and measure choice behaviours later on, the single cost coefficient is kept in the SP model, because the difference between two cost coefficients is not very much. Therefore, the final estimation results are based on the single cost coefficient (Yuan per month) and listed in Table 7.15.

All estimation results in Table 7.14 have correct sign and most coefficients are statistically significant at 95 per cent level. It can be seen that the travel cost, boarding time and deposit estimates present a correct sign (negative) as discussed in Section 7.3.2. For dummy variables, when checking the base we defined before the model estimation, the sign of estimated coefficients and relative magnitude between dummy variables for one attribute are correct as expected. In general, as discussed for the SP separate models in the last section, if dummies are better than the base, then estimates present positive sign, because respondents perceived fare payment methods with better services and the relative utility may be increased, and *vice versa*. Secondly, the size of estimated values can tell the information about the relativities between two dummies, among positive dummies, the greater the estimate is, the better the service level can present. Thus it is preferred by respondents. On the contrary, the more negative the dummy is, the worse the dummy represents and the relative utility may become lower. For example, three dummy variables for 'multifunction' are positive, which means that these three dummies are better than the base ('no multifunction, only for PT fare payment'). And between dummy variables, estimated values with more multifunction of smart cards are greater than those with less function (Dummy3>Dummy2>Dummy1), indicating that most respondents rationally perceived better services of fare payment. The similar explanation can be applied in '*geographic areas covered*', '*top-up options*', etc. Another example is that dummies for '*PT service route covered*' in travel cards and smart cards are negative, indicating respondents more like the base ('unlimited routes applied') rather than those dummies with 'limited routes' or 'unlimited routes but extras charges needed'. And with the increase of extra charge for service routes, the estimate of the dummy becomes more negative. It is common that respondents would not like to use the fare payment method with higher travel cost and then the relative utility of this payment mean is lower than other payment options.

Table 7.14 Estimation Results of Combined Four SP Data Sets with Three Scale Factors

Variables		Description	Parameters	T-Ratios
Travel Cost-1:	Alternative Specific Variable: pay monthly	**Yuan per month**	-0.1282	-11.48
Travel Cost-2:	Alternative Specific Variable: pay per ride	**Yuan per trip** (converted to Yuan per month)	-0.1901	-12.24
Boarding time:	Used when comparing cash with other two payments	**Seconds**	-0.0126	-3.88
Change (dummy variable):	Alternative Specific Variable: Cash	*Passengers cannot get changes **(Base)***	0.2168	2.54
Deposit:	Alternative Specific Variable: Smart cards	**Yuan**	-0.0184	-7.04
Overdraft (dummy variable):	Alternative Specific Variable: Smart cards	*Smart cards have overdraft function **(Base)***	0.7141	6.01
PT services covered-TC:	Alternative Specific Variable: Travel cards			
Dummy 1:		**Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.9147	-6.10
Dummy 2:		**Unlimited routes with extra charge: 10% more than limited services*	-0.8917	-6.93
Dummy 3:		**Unlimited routes with extra charge: 15% more than limited services*	-1.1502	-7.80
		Choice Base: Unlimited routes without any extra charges		
PT services covered-SC:	Alternative Specific Variable: Smart cards			
Dummy 1:		** Unlimited routes with extra charge: 10% more than limited services.*	-0.1104	-0.54
Dummy 2:		**Unlimited routes with extra charge: 15% more than limited services*	-0.4458	-2.04
Dummy 3:		**Unlimited routes with extra charge: 20% more than limited services*	-0.7353	-4.41
		Choice Base: Unlimited routes without any extra charges		

Table 7.14 (Continued)

Multifunction:	Alternative Specific Variable: Smart cards			
Dummy 1:		**Shopping, telephone, entertainment*	0.3300	2.61
Dummy 2:		**Shopping ,telephone ,entertainment, parking and tolling*	0.9935	5.83
Dummy 3:		**Shopping ,telephone ,entertainment, parking, tolling and banking*	1.3852	7.22
		Choice Base: None. only for public transport		
Geographic areas covered	Alternative Specific Variable: Smart cards			
Dummy 1:		**Dalian Urban and rural areas*	0.5314	3.81
Dummy 2:		**Dalian plus other nearby cities*	0.9363	3.93
Dummy 3:		**Within one province*	1.3852	7.22
		Choice Base: Only at Dalian urban area		
Top-up methods-TC	Alternative Specific Variable: Travel cards			
Dummy 1:		**Ticket offices and agencies*	0.0368	0.19
Dummy 2:		**Ticket offices and agencies and banks*	1.0676	4.09
		Choice Base: Only at Ticket offices		
Top-up methods-SC	Alternative Specific Variable: Smart cards			
Dummy 1:		**Ticket offices, banks , agencies*	0.4455	3.17
Dummy 2:		**Ticket offices ,banks, agencies, self-adding value machine*	0.9907	6.79
Dummy 3:		**Ticket offices, banks, agencies, self-adding value machine, telephone and Internet*	1.9010	3.39
		Choice Base: Only at ticket offices		
ASC-smart cards in SP2			-0.7226	-4.62
ASC-smart cards in SP4			-0.0809	-0.41
θ_1		Scale factor for SP2	0.9813	6.65
θ_2		Scale factor for SP3	1.4673	10.11
θ_3		Scale factor for SP4	1.0960	8.81
Likelihood:	-2246.5753			
Rho-squared w.r.t. cons.:	0.4434			
Rho-squared w.r.t. zero:	0.4671			
No. of observations:	5593			

Table 7.15 Estimation Results of Combined Four SP Data Sets with Scale Factor for SP 3

Variables		Description	Parameters	T-Ratios
Travel Cost:	Alternative Specific Variable: pay monthly	**Yuan per month**	-0.1297	-15.89
Boarding time:	Used when comparing cash with other two payments	**Seconds**	-0.0131	-4.66
Change (dummy variable):	Alternative Specific Variable: Cash	*Passengers cannot get changes (**Base**)*	0.2205	2.79
Deposit:	Alternative Specific Variable: Smart cards	**Yuan**	-0.0193	-7.36
Overdraft (dummy variable):	Alternative Specific Variable: Smart cards	*Smart cards have overdraft function (**Base**)*	0.7388	6.50
PT services covered-TC:	Alternative Specific Variable: Travel cards			
Dummy 1:		**Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.9763	-7.20
Dummy 2:		**Unlimited routes with extra charge: 10% more than limited services*	-0.9588	-8.76
Dummy 3:		**Unlimited routes with extra charge: 15% more than limited services*	-1.2231	-9.04
		Choice Base: Unlimited routes without any extra charges		
PT services covered-SC:	Alternative Specific Variable: Smart cards			
Dummy 1:		** Unlimited routes with extra charge: 10% more than limited services.*	-0.1655	-0.82
Dummy 2:		**Unlimited routes with extra charge: 15% more than limited services*	-0.4778	-1.91
Dummy 3:		**Unlimited routes with extra charge: 20% more than limited services*	-0.8133	-4.01
		Choice Base: Unlimited routes without any extra charges		

Table 7.15 (Continued)

Multifunction:	Alternative Specific Variable: Smart cards			
Dummy 1:		**Shopping, telephone, entertainment*	0.3379	2.68
Dummy 2:		**Shopping ,telephone ,entertainment, parking and tolling*	1.0209	5.60
Dummy 3:		**Shopping ,telephone ,entertainment, parking, tolling and banking*	1.4653	10.08
		Choice Base: None. only for public transport		
Geographic areas covered	Alternative Specific Variable: Smart cards			
Dummy 1:		**Dalian Urban and rural areas*	0.5541	4.75
Dummy 2:		**Dalian plus other nearby cities*	0.9747	4.42
Dummy 3:		**Within one province*	1.2883	10.10
		Choice Base: Only at Dalian urban area		
Top-up methods-TC	Alternative Specific Variable: Travel cards			
Dummy 1:		**Ticket offices and agencies*	0.0280	0.14
Dummy 2:		**Ticket offices and agencies and banks*	1.1294	4.35
		Choice Base: Only at Ticket offices		
Top-up methods-SC	Alternative Specific Variable: Smart cards			
Dummy 1:		**Ticket offices, banks , agencies*	0.4585	2.90
Dummy 2:		**Ticket offices ,banks, agencies, self-adding value machine*	1.0131	7.30
Dummy 3:		**Ticket offices, banks, agencies, self-adding value machine, telephone and Internet*	1.9310	4.03
		Choice Base: Only at ticket offices		
ASC-smart cards in SP2			-0.7619	-5.38
ASC-smart cards in SP4			-0.0787	-0.38
θ_4		Scale factor for SP3	1.4122	11.14
Likelihood:	-2247.6853			
Rho-squared w.r.t. cons.:	0.4431			
Rho-squared w.r.t. zero:	0.4598			
No. of observations:	5593			

When comparing estimated results across different variables, the sensibility of variables to the demand can be found:

- Top-up methods of smart cards would have the biggest effects on choosing smart cards, because individual utility could vary more sensibly than any other variables due to the variation of top-up methods of smart cards. In particular, from dummy 2 to dummy 3, the relative utility would change very significantly (1.0131 to 1.931) when fixing other variables and only considering the single factor (please see Table 7.15). Regarding how influence of payment variables could be on the demand, Section 7.6.1 discusses in detail.
- PT services covered by travel cards is the second variables with the relatively remarkable effects on the payment demand, particularly when the attribute changes from the base (unlimited routes without extra cost) to other categories (limited services or unlimited services with extra cost).
- Compared with these two sensible variables, PT services covered by smart cards might have the least influence on trading off between smart cards and other payment options. The relative utility based on the change of levels of this attribute is not as significant as PT services covered by TC and top-up methods of SC.

In Table 7.14, three scale factors (θ_1, θ_2 and θ_3) are used for scaling SP 2, 3 and 4 in the lower level of the hierarchical structure. Three scale factors are 0.9813, 1.4673 and 1.0960 respectively for SP-2, 3 and 4. From the estimated scales, we can deem that SP-1 and SP-2 and SP-4 should be put on the same level (the upper level) of the hierarchical structure, because θ_1 for SP-2 and θ_3 for SP-4 are very close to 1.0 like SP-1 that is set on the upper level. SP-3 is still set on the lower level. Therefore, the new hierarchical structure can be designed as: SP-1, 2 and 4 on the upper level and SP-3 on the lower level with a scale factor. Estimation results for the new developed model are listed in Table 7.15. All discussions in the later stage also are based on the estimation results in Table 7.15. From Table 7.15, we can observe that coefficient sign, plausibility of absolute value and relativities between levels of dummy variables also present good explanation as in Table 7.14. After combining scales, most estimates' absolute values are slightly greater than results in Table 7.14. This is due to the presence of scale factors for SP 2 and 4 in Table 7.14, but in Table 7.15 these two SP games have been reset to the upper level with SP 1. The scale factor of SP 3 in Table 7.15 is very close to the scale factor in Table 7.14.

The goodness of fit in the combined model is measured by the Rho-squared value with respect to constants (See Table 7.14 and 7.15). In these two models, the Rho-squared value is 0.4434 in the model with three scales and 0.4431 in the model with one scale only for SP 3. Both values are not far from the range between 0.2 and 0.4. Therefore, we regard the goodness of fit of the combined model acceptable.

7.4. Data Enrichment Analysis with the RP and SP Data

In Section 7.2 and 7.3 the separated models for the RP and SP data have been discussed and estimation results have been presented with correct sign and most of them have statistical significance. In this section, the data enrichment is conducted for these two different data sources. Through combining the RP and SP data, we can make full use of advantages of these two sorts of data in modelling choice behaviour and forecasting user demand. By introducing 'scale factor', how the RP and SP data are reliable can be examined and 'true' utilities can be estimated. In this section, two estimation approaches (i.e., sequential estimation and simultaneous estimation) available for the joint RP and SP analysis are discussed and compared with each other to determine which model is suitable for the further analysis.

7.4.1. Sequential Estimation of the RP and SP Data

The sequential estimation approach is firstly employed in this section. The principal advantage of the sequential estimation is because the sequential estimation is conducted separately (e.g., SP estimation first, then RP estimation), two data sources are not required to be combined together. This estimation approach will keep the same relative valuation of attributes in SP models, just changing the scale. So this estimation approach may be suitable if the relative values from the SP models are satisfactory before combining. The estimation significance will remain the same level as the single data source model before the combination. In the sequential estimation, the first task is to determine common attributes between the RP and SP models so that these common attribute estimations in the SP model can be entered in the RP utility model to calculate the scale factor (more details in Section 4.4.3, Chapter 4). In this research, the common attributes across these two data sources are "*Travel Cost*" and "*Boarding Time Savings*".

First of all, we use estimated coefficients of these two common attributes in the SP models to replace the cost coefficient and boarding time coefficient in the RP model to generate new utility (NU) for three RP alternatives as follows:

$$NU = \alpha_{sp}\ Cost_{rp} + \beta_{sp}\ BoardingTime_{rp} \tag{7.4}$$

Then introduce NUs to the RP utility model as equation (4.17) in Chapter 4, in which the RP utility function is composed by the NU and the RP specific variables. Finally, the sequential estimation is carried out by ALOGIT software to calculate the scale factor. In this model, the scale factor is 0.5608 with statistical significance at 95% level in this case.

It should be noted that in the sequential estimation, the SP data is firstly estimated as Table 7.15 and then by introducing the scale factor to model the RP data, therefore, results presented in Table 7.16 contain estimates of RP specific variables and the scale factor. Although in both the RP and SP survey, 'top-up methods' was used, due to different levels designed in the SP survey this attribute is still regarded as specific variables in the RP and SP data, thus 'top-up' variable for two data sets is separately presented in Table 7.16.

Table 7.16 Sequential Estimation Results

Variables	Estimated Parameters (T-ratios)
1. Overall assessment: **Dummy 1:** Totally unsatisfied & Unsatisfied **Dummy 2:** Satisfied & Totally satisfied **Base:** Neutral	 -0.4203 (-2.2) 0.3246 (2.1)
2. Cash ticket type: **Dummy variable:** Zonal fare **Base:** Flat fare	 0.7338 (3.3)
3. Seat Availability by using travel cards or smart cards, comparing with cash **Dummy variable:** Slightly better or Better or Much better **Base:** No difference	0.6915 (4.0)
4. Top-up/purchase methods of travel cards **Dummy variable:** ticket offices **or** banks **Base:** Both ticket offices and banks	 -0.4371 (-2.1)
5. Top-up/purchase methods of smart cards **Dummy 1:** At ticket offices **Dummy 2:** Banks **Dummy 3:** Agencies **Base:** two or three top up methods used	 -1.743 (-5.8) -1.376 (-5.2) -0.9295 (-3.0)
6. Easiness of topping up/purchasing **Dummy 1:** Very difficult & Difficult **Dummy 2:** Easy & Very easy **Base:** Neutral	 -0.9246 (-3.0) 0.5916 (2.6)
ASC-travel cards:	-1.783 (-7.1)
ASC-smart cards:	-1.18 (-4.1)
Scale Factor (θ):	0.5608 (13.7)
Log likelihood at zero:	-793.319
Log likelihood:	-489.8909
No. of Observations:	782
Rho-squared value w.r.t constants:	0.3799

From Table 7.16, it can be seen that the scale factor is less than 1.0 in the sequential estimation, which means that the SP data have less error than the RP data. It may be because the SP survey specifies the choice context (designed under control) better than the RP survey. After scaling, estimates of the RP specific variables remain the same sign and statistical significance as the stand-alone model. In addition, relativities between these dummy variables also can explain respondents' rational perceptions in terms of the size of estimates. The goodness of fit measured by Rho-squared value with respect to constants is 0.3799, which is acceptable and regarded as the good model fits in the sequential estimation.

After obtaining the scale factor in the sequential estimation, the next task is to multiply the SP data by the scale (θ) to get the modified SP data set, then use the new SP data to carry out forecasting.

Moreover, in order to compare different estimation approaches for modelling the joint RP and SP data, besides the sequential estimation, another estimation approach, the simultaneous estimation, also is used. In the mean time, the results by these two approaches are compared in the following section to identify whether these two methods are different or not due to different estimation procedure, and which model is preferred for the further analysis.

7.4.2. Simultaneous Estimation of the RP and SP Data

The simultaneous estimation approach is based on a hierarchical structure as shown in Figure 4.5, Chapter 4. In the hierarchical structure, three RP alternatives are set on the upper level with the standard MNL model structure, because in Section 7.2.2, the scale factor is very close to 1.0 in the HL model for the RP alternatives, and we can deem that the MNL model is almost the same as the HL model for the RP data. In the lower level of Figure 4.5, eight SP alternatives from four different binary choice experiments are allocated. In Section 7.3.3, the final hierarchical structure for combining the SP data suggests that SP-1, 2 and 4 are in the upper level; SP 3 is set in the lower level and scaled by θ, therefore, in the hierarchical structure of the simultaneous estimation for combining the RP and SP data, two scale factors (θ_1 and θ_2) are used to distinguish two different sub-nests within the SP data sets (SP1, 2 and 4 scaled by θ_1 and SP3 scaled by θ_2).

The simultaneous estimation results for the combined RP and SP data are presented in Table 7.17. From Table 7.17, we can observe that the estimated results have the correct sign and most of them are statistically significant at 95% level. Compared with the sequential estimation approach, several features about the simultaneous estimation outputs should be pointed out:

- Estimates of common attributes ('cost' and 'time'): in the simultaneous estimation, the model deals with the scale between the RP and SP data in one stage, therefore, the estimates of cost and time are presented after scaling in Table 7.17. But the sequential estimation separately models the RP and SP data (e.g., SP first, and then RP) and produces the scale factor. The scale (0.5608) by the sequential estimation is required to multiply the SP data to adjust the SP data when using the SP data to forecast. By comparing the estimates of common attributes in the simultaneous estimation and modified cost and time coefficients of the sequential estimation, we can see that estimated coefficients in two different models are very close to each other (See Table 7.18 below).

Table 7.17 Simultaneous Estimation Results

Variables	Parameter Estimations	T-ratios
1. Travel cost (Yuan)—RP and SP	-0.1064	-13.7
2. Boarding time (second)—RP and SP	-0.007896	-4.8
3. Overall assessment-RP:		
Dummy 1: Totally unsatisfied & Unsatisfied	-0.4726	-2.4
Dummy 2: Satisfied & Totally satisfied	0.3236	2.1
Base: Neutral		
4. Cash ticket type-RP:		
Dummy variable: Zonal fare; **Base:** Flat fare	0.4500	2.1
5. Seat Availability by using travel cards or smart cards, comparing with cash-RP		
Dummy variable: Slightly better or Better or Much better	0.6993	4.0
Base: No difference		
6. Top-up/purchase methods of travel cards-RP		
Dummy variable: ticket offices **or** banks	-0.4488	-2.1
Base: Both ticket offices and banks		
7. Top-up/purchase methods of smart cards-RP		
Dummy 1: At ticket offices	-1.734	-5.7
Dummy 2: Banks	-1.369	-5.2
Dummy 3: Agencies	-0.8823	-2.7
Base: two or three top up methods used		
8. Difficulty of topping up/purchasing-RP		
Dummy 1: Very difficult & Difficult	-0.8822	-2.8
Dummy 2: Easy & Very easy	0.6058	2.6
Base: Neutral		
9. Whether passengers can get changes if they pay bid money value: 0: No; **1:** Yes --- **Cash (SP)**	0.1455	2.0
10. Deposit –Smart cards(Yuan)-SP	-0.01277	-7.2
11. Service-Travel cards SP:		
1: Limited routes: Only one bus, or light rail route service in urban area (servicetc1);	-0.6703	-6.7
2: Unlimited routes with extra charge: 10% more than limited services (Servicetc2);	-0.7355	-6.5
3: Unlimited routes with extra charge: 15% more than limited services (Servicetc3)	-0.8347	-7.5
Base: Unlimited routes without any extra charges		
12. Service-Smart cards-SP:		
1: Unlimited routes with extra charge: 10% more than limited services (servicesc1);	-0.1585	-1.1
2: Unlimited routes with extra charge: 15% more than limited services (servicesc2);	-0.2713	-2.0
3: Unlimited routes with extra charge: 20% more than limited services (servicesc3)	-0.5121	-3.5
Base: Unlimited routes without any extra charges		
13. Overdraft-smart cards-SP: Overdraft function in smart cards: 0: No; 1 Yes	0.4427	4.6
14. Multifunction-Smart cards-SP:		
1: Shopping, telephone, entertainment (mf1);	0.2469	2.6
2: Shopping, telephone, entertainment, parking and tolling (mf2);	0.6779	6.1
3: Shopping, telephone, entertainment, parking, tolling and banking (mf3)	0.9473	7.6
Base: No. only for public transport		

(Table 7.17 continued)

15. Geographic Area-Smart cards-SP: **1:** Urban and rural areas (Geo1); 2: Dalian and other nearby cities (Geo2); 3: Within one province (Geo3) **Base:** Only urban area	 0.3639 0.6281 0.8572	 3.6 5.1 6.6
16. Top-up-travel cards-SP: **1:** Ticket offices and agencies (topuptc1); 2: Ticket offices and agencies and banks (topuptc2) **Base:** Ticket offices	 0.09557 0.7637	 0.6 3.6
17. Top-up-smart cards-SP: **1:** Ticket offices, banks, agencies (topup1); 2: Ticket offices, banks, agencies, self-adding value machine (topup2); 3:Ticket offices, banks, agencies, self-adding value machine, telephone and Interne (topup3) **Base:** Only at ticket offices	 0.3520 0.6965 1.402	 3.0 4.9 5.7
ASC-travel cards (RP):	0.2169	1.1
ASC-smart cards (RP):	0.6982	2.8
ASC-smart cards (SP2):	-0.5414	-5.0
ASC-smart cards (SP4):	-0.1985	-1.5
θ1 (SP1, 2 and 4):	1.453	12.2
θ2 (SP3):	1.833	10.9
Likelihood:	-2797.3845	
No. of Observations:	6375	
Rho-squared value w.r.t. cons:	0.4204	

Table 7.18 Comparison of Estimates of Common Attributes in Different Models

	Pure RP	Pure SP	The Sequential	The Simultaneous
Travel Cost	-0.108	-0.1297	-0.1122	-0.1064
Boarding Time	0.00678	-0.0131	-0.00835	-0.007896

- It should be noted that the sign of time coefficient for the pure RP model is positive, while the pure SP, the sequential and simultaneous estimation present negative estimates in Table 7.18. This is due to the different expression of question about boarding time difference in the RP and SP survey. In the RP survey, we asked respondents "how much quicker is the boarding time by using card payment methods than using cash". Therefore it is reasonable that the time coefficient has a positive sign, which means that the quicker the boarding time by using cards, the higher the utility of the card payment. However, in the SP survey, in order to reduce the number of attributes so as to reduce the complexity of the survey design, we used "how much slower of cash payment than card payment" to present the boarding time difference. Actually both designs for boarding time have the same effect on the utility models, except the different signs. Absolute values of time coefficients in the RP and SP can still be comparable. Therefore, according to Table 7.18, it can be safely concluded that estimated time coefficients in the sequential and simultaneous estimation are very close to the pure RP model (absolute values).

- When comparing other estimates of RP/SP specific variables in the simultaneous estimation with results by the pure RP and pure SP estimation results, we can see that there is a great similarity in the values of each corresponding parameter between Table 7.17, 7.4 (the pure RP model) and 7.15 (the pure SP model) . However, the pure RP and SP estimation yields parameters with slightly higher t-statistics. This may be due to the fact that the simultaneous estimation method uses the same sample size to jointly estimate more parameters.
- The same as the pure RP and SP models and sequential estimation approach, the simultaneous estimation approach also can achieve consistent and correct sign for coefficients and reasonable relativities between dummy variables within an attribute (Please see Table 7.4, 7.15 and 7.17).
- Scale factors: in the simultaneous estimation method, two scale factors are introduced. θ_1 is used to scale alternatives in SP 1, 2 and 4, θ_2 for scaling SP 3. In the simultaneous estimation approach, two scales are all greater than 1.0 (1.453 and 1.833), which correspond to the scale factor less than 1.0 in the sequential estimation, because the scale factor in the sequential estimation is the inverse of the scale factor in the simultaneous estimation. It can be found that inversed values of 1.435 and 1.853 (0.6882; 0.5397) are very similar with the scale value in the sequential estimation (0.5608). Moreover, the scale greater than 1.0 in the simultaneous estimation also means there is bigger error in the RP data than in the SP data. Therefore, from this point, the sequential and simultaneous estimation achieved the similar results to scale different data sources in this research.

However, the simultaneous estimation approach allows for controlling the problem of the scale parameter in one stage and avoids separately inputting the RP and SP data, therefore, the estimation results by the simultaneous estimation approach are finally used to compare with estimation results by the pure RP and pure SP models to examine the model validity and carry out the demand forecast.

7.5. Measuring the RP and SP Choices

After getting estimated coefficients in the discrete choice modelling stage, the next stage for the data analysis is to measure these estimations to identify respondents' perceptions and choice behaviour. The following measurements need to be examined in this section, including:

- Valuation of boarding time savings (VOBTS)
- Valuation of qualitative attributes (PT service, multifunction, geographic areas covered, top-up/purchase methods, *etc*)

Another purpose in this section is to compare the valuation of attributes from different models (i.e., the pure RP model, pure SP model and joint RP/SP model by the simultaneous estimation approach) to verify the estimation.

7.5.1. Value of Boarding Time Savings

First of all, VOBTSs for the pure RP, pure SP and the joint RP and SP model discussed previously, are presented and compared in Table 7.19.

Table 7.19 Value of Boarding Time Savings in Different Models

Models	VOBTS (t-ratio)
Pure RP Model	3.6yuan/min (1.76)
Pure SP Model	6.06yuan/min (12.89)
Joint RP and SP with the Simultaneous Approach	4.45yuan/min (9.84)

In the pure RP MNL model, the value of boarding time savings is about 3.6yuan/minute (0.06yuan/second). In the pure SP model, the VOBTS is 6.06yuan/minute (0.101yuan/second), which is greater than that of the pure RP model. Compared with the pure RP and SP models, the VOBTS in the joint RP/SP model with the simultaneous estimation is about 4.45yuan/minute (0.074yuan/second), which is higher than the VOBTS in the pure RP model and lower than the value in the pure SP model. Although in China there is no evidence of 'value of time' to be compared with the result from this research, the possible reason for this result would be that as aforementioned the estimation results in the joint RP/SP analysis can be regarded as the average of the estimation results by the pure RP and SP models. In the RP survey all analyses are based on respondents' actual choice behaviours, particularly for the smart card payment, new features, such as multifunction, flexible top-up/purchase options and wider geographic areas covered, *etc.*, have not been fully implemented by operators and realised by PT passengers. On the contrary, in the SP survey, after introducing new features of payment alternatives in the survey design, the travel cost may be increased to some extent as compensation for the better services in some scenarios. Therefore, to get better fare payment services, it may be reasonable that the respondents were willing to pay more in the SP survey for their quicker boarding time than in the RP survey.

It also should be noted that in Table 7.19, the t-ratios of VOBTS in the pure SP model and the simultaneous estimation are statistically significant with 95% confidence level. The t-ratio of the VOBTS in the pure RP model is slightly insignificant due to the insignificance of the estimated time coefficient.

7.5.2. Value of Other Attributes

According to Equation 4.36 in Chapter 4, we can get monetary valuations for dummy variables. The valuation of qualitative attributes can provide respondents' perceptions to different levels within an attribute. Table 7.20 lists monetary valuations for some attributes, including 'whether passengers can get change back if paying by cash', 'PT service routes covered by TC', 'PT service routes covered by SC', 'Multifunction', 'Overdraft', 'Geographic areas covered' and 'Top-up/purchase methods'.

Table 7.20 Valuation of Attributes for Mixed RP and SP Data

Variables	Estimation	Valuation of Attributes
Travel cost:	-0.1064	N/A
Getting change back if paying by cash. Base: No	0.1455	1.37yuan (2.1)
PT service routes covered:		
TC: Servicetc1: *Limited routes: Only one bus, or light rail route service in urban area. But passengers still can use cash to pay other PT services.*	-0.6703	6.30yuan (4.3)
Servicetc2: *Unlimited routes with extra charge: 10% more than limited services*	-0.7355	6.91yuan (4.87)
Servicetc3: *Unlimited routes with extra charge: 15% more than limited services*	-0.8347	7.84yuan (5.21)
Base: Unlimited routes without any extra charges		
SC: Servicesc1: *Unlimited routes with extra charge: 10% more than limited services.*	-0.1585	1.49yuan (1.87)
Servicesc2:*Unlimited routes with extra charge: 15% more than limited services*	-0.2713	2.55yuan (2.04)
Servicesc3: *Unlimited routes with extra charge: 20% more than limited services*	-0.5121	4.81yuan (2.98)
Base: Unlimited routes without any extra charges		
Multifunction:		
Mf1: *Shopping, telephone, entertainment*	0.2469	2.32yuan (2.23)
Mf2:*Shopping ,telephone ,entertainment, parking and tolling*	0.6779	6.37yuan (5.14)
Mf3:*Shopping ,telephone ,entertainment, parking, tolling and banking*	0.9473	8.9yuan (6.99)
Base: None. only for public transport		
Overdraft, Base: No	0.4427	4.16yuan(3.69)
Geographic areas covered:		
Geo1: *Dalian Urban and rural areas*	0.3639	3.42yuan (3.11)
Geo2: *Dalian plus other nearby cities*	0.6281	5.9yuan (4.35)
Geo3: *Within one province*	0.8572	8.05yuan (5.18)
Base: Only at Dalian urban area		
Top-up/purchase methods:		
TC: Topuptc1: *Ticket offices and agencies*	---	---
Topuptc2: *Ticket offices and agencies and banks*	0.7637	7.18yuan (3.2)
Base: Only at Ticket offices		
SC: Topupsc1: *Ticket offices, banks , agencies*	0.3520	3.31yuan (2.1)
Topupsc2: *Ticket offices, banks, agencies, self-adding value machine*	0.6965	6.55yuan (3.03)
Topupsc3: *Ticket offices, banks, agencies, self-adding value machine, telephone and Internet*	1.402	13.18yuan(4.88)
Base: Only at ticket offices		

In Table 7.20, estimated coefficients are presented by dummy variables relative to the base predefined. Meanwhile, the t-ratios of these valuations of attributes are listed to indicate the statistical significance of these monetary values. Discussions about these valuations of attributes of fare payment alternatives are as follows:

Getting change back if passengers cannot pay exact money

The valuation of "whether passengers can get change back if they cannot pay exact money" can be used to measure the convenience of cash fare payment for passengers. As can be seen in Table 7.20, the monetary valuation of this attribute is about 1.37yuan, equivalent to about 9p in GBP. Because we defined the unit of travel cost in the model is yuan per month, 1.37yuan can indicate that those passengers who perceived "Yes, PT drivers can return change back when passengers cannot pay exact cash" were willing to pay 1.37yuan per month for cash payment, compared with those chose cash fare payment with "No change back from PT drivers" (Base).

Although the evidence of the monetary valuation on this variable is not available currently, the result can be explained by the cash fare structure in use. If cash users could get change back when they cannot pay exact cash, the cash fare structure could be more flexible, like the PT fare structure in the UK (e.g., a single fare could be 90p, £1, £1.20, etc.) rather than the current application in Dalian, China (i.e., 1yuan for flat fare). Therefore, under such circumstances (change is refundable), a cash user would travel more, and their willingness to pay for cash fare would be related with their more travel demand.

The t-ratio of valuation of 'getting change back if paying by cash' is 2.1, statistically significant that 95% level.

PT service routes covered

The valuations of 'PT service routes covered' are separately measured for travel cards and smart cards due to different levels designed for these two payment options. For the service route covered by TC and SC, we set 'unlimited PT routes covered' as the base, which can be viewed as the most beneficial level to travel card/smart card users. We observe that all the dummies are presented negative sign relative to the base. So those respondents chose travel card/smart card payment with higher travel cost for 'unlimited PT service' had relatively higher WTP (i.e., 'Servicetc2, Servicesc2 and Servicesc3' in Table 7.20). For those chose travel cards with 'limited PT routes', they also would like to pay 6.3yuan/month, which is more than 'unlimited service routes without any extra charge'.

When comparing the size of valuations of services, we can see that monetary values increase with the change of service routes from 'limited' to 'unlimited' (with some extra cost), not only for travel cards but also for smart cards. Therefore, it can be implied that the willingness to pay of card users would like to pay more for better PT route services.

Regarding the plausibility of valuation of PT route services, some evidence of travel cards in the RP survey would explain these results. In the actual use of travel cards, the cost of a travel card with limited service route is about 29yuan per month, the extra cost by some other payment methods due to the restriction of the travel card is about 15yuan per month. So the total cost per month is about 44yuan. Normally a travel card with unlimited route (but without extra charge) costs about 40yuan per month in Dalian, 4yuan less than the cost by using the travel card with limited services plus extra cost by other payment methods. The 4yuan difference is close to 6.3yuan in Table 7.20 ('Servicetc1'), which indicates the plausibility of the estimation result.

Multifunction

Valuations of multifunction are also calculated by dummy variables. It can be seen from Table 7.20 that the valuation of 'mf3' is greater than 'mf2', 'mf2' greater than 'mf1'. The dummy variable 'mf3' is designed as the highest level with the most extra functions compared with other dummies and the base. Therefore, when respondents chose smart card payment with some multifunction, they were fully aware that smart cards could bring them convenience and benefits for their journey and some other social services, therefore, they were willing to pay more for smart cards with extra services than without (or less) extra services.

Regarding the relativities between monetary valuations of multifunction, there is no evidence available for the comparison at the moment in Dalian, but through the definition of the attribute levels, we still can get some clues for the difference of valuations. From 'Mf1' to 'Mf3', more and more extra functions are added in the smart card ticketing. Particularly for the functions of 'entertainment: admission fee payment', 'parking fee payment' and 'tolling', they would trigger more accesses of users under the integrity of these social services by one smart card. Therefore, it is understandable that smart card users would pay more for the better services. Moreover, some banking services are not free of charge in China, such as withdrawing across different banks' cash machines, so it is reasonable that the willingness to pay for 'Mf3' is greater than 'Mf2' and 'Mf1' in Table 7.20.

Overdraft

The valuation of 'overdraft' is 4.16yuan/month, which means that if respondents choose smart cards with overdraft function, they would like to pay 4.16yuan per month. This kind of policy is particularly convenient for those passengers who pay by smart cards but the credit left in the card is not enough to pay a single ticket. The overdraft policy can guarantee the last trip for card users.

When comparing with the actual use of smart cards in Dalian, this value is slightly higher than the potential cost under the situation if 'overdraft' is not applicable, because when the credit left in the smart card is not enough for a single ticket, the potential cost by cash would be 1-2yuan (if we assume the passengers could take 1-2 rides for one trip).

However, besides the monetary value can be measured, the potential convenience by using smart cards with the overdraft facility would not be exactly compared with the situation of 'non-overdraft'.

Geographic areas covered

Like those dummies in 'multifunction' variable, the similar outcomes of monetary valuations of 'geographic areas covered' also present respondents' perceptions towards the smart card payment combined with the attribute-level of 'geographic areas covered'. From Table 7.20 we find that the wider areas smart cards can cover, the higher the value of the dummy would be. The higher monetary valuation also indicates the higher WTP of respondents when they use smart cards with covering wider geographic areas.

The differences of valuations of three dummies can be explained by the differences of fares if travelling different areas. To link Dalian urban area with the rural area, the actual cost is about 5yuan in the current situation, which is close to the monetary value of 'Geo1' (3.42yuan). And if travelling from Dalian to neighbouring cities, the current travel cost is about 10yuan, also corresponding to 'Geo3': 8.05yuan.

Top-up/Purchase Methods

For top-up/purchase methods, we designed different levels for travel cards and smart cards respectively. Among these dummies, 'topupsc3' for smart cards has the highest monetary valuation (13.18yuan/month), indicating respondents' highest WTP for using smart cards with the most various top-up/purchase options. Such high monetary valuation would be explained by the extra cost avoided due to using a variety of top-up/purchase options, because if these top-up facilities were introduced, smart card users may have lots of conveniences to top up/buy smart cards, avoiding some extra efforts (such as going to some ticket office) for topping up their smart cards under the situation of 'only at ticket offices'.

Similar with other variables with two or three dummies discussed above, along with the attribute level becoming better compared with the base, the valuation of "top-up" dummies for TC and SC increases. This means respondents' WTP also goes up with the attribute becoming better than the base.

It also should be noted that because of the insignificant estimation in statistics for 'Topuptc1' the value of this dummy variable is not taken into account. Only those significant estimations are measured in this section.

7.5.3 Importance of Attribute

To sum up, from valuation of attributes in Table 7.20, we can determine the most and least important features among these alternative attributes. The following features are relatively important compared with others when respondents trade off different payment methods, including:

- Services covered by travel cards;
- Multifunction of smart cards;
- Geographic areas can be covered by smart cards; and
- Top-up/purchase methods of smart cards.

because on average these attributes have higher monetary valuations than others.

Compared with the most important features of payment methods, two features: 'whether passenger can get change back if paying cash' and 'service route covered by smart cards' have relatively low monetary valuations. It can be concluded that when respondents trade off between cash and other payment methods, 'whether they can get change back if they cannot pay exact fare' would not influence their decision as much as those four features listed above. For smart cards, compared with the feature of service routes covered by smart cards, respondents would more focus on other attributes, such as multifunction, wider geographic areas covered, *etc*.

Meanwhile, among these most important attributes, the level is most preferred can also be identified through Table 7.20. As can be seen, 'Topupsc3' has the highest monetary valuation, indicating that a variety of topping up/purchase options is most preferred and respondents are willing to pay much higher than other features, because various topping-up/purchase options can bring convenience to card users and increase the accessibility to public transport for all passengers.

7.6. Market Share Forecasting

In the modelling application stage, the main task is to use estimated choice model to carry out demand forecasting. The forecast demand of fare payment choices is based on the aggregated market share after obtaining respondents' individual choice probabilities. Regarding the technique of aggregating individual choice probability, the Department for Transport (2004) states that consistent estimates of market share can be obtained by using sample enumeration. Sample enumeration involves calculating, for each decision maker in the sample, the probability of choice for each alternative in the choice set. These probabilities are then aggregated over decision-makers, and the average probabilities can be obtained by dividing through by the sample size. Sample enumeration is used in this study to allow application of the model.

The market share forecasting model is based on the estimated coefficients derived from the joint RP and SP model (the simultaneous estimation) firstly. Choice probabilities are presented by three fare payment alternatives in the RP/SP survey. Besides demand forecasts by estimated coefficients in utility functions, another task in the forecasting analysis is to examine the influence of travel cost (fare) on the users' choice behaviour (i.e., fare elasticities analysis), based on the estimates of travel cost by the joint RP and SP model with the simultaneous estimation approach.

7.6.1. Market Share Forecasting for Different Fare Payment Methods

First of all, the individual choice probabilities of fare payment alternatives were obtained according to Equation 4.6 in Chapter 4 and estimated parameters in the utility models for the RP and SP data (Table 7.17). These probabilities are influenced by the system characteristics (fare payment attributes and levels). The predicted aggregate market share using fare payment methods for the RP and SP survey is an average of the individual choice probabilities of the whole data set used for the model estimation. As discussed previously, the joint RP/SP model by the simultaneous estimation is used, therefore, the market share forecasting also will be based on the joint RP/SP model.

In order to analyse the change of the user demand based on the change of variables, in the forecasting analysis, some scenarios are used, in which one variable of smart cards will be changeable and others will be fixed on a given level, because if combinations of payment variables were used, the forecasting model would become very massive and complicated. Secondly, because the objective of this research is to forecast users' demand so as to evaluate benefits of smart cards, the forecasting analysis mainly focuses on effects of attributes of smart cards on PT users' payment demand. The following factors are considered in the demand analysis, including:

- Travel cost;
- Boarding time difference;
- PT service routes covered;
- Overdraft;
- Deposit;
- Multifunction;
- Geographic areas covered;
- Top-up/purchase options.

<u>Travel cost</u>

Table 7.21 Demand Forecast Based on Cost Variable

	20yuan	40yuan	60yuan	80yuan	100yuan	120yuan	150yuan
Cash	9%	17.2%	43.2%	58.9%	61.5%	61.7%	61.7%
TC	7%	14.3%	33.2%	35.1%	37.7%	38.3%	38.3%
SC	84%	68.5%	23.6%	6%	0.8%	0	0

In Table 7.21, only the cost variable of smart cards is considered, which is gradually added from 20yuan to 150yuan. The travel cost of cash and travel cards are set to the average level according to the RP survey (48.31yuan and 52.26yuan, respectively). Meanwhile, other variables of cash, travel cards and smart cards are fixed on the base level or current situation. Regarding the base level of the three payment options in the RP and SP survey, please refer to Table 7.4, 7.6-7.9.

From Table 7.21, it can be seen that with the increase of travel cost of smart cards the market share of smart cards significantly reduces (from 84% when at 20yuan level to 0 when at 150yuan level). Particularly when the cost of smart cards increases from 60yuan to 100yuan, the market share of smart cards almost becomes zero at 100yuan level. When the cost is less than 40yuan, the market share of smart cards is more than other two payment options. When the cost of smart cards is at 60yuan level, the shares of cash and travel cards significantly increase and exceed the share of smart cards. Therefore it can be implied that 60yuan and 100yuan would be two critical levels, which could influence PT users' choices of smart cards when other variables of three payment methods keep the current or the base level. This result would be reasonable, because the average costs of cash and travel cards are around 50yuan. When the cost of smart cards is more than 50yuan, it would be possible for the majority of smart card users to switch to other two payment options. Moreover, the base level of smart cards cannot show a very obvious advantage over cash and travel cards, such as deposit (30yuan), multifunction (only for PT fare payment) and geographic areas covered (only Dalian urban area), etc. So when the cost of smart cards increases, smart card users would readily change from smart cards to cash or travel card payment.

Besides the change of smart card share, we find the changes of cash and travel cards towards the increase of smart card cost are very different. Compared with the mild increase of the travel card market share, the change of the market share of cash payment is more significant, particularly when the cost is over 60yuan, which means that more smart card users would switch to cash payment in case the travel cost of smart cards goes up. The reason for this result would be that the prepayment of travel cards and less flexibility than cash may cause the relative low utility of travel cards, compared with cash.

<u>Boarding time difference</u>

Table 7.22 Demand Forecast Based on Boarding Time Variable

	0second	5second	10seconds	20seconds	40seconds	60seconds	80seconds
Cash	52.4%	49%	44.3%	37.2%	34.3%	29.4%	25.9%
TC	28.3%	29.9%	32.8%	33.1%	33.9%	34.1%	35.9%
SC	19.3%	21.1%	22.9%	29.7%	31.8%	36.5%	38.2%

The boarding time difference presents seven levels in Table 7.22 (from 0s to 80s). As can be seen, the boarding time factor also would influence the choice of fare payment methods. With the increase of boarding time difference (quicker boarding time by using smart cards), the share of cash payment goes down. Meanwhile, when the boarding time is over 40s, PT users would prefer to use card payment (travel cards or smart cards). However, when the boarding time difference is varied from 0s to 40s, the use of cash payment would still be dominant in the three payment methods. Therefore, we can imply that 40-60s of boarding time difference would be a critical level to make the PT users' perception change. But admittedly, the effect of boarding time difference on the smart card demand is not as significant as that of the travel cost of smart cards in Table 7.21.

Among the three fare payment methods, the change of boarding time difference of smart cards does not seem to have an obvious influence to the use of travel cards, because from 0s to 80s, the share of travel cards only increases about 7%, much lower than the changes of cash and smart cards. Some other reasons would cause this result, including the travel cost, PT service routes covered, and special features of smart cards.

PT service routes covered (SC)

Table 7.23 Demand Forecast Based on Service Routes Variable

	Unlimited route without extra charge (Base)	**Unlimited route with extra 10% extra charge**	**Unlimited route with extra 15% extra charge**	**Unlimited route with extra 20% extra charge**
Cash	32.3%	30.2%	29.7%	29.2%
TC	17.6%	27.5%	32.1%	34.6%
SC	50.1%	42.3%	38.2%	36.2%

Table 7.23 lists the forecast result when the PT service routes variable of smart cards changes. At the base level, the market share of smart cards is about 50%, indicating that the majority would prefer to use the smart card ticketing when it does not have extra charges for the unlimited PT service routes. With the increase of extra charge for the unlimited PT services, it is not surprising that the market share of smart cards gradually drops down. Correspondingly, more PT users would switch from smart cards to travel cards when smart cards charge more. Although travel cards limit the PT service route in the current situation, the travel cost of travel cards is lower. Compared with the interaction between smart cards and travel cards, the effect of PT services covered by smart cards on the use of cash payment is relatively mild, because the share of cash only varies between 32.3% and 29.2%. The reason for this result would be the low substitutability of cash to smart cards. That is to say whatever the PT service routes covered by smart cards, the demand of the cash payment may not change too much.

Overdraft (SC)

Table 7.24 Demand Forecast Based on Overdraft Variable

	No (Base)	Yes
Cash	31.2%	27.6%
TC	32%	22.2%
SC	36.8%	50.2%

When changing the feature of overdraft attribute of smart cards, the market share also shows the significant change for the demand of the smart card ticketing. The increase of the market share of smart cards (from 36.8% to 50.2%) indicates that when the overdraft facility is introduced, much more PT users would choose smart cards. Compared with smart cards, travel card users would switch from travel cards to smart cards with overdraft function, because travel cards are required to renew (top-up) on a fixed day and cannot be overdrawn. However, the market share of cash would not change too much (31.2% to 27.6%) when the overdraft variable changes, because cash is highly available for any PT users. Therefore, whether the smart card has overdraft facility or not would not become the dominant factor for cash users to switch from cash to smart cards.

Deposit (SC)

Table 7.25 Demand Forecast Based on Deposit Variable

	0yuan	20yuan	30yuan	50yuan
Cash	26.9%	27.7%	31.2%	33.1%
TC	19.9%	27.6%	32%	34.6%
SC	53.2%	44.7%	36.8%	32.3%

Similar with 'PT service route covered' and 'Overdraft' variables, at the base level, the use of smart cards almost occupies more than 50% of the three payment methods in the market place. With the increase of deposit, the share of smart cards gradually reduces, while the shares of cash and travel cards go up. But compared with the factors discussed above, the influence of 'deposit' attribute would not be as significant as others, because when the deposit is 30yuan or 50yuan, the shares of the three payment options are almost the same (around 30%). The reason for this result is that the deposit can be refundable in most smart card projects and most smart card users have been aware of this point.

Multifunction (SC)

Table 7.26 Demand Forecast Based on Multifunction Variable

	Only for PT fare payment	Shopping, telephone, entertainment	Shopping, telephone, entertainment, parking and tolling	Shopping, telephone, entertainment, parking, tolling and banking
Cash	31.2%	29.9%	27.9%	23.1%
TC	31.7%	25.5%	22.3%	18.6%
SC	37.1%	44.6%	49.8%	58.3%

Table 7.26 presents demand forecast results based on the multifunction variable of smart cards. Through the change of the market share of smart cards with the variation of multifunction, we can see the importance of this attribute to the demand of smart cards. At the base level, the share of smart cards is slightly more than other two payment methods, but when some functions are added, the shares of the three payment methods become obviously different. When the multifunction is at the best level, the percentage of the use of smart cards reaches the highest level (58.3%). In addition, the change of the market shares of cash and travel cards show how cash and travel card users' preference change: more travel card users would switch to smart cards due to the multifunction of smart cards, because except the similar basic features of two card payment options, the introduction of multifunction can improve the service quality of the smart card payment, therefore, the percentage of travel cards would significantly reduce.

Geographic areas covered (SC)

Table 7.27 Demand Forecast Based on Geographic Areas Variable

	Only in Dalian urban area	Dalian Urban and rural areas	Dalian and other nearby cities	Within one province
Cash	29.8%	27.9%	25.5%	22.3%
TC	33.6%	27%	23.3%	21.5%
SC	36.6%	45.1%	51.2%	56.2%

Through Table 7.27, the importance of 'geographic areas covered' attribute also can be seen: at the best level, the percentage of the smart card use is dominantly more than other two payment methods. The reduction of the market share of travel cards (from the base level to the best level of geographic areas covered) is more significant than that of cash payment. It can be concluded that due to the high substitutability of smart cards to travel cards, the improvement of geographic areas covered by smart cards would result in the switch of travel cards to smart cards. Compared with travel cards, the influence of this attribute to cash payment is relatively mild (29.8% to 22.3%). This is also due to the high availability of cash to any PT users, wherever it is used.

Top-up/purchase options (SC)

Table 7.28 Demand Forecast Based on Top-up/purchase Options Variable

	Only at ticket offices	Ticket offices, banks, agencies	Ticket offices, banks, agencies, self-adding value machine	Ticket offices, banks, agencies, self-adding value machine, telephone and Internet
Cash	30.3%	28.6%	27.2%	26.1%
TC	33.2%	32%	28.1%	24.3%
SC	36.5%	39.4%	44.7%	49.6%

In Table 7.28, the demand forecast result shows that with the top-up/purchase options becoming more, the use of smart cards would increase (from 36.5% at the based level to 49.6% at the best level). Meanwhile, the market share of travel cards would decrease more than the reduction of the market share of cash payment (i.e., for travel cards, about 11% reduction; for cash, about 4% reduction). The demand forecast result indicates the interaction between smart cards and other two payment methods: because the top-up/purchase options of travel cards is set to the base level ('ticket offices') when improving the top-up/purchase options of smart cards, the convenience of the smart card ticketing with a variety of top-up/purchase options would attract more travel card users to choose smart cards. For cash users, as discussed previously due to the high availability of cash, their current choice would not be affected too much by the top-up/purchase options of smart cards.

Beyond the single factor analysis, scenarios of multiple changes of smart cards also can be introduced in the forecasting analysis. However, combinations of multiple variables would make the forecast model very complicated, therefore in order to simplify the model, the 'multiple changes' analysis only considers two scenarios: (1) let all variables of smart cards remain the best level (except that the travel cost is set to the average level, because it is obviously impossible to let the travel cost of smart cards be zero. Therefore giving an average level is more reasonable.); (2) let all variables of smart cards remain the medium level. In the two scenarios, cash and travel cards still keep the current situation. Table 7.29 lists the forecast results of these two scenarios.

Table 7.29 Demand Forecast—Multiple Changes of Variables of SC

	Cash	Travel Cards	Smart Cards
Scenario 1 (best level of SC)	12.7%	9.1%	78.2%
Scenario 2 (medium level of SC)	22.5%	19.9%	57.6%

From Table 7.29, it can be seen that when the variables of smart cards (except the travel cost) are set to the best level, and other two payment methods are presented as the current situation, the market share of smart cards are dominantly more than that of cash and travel cards, about 78.2%. Even if the smart card ticketing is presented by the medium level, the share of smart cards is still more than 50% in the market place. Therefore, we can conclude that the multiple effect of the smart card payment variables is encouragingly positive to the PT users' payment method choices. If the relevant facilities of smart cards are applied at the same time, the traditional fare payment users would switch to smart cards due to much more conveniences by using the novel payment option.

Meanwhile, not only in the single factor analysis, but also in the multiple factor analysis, we find that in general the share of travel cards would reduce significantly when the payment features of smart cards become better. The high substitutability of smart cards to travel cards results in the stronger competition between travel cards and smart cards. Therefore, with the improvement of the smart card application in the future, travel card users would be the main source to switch to the smart card ticketing. However, the cash user group would be less influenced by smart cards than travel cards, due to some factors, such as cash users' travel frequency, the availability of cash fare, etc.

7.6.2. Fare Elasticities

The predicted market share in the last section provides information of respondents' demand changes by trading off the combination of fare payment attributes. In this section, how the demand can be influenced with respect to changes of travel cost is analysed by travel cost elasticities. Different from the market share forecast, elasticities provide information about the sensitivity (or extent) of changes on payment demand based on a given market size rather than detailed values (i.e., predicted market share).

Two types of elasticities: own and cross elasticities, are used for different purposes in this research. Through own elasticities, changes of a fare payment demand with respect to its own travel cost can be obtained, while cross elasticities tell information about demand changes of a fare payment method with respect to changes of travel cost of other payment methods. The aggregate own and cross fare elasticities for three models (the pure RP, pure SP and joint RP and SP models) are presented in Table 7.30, 7.31 and 7.32, respectively.

Table 7.30 Own and Cross Fare Elasticities in the Pure RP Data

	Cash	Travel Cards	Smart Cards
Cash	-2.4896	1.0566	0.9639
Travel Cards	1.2436	-1.8199	0.8618
Smart Cards	1.041	0.838	-1.7335

Table 7.31 Own and Cross Fare Elasticities in the Pure SP Data

	Cash	Travel Cards	Smart Cards
Cash	-1.832	0.7516	0.9106
Travel Cards	1.2689	-1.1099	1.8618
Smart Cards	1.8037	1.5147	-1.2195

Table 7.32 Own and Cross Fare Elasticities in the Joint RP and SP Data

	Cash	Travel Cards	Smart Cards
Cash	-1.9621	0.8016	0.9219
Travel Cards	1.2606	-1.4162	1.6455
Smart Cards	1.6091	1.3365	-1.4637

Rows of three tables above represent alternatives were used to be compared with alternatives in columns (i.e., elasticities of demand of alternatives in the columns of the tables with respect to travel cost of the alternatives in the rows). For example, in Table 7.30, 1.2436 indicates a cross elasticity of demand of cash with respect of the change of travel cost of travel cards. Negative values mean the own elasticities of payment alternatives relative to their own travel cost. Positive values represent the cross elasticities of payment methods with respect to the change of travel cost of other payment alternatives, because in this research, three fare payment methods are substitutable with each other, positive cross elasticities can be explained that as the travel cost of one fare payment goes up the demand of the other will increase.

Absolute values of elasticities greater than 1.0 indicate the elastic relationship between demand of payment alternatives and travel cost (own and cross elasticities) and the percentage of change in demand for the payment method is greater than that in travel cost. Absolute values of elasticities less than 1.0 represent the inelastic relationship between demand of payment alternatives and the change of travel cost of their own or others, which means that the percentage change in demand is smaller than that in travel cost. In this research, we expect that elasticities are high because strong competition may exist in the three payment options.

From Table 7.30, we can see that absolute values of own elasticities of three payment methods in the pure RP data are greater than 1.0. That means the demand of payment methods being used is very sensitive to any changes of their own travel cost. The cross elasticities of cash payment demand with respect to the cost of travel cards and smart cards are also greater than 1.0, indicating that for current cash fare users, any slight increase or decrease of travel cost of travel cards and smart cards can result in an obvious change of cash users' choice behaviour toward travel cards and smart cards. It also shows that travel cards and smart cards are two sensitive competitors of cash payment.

The travel card payment is elastic to cash payment, but inelastic to smart card payment. Because travel cards and smart cards have a number of common or similar features in the current applications, such as faster boarding time than cash payment and value stored in advance for cards, and they can also achieve travel cost saving for frequent users, therefore, the change of travel cost of smart cards may not result in a greater change of demand of travel cards.

For smart card payment, the cross elasticities less than 1.0 with respect to both cash and travel cards show inelastic relationship between the demand of smart cards and travel cost of cash and travel cards. Therefore, the existing demand of smart cards is relatively stable and would not be influenced by changes of cash and travel cards too much.

Fare elasticities in Table 7.31 for the pure SP data show a very similar result on the own elasticities with the pure RP data. Absolute values of the own elasticities of three fare payment are greater than 1.0, which means the demand of fare payment methods with respect to change of travel cost of their own is elastic and users would like to switch to another payment methods when the travel cost changes.

In Table 7.31, cross elasticities of cash payment demand with respect to cost of travel cards or smart cards are also greater than 1.0, but cross elasticities for card payment demand with respect to cost of cash payment are smaller than 1.0. This reveals that similar with the RP data, cash users would like to switch to cashless payment when travel cost of travel cards or smart cards goes down, and vice versa. And card payment users tend to still use travel cards/smart cards when the travel cost of cash changes. Absolute values of cross elasticities between travel cards and smart cards are all greater than 1.0, which indicates that in the pure SP data the strongly competitive and substitutable relationship between travel cards and smart cards, because two types of cashless payment methods share some similar key features (e.g., quicker boarding time than cash, prepaid fare, travel cost saving for frequent travellers, etc.) and when respondents trade off these two payment options in SP situations, any changes of travel cost of one payment could cause switch from one kind of payment to another.

Elasticitis for the joint RP/SP data in Table 7.32 also present the similar results with the pure SP data, therefore, the same explanation as discussed in the pure SP data can be applied to results in the joint RP/SP model.

Comparing elasticities by three models, we can conclude some common characteristics of travel cost elasticities in the RP and SP data:

- Own elasticities: all own elasticities are greater than 1.0 in three models, indicating payment demand with respect to changes of their own travel cost is very sensitive.
- Cross elasticities between cash and card payment methods: not only in the RP but also in the SP and joint RP/SP models, all cross elasticities of cash with respect to travel cost of other two card payment methods are greater than 1.0. It can be implied that the cash user group is very sensitive and readily affected by changes of travel cost of other fare payment methods. However, cross elasticities of smart cards with respect to change of cash cost in three models are all less than 1.0, indicating relative to change of cash travel cost, the demand of smart cards tends to be stable.
- Cross elasticities between two card payment methods: in the pure SP and joint RP/SP models, all cross elasticities between two card payment methods are greater than 1.0. Therefore such high substitutability between travel cards and smart cards can be clearly seen, while in the pure RP model, respondents' actual choice behaviours look more stable between these two card payment methods, because the cross elasticities between

two card payment methods are less than 1.0. Current travel card smart card users would not consider switching to another card payment when the travel cost of another card payment reduces.

7.7. Model Validation

The validity of forecasted results is examined by using the validation sample, because the validation sample did not involve the model estimation and was retained to the well-estimated choice model. In order to examine the validity of the choice models, a comparison of market shares between the control sample and validation sample is conducted. The result of forecast market shares by the validation sample are summarised in Table 7.33-7.41. Meanwhile, to make the results from two different data sets comparable, all scenarios used in Section 7.6.1 are also considered in Table 7.33-7.41, including the single factor change (e.g., cost, boarding time, etc) and multiple changes of smart card attributes (at the best level and medium level).

Table 7.33 Demand Forecast Based on Cost Variable by Validation Sample

	20yuan	40yuan	60yuan	80yuan	100yuan	120yuan	150yuan
Cash	10.4%	18.5%	43.4%	59.1%	61.6%	60.5%	60.5%
TC	7.5%	15.1%	34.3%	35.3%	37.7%	39.5%	39.5%
SC	82.1%	66.4%	22.3%	5.6%	0.7%	0	0

Table 7.34 Demand Forecast Based on Boarding Time Variable by Validation Sample

	0second	5second	10seconds	20seconds	40seconds	60seconds	80seconds
Cash	53%	49.8%	44.8%	37.4%	35%	30.2%	26.1%
TC	28.5%	30.2%	34%	33.6%	33.9%	33.7%	36.6%
SC	18.5%	20%	21.2%	29%	31.1%	36.1%	37.3%

Table 7.35 Demand Forecast Based on Service Routes Variable by Validation Sample

	Unlimited route without extra charge (Base)	Unlimited route with extra 10% extra charge	Unlimited route with extra 15% extra charge	Unlimited route with extra 20% extra charge
Cash	33.7%	31.6%	30%	30.2%
TC	18%	27.7%	32.1%	35%
SC	48.3%	40.7%	37.9%	34.8%

Table 7.36 Demand Forecast Based on Overdraft Variable by Validation Sample

	No (Base)	Yes
Cash	31.8%	28.9%
TC	33.1%	22.2%
SC	35.1%	48.9%

Table 7.37 Demand Forecast Based on Deposit Variable by Validation Sample

	0yuan	20yuan	30yuan	50yuan
Cash	28.6%	28.5%	31.8%	33.8%
TC	20.5%	28.3%	33%	35.6%
SC	50.9%	43.2%	35.2%	30.6%

Table 7.38 Demand Forecast Based on Multifunction Variable by Validation Sample

	Only for PT fare payment	Shopping, telephone, entertainment	Shopping, telephone, entertainment, parking and tolling	Shopping, telephone, entertainment, parking, tolling and banking
Cash	32.2%	30.1%	28.4%	24%
TC	32.4%	27.1%	22.6%	18.9%
SC	35.4%	42.8%	49%	57.1%

Table 7.39 Demand Forecast Based on Geographic Areas Variable by Validation Sample

	Only in Dalian urban area	Dalian Urban and rural areas	Dalian and other nearby cities	Within one province
Cash	31.1%	29%	26.4%	22.5%
TC	34%	27.8%	24.3%	22.5%
SC	34.9%	43.2%	49.3%	55%

Table 7.40 Demand Forecast Based on Top-up/purchase Options Variable by Validation Sample

	Only at ticket offices	Ticket offices, banks, agencies	Ticket offices, banks, agencies, self-adding value machine	Ticket offices, banks, agencies, self-adding value machine, telephone and Internet
Cash	31.5%	29.2%	28.2%	27%
TC	34.3%	32.3%	29.3%	25%
SC	34.2%	38.5%	42.5%	48%

Table 7.41 Forecast Market Share of Multiple Changes in Validation Sample

Scenario	Cash	Travel Cards	Smart Cards
Best level of SC	13.9%	10%	76.1%
Medium level of SC	24.1%	20.1%	55.8%

In Table 7.33-7.41, the market shares for the validation sample are very close to the predicted results by the control sample in Table 7.21-7.29, indicating the validity of estimation results by the joint RP/SP model, except that the share of smart cards in the control sample is slightly lower than the validation sample. The reason for such result would be due to the random selection of the validation sample. From the analysis results for the model validation, we can conclude that the joint RP/SP model can perfectly replicate the actual situation when comparing forecasted and observed market shares of three payment methods by using the validation data, which were not involved into the model estimation.

7.8. Market Segmentation Analysis

An important aspect in analysing PT users' choice behaviour and carrying out demand forecast is to characterise the fare payment market based on different user groups. The purpose of the market segmentation analysis in this chapter is to analyse groups of PT users having similar needs and preferences for three fare payment methods, so as to assist in the development of PT ticketing service and marketing plans. In this context, one would expect, for example, some passengers might have a strong preference for a sort of convenient fare payment bringing them quicker boarding time, while others might be more concerned about fare payment method(s) that can meet their different travel frequency so as to save their travel cost. Such information is invaluable in developing PT services and marketing plans in the future.

In this research, detailed socio-economic variables that can be used in the segmentation analysis have been introduced in Chapter 5: Survey Design (Please see Table 5.3 in Chapter 5), including age, gender, educational background, employment status, household income, availability of private transport and attitude of willingness to prepay. If all seven factors were used in the segmentation analysis, the model would become quite massive and complicated, therefore, of these seven variables, household income, gender, and age are selected as 'the important factors to examine the effects of socio-economic variables to choice behaviour, because comparing with other socio-economic variables, age and sex factors are relatively stable and the segmentation analysis therefore can be reasonable to identify the heterogeneity of choice behaviour between different groups. For the household income factor, because it has been widely used in the previous studies, it is also considered in this research as one of important factors.

When we carry out the segmentation analysis, ideally, zero correlations between socio-economic variables are required beforehand, so that the comparison of effects of socio-economic variables is independent (Wardman, 1988). Therefore, the first task is to test the correlation between socio-economic variables. Table 7.42 lists whether the null hypothesis of the independence of two attributes can be rejected at 95% level of confidence, according to the chi-squared statistics derived from contingency tables of three socio-economic variables.

Table 7.42 Correlations of Segmentation Variables

	Age	Sex
Sex	0.06	
Household Income	0.4496	0.4740

It can be seen from Table 7.42 that in all the three pairwise comparisons of segmenting variables, the null hypothesis of independence can be rejected. In particular, between age and sex, the correlation is considerably low. Correlations between age and household income, sex and household income are less than 0.5, which can also be accepted as a low correlation. Therefore, the segmentation analysis based on these three socio-economic variables is feasible.

The segmentation models are presented separately based on different categories of socio-economic variables. In each model, only one category is involved to carry out the model estimation. Estimation results for each socio-economic category of three variables are attached in Appendix C. Meanwhile, in order to identify the effects of different models with socio-economic variables, valuations of attributes are listed in Table 7.43, 7.45 and 7.47 for age, sex and household income, respectively. Moreover, values of attributes without segmenting also are included in these three tables to reveal the heterogeneity of individual choice of payment methods in different segments.

As discussed in Table 7.20 for the model without segmentation, in Table 7.43, 7.45 and 7.47 same monetary measures are used, including value of boarding time savings (VOBTS), value of qualitative attributes, such as multifunction, geographic areas covered, top-up/purchase options, etc. Regarding the definition of those dummy variables, please refer Table 7.20.

Beyond the discussion of valuation of attribute by different segments, some demand forecast is carried out based on different segments. In order to simplify the model, only two scenarios as discussed in the previous section, are considered in the segmentation analysis: the best level of smart cards and the medium level of smart cards. The relevant results are listed in Table 7.44, 7.46 and 7.48.

7.8.1 Age

Table 7.43 lists and compares valuations of attributes based on the age segments and non-segmentation. For the segmentation analysis, four age groups generated four separate models. From Table 7.43, we can see that in general valuations of attributes gradually increase with age, but the variation of monetary value is little across age groups for fare payment variables. In the segments of aged 26-35, 36-45 and over 46, most monetary values are above the valuations without segmentation, except three attributes ('boarding time savings', 'getting changes back if pay by cash' and 'PT service routes covered'). Compared with the valuation of attributes without segmentation, among these four age groups respondents aged over 45 would have the highest willingness to pay for these attributes listed in Table 7.43. However, valuations of attributes for respondents aged 16-25 are all less than valuations without segmentation, indicating that compared other groups, the willingness to pay for better PT service and payment convenience in the group of 16-25 is lower than any other three groups of people. A possible reason is that younger people tend to be less well paid than other age groups, the effects attributed to age factor would stem from income.

Compared with attributes of 'multifunction', 'overdraft', 'geographic areas covered' and 'top-up/purchase methods', those three attributes ('boarding time savings', 'getting changes back if pay by cash' and 'PT service routes covered') show slightly different perceptions of respondents to these three attributes according to different age segments. From Table 7.43, we can see that two age groups (16-25 and 26-35) have a lower monetary valuation than the relevant overall valuations of attributes.

Table 7.43 Segmentation Results by Age Factor: Value of Attributes

Value of Attributes	All	Aged 16-25	Aged 26-35	Age 36-45	Over 45
VOBTS	4.45 (9.84)	3.96 (8.23)	4.31 (9.34)	4.59 (9.66)	4.72 (8.33)
Getting change back if paying cash	1.37 (2.1)	1.16 (1.9)	1.32 (2.1)	1.48 (2.0)	1.51 (2.0)
PT service routes covered:					
TC: Servicetc1:	6.30 (4.3)	(4.1)	6.28 (4.6)	6.39 (4.2)	6.42 (4.2)
Servicetc2	6.91 (4.87)	6.78 (3.5)	6.79 (3.9)	6.98 (4.9)	7.02 (4.7)
Servicetc3	7.84 (5.21)	7.7 (4.6)	7.83 (5.0)	7.93 (5.1)	7.99 (4.8)
SC: Servicesc1	1.49 (1.87)	1.33 (1.8)	1.45 (1.9)	1.53 (1.9)	1.58 (2.0)
Servicesc2	2.55 (2.04)	2.36 (2.0)	2.49 (2.1)	2.67 (2.0)	2.73 (2.0)
Servicesc3	4.81 (2.98)	4.67 (2.8)	4.76 (2.9)	4.87 (2.7)	4.94 (2.9)
Multifunction: Mf1:	2.32 (2.23)	2.22 (2.2)	2.36 (1.9)	2.41 (2.0)	2.29 (2.0)
Mf2	6.37 (5.14)	6.12 (4.7)	6.43 (5.0)	6.55 (5.5)	6.28 (4.9)
Mf3	8.9 (6.99)	8.83 (6.3)	8.93 (6.8)	9.02 (6.2)	8.73 (6.0)
Overdraft	4.16 (3.69)	3.89 (3.0)	4.23 (3.3)	4.29 (3.9)	4.41 (3.5)
Geographic areas covered: Geo1:	3.42 (3.11)	3.32 (3.0)	3.55 (3.3)	3.47 (3.4)	3.51 (2.9)
Geo2	5.9 (4.35)	5.86 (4.3)	6.15 (4.5)	5.98 (4.2)	5.95 (4.1)
Geo3	8.05(5.18)	7.93 (5.2)	8.22 (5.0)	8.17 (5.1)	8.09 (5.0)
Top-up/purchase methods:					
TC: Topuptc1:	---	--	--	--	--
Topuptc2	7.18 (3.2)	6.87 (3.2)	7.21 (3.3)	7.27 (3.0)	7.33 (2.9)
SC: Topupsc1:	3.31 (2.1)	3.18 (2.0)	3.36 (1.9)	3.39 (2.0)	2.48 (2.0)
Topupsc2:	6.55 (3.03)	6.43 (2.8)	6.63 (3.1)	6.69 (3.0)	6.73 (2.9)
Topupsc3	13.18(4.88)	12.52 (4.4)	13.29 (4.3)	13.35 (4.1)	13.68 (4.2)

Table 7.44 Smart Card Demand Forecast by Age Factor

	All	Aged 16-25	Aged 26-35	Age 36-45	Over 45
Scenario 1	78.2%	68.7%	73.2%	78.8%	79.6%
Scenario 2	57.6%	52.1%	55.6%	58.2%	60.9%

Table 7.44 lists the market share of smart card demand for the four age segments. It can be seen that with the increase of age the predicted share of smart card use gradually increases in the two scenarios, corresponding to the valuation of attributes for the four age segments. This result indicates that the older age group more would like to choose smart cards than the younger group. Moreover, the shares of aged 16-25 and aged 26-35 are below

the predicted share of the non-segmentation analysis, while the shares of aged 36-45 and over 45 are all over the overall prediction.

7.8.2 Sex

Table 7.45 Segmentation Results by Sex Factor: Value of Attributes

Value of Attributes	All	Male	Female
VOBTS	4.45 (9.84)	4.66 (10.02)	3.87 (9.23)
Getting change back if paying cash	1.37 (2.1)	1.55 (2.1)	1.12 (2.0)
PT service routes covered:			
TC: Servicetc1:	6.30 (4.3)	6.52 (4.4)	6.11 (4.1)
Servicetc2	6.91 (4.87)	7.13 (4.9)	6.78 (3.5)
Servicetc3	7.84 (5.21)	7.99 (5.3)	7.7 (4.9)
SC: Servicesc1	1.49 (1.87)	1.63 (1.9)	1.31 (1.8)
Servicesc2	2.55 (2.04)	2.87 (2.0)	2.35 (2.0)
Servicesc3	4.81 (2.98)	4.98 (2.9)	4.57 (2.8)
Multifunction: Mf1:	2.32 (2.23)	2.51 (2.1)	2.2 (2.1)
Mf2:	6.37 (5.14)	6.65 (5.2)	6.11 (4.6)
Mf3:	8.9 (6.99)	9.22 (6.1)	8.73 (6.0)
Overdraft	4.16 (3.69)	4.39 (3.7)	3.78 (3.0)
Geographic areas covered: Geo1:	3.42 (3.11)	3.58 (3.2)	3.21 (3.0)
Geo2:	5.9 (4.35)	6.2 (4.2)	5.81 (4.3)
Geo3:	8.05(5.18)	8.21 (5.0)	7.83 (5.2)
Top-up/purchase methods:			
TC: Topuptc1:	---	--	--
Topuptc2	7.18 (3.2)	7.33 (3.0)	6.82 (3.2)
SC: Topupsc1:	3.31 (2.1)	3.48 (2.0)	3.08 (2.0)
Topupsc2:	6.55 (3.03)	6.79 (2.9)	6.4 (2.8)
Topupsc3	13.18 (4.88)	13.45 (4.6)	12.62 (4.9)

The segmentation analysis for sex factor is based on two separate models for males and females. The effect of gender factor also presents an obvious difference between male and female respondents when comparing valuation of attribute with each other. From Table 7.45, we observe that generally values of attributes for male respondents are greater than the monetary valuations without segmentation, while valuations for female respondents are all less than these overall valuations. Therefore, for getting better PT services and payment convenience (such as quicker boarding time, wider areas smart cards can cover, various top-up/purchase methods, *etc.*), male people would like to pay more than female. The same reason as discussed for age factor could be applied to explain such heterogeneity of choices between male and female due to the differentiation of income.

Table 7.46 Smart Card Demand Forecast by Sex Factor

	All	Male	Female
Scenario 1	78.2%	81.2%	73.2%
Scenario 2	57.6%	62.5%	49.8%

Similar with the valuation of attribute, the heterogeneity of user preference across sex groups also can be reflected from the forecast market share of smart cards in Table 7.46. In the two scenarios, the share of smart card use of males is greater than that of females (about 81.2% and 62.5%, respectively).

7.8.3 Household Income

Table 7.47 Segmentation Results by Household Income Factor: Value of Attributes

Value of Attributes	All	Less than ¥1500	¥1500-¥2999	¥3000-¥3999	>¥4000
VOBTS	4.45 (9.84)	3.78 (8.63)	4.38 (9.4)	4.61 (9.71)	4.8 (9.03)
Getting change back if paying cash	1.37 (2.1)	1.15 (1.9)	1.36 (2.0)	1.46 (2.0)	1.52 (2.0)
PT service routes covered:					
TC: Servicetc1:	6.30 (4.3)	6.15 (4.0)	6.33 (4.6)	6.4 (4.2)	6.5 (4.2)
Servicetc2	6.91 (4.87)	6.68 (3.5)	6.88 (3.8)	7.04 (4.9)	7.22 (4.9)
Servicetc3	7.84 (5.21)	7.62 (4.6)	7.83 (5.1)	7.96 (5.1)	8.2 (4.9)
SC: Servicesc1	1.49 (1.87)	1.29 (1.8)	1.42 (1.9)	1.59 (1.9)	1.7 (2.0)
Servicesc2	2.55 (2.04)	2.36 (2.0)	2.48 (2.0)	2.77 (2.2)	2.89 (2.0)
Servicesc3	4.81 (2.98)	4.57 (2.3)	4.78 (3.0)	4.96 (2.5)	5.11 (2.9)
Multifunction: Mf1:	2.32 (2.23)	2.18 (2.3)	2.39 (1.9)	2.45 (2.0)	2.56 (2.2)
Mf2	6.37 (5.14)	6.1 (4.9)	6.46 (5.1)	6.56 (5.5)	6.67 (4.9)
Mf3	8.9 (6.99)	8.63 (6.8)	8.99 (6.8)	9.15 (6.6)	9.26 (6.1)
Overdraft	4.16 (3.69)	3.8 (3.6)	4.26 (3.3)	4.39 (3.6)	4.52 (3.6)
Geographic areas covered: Geo1:	3.42 (3.11)	3.22 (3.0)	3.48 (3.3)	3.58 (3.2)	3.7 (3.0)
Geo2	5.9 (4.35)	5.66 (4.3)	6.04 (3.9)	6.18 (4.1)	6.23 (4.0)
Geo3	8.05(5.18)	7.73 (5.1)	8.12 (5.0)	8.27 (4.6)	8.49 (4.8)
Top-up/purchase methods:					
TC: Topuptc1:	---	--	--	--	--
Topuptc2	7.18 (3.2)	6.77 (3.2)	7.25 (3.3)	7.39 (3.1)	7.46 (3.0)
SC: Topupsc1:	3.31 (2.1)	3.18 (2.0)	3.38 (1.9)	3.46 (2.1)	3.57 (2.1)
Topupsc2:	6.55 (3.03)	6.31 (2.99)	6.65 (3.0)	6.72 (3.0)	6.85 (2.8)
Topupsc3	13.18(4.88)	12.62 (4.3)	13.26 (4.1)	13.33 (4.3)	13.65 (4.2)

The segmentation results for household income are based on four different income groups. From Table 7.47, we can see a strong influence of income factor on the valuations of attributes when splitting respondents by different household income levels. Similar with results analysed for age factors previously, valuations of attribute for household income segmentation indicate that with the increase of household income, the valuations of attributes also accordingly increase as expected. It can be implied that those with higher income would like to pay more for these better services, boarding time savings, and so forth. Meanwhile, it is worth noting that generally the largest increase of monetary valuations between two close-by groups is in the movement from the segment of less than 1500yuan per month to the segment of 1500-2999yuan per month, which indicates that respondents' willingness to pay tends to change obviously from the group of less 1500yuan to 1500-2999yuan.

Table 7.48 Smart Card Demand Forecast by Household Income Factor

	All	Less than ¥1500	¥1500-¥2999	¥3000-¥3999	>¥4000
Scenario 1	78.2%	66.5%	72.1%	78.3%	82.9%
Scenario 2	57.6%	48.9%	53.6%	58.1%	61.6%

Table 7.48 presents forecast results of smart cards for four household income segments. The heterogeneity of user preference due to different income backgrounds does exist across the four income segments. It is not surprising that with the increase of household income, the share of smart card also increases. Therefore it can be implied that users with high household income would like to pay more for the better payment services.

To sum up, from the segmentation analysis by three socio-economic variables, we can conclude as follows:

- For age factor, the valuations of attributes increase with age, and the younger respondents have the lowest willingness to pay in these four age groups. Most perceptions of respondents to payment attributes tend to change from the group of 16-25 to 26-35, compared with the overall monetary valuations without segmenting. However, it should be noted that the variation across different age groups is little, although some changes do exist. Through comparing the forecast result of market share of smart cards, we can see that with the increase of age, the use of smart cards also would go up.
- Sex factor also has the strong influence on the values of attributes. Male respondents would like to pay more for better PT services and payment convenience.
- Household income can also explain respondents' preferences in different income groups. People in the lowest income group would have the lowest willingness to pay. Meanwhile, it can be seen that respondents' choice behaviour would begin to change obviously between the group of less than 1500yuan and 1500-2999yuan. On the other hand, due to some correlation existing between age and income, sex and income, though such correlation is not very high, it still could partly explain the heterogeneity of choices between the young group and other age groups, male and female groups.

7.9. Principal Findings and Conclusions

The evaluation study was carried out in Chapter 7 by analysing the RP and SP data with the discrete choice models. The principal findings and conclusions in this chapter can be summarised as follows:

- First of all, this research focused on PT users' demand analysis to have an insight into respondents' perceptions toward smart cards as well as cash and travel cards, so as to reveal benefits and effectiveness of smart cards for public transport. The forecast market shares for three payment methods in the joint RP/SP model present changes of the user demand when factors of smart cards change (single-factor and multiple-factor changes). From the demand analysis, it can be found that respondents showed the strong preference to use smart card payment when smart card variable(s) is (are) at good level(s), particularly for smart card features, such as multifunction, geographic areas covered, top-up methods. From these scenarios, it can be concluded that with the payment attribute(s) of smart cards becoming better, the share of smart cards would increase and be dominantly more than other two payment methods.

- Secondly, preference data were measured in this chapter to obtain respondents' willingness to pay and how much different features of the smart card payment would benefit users in terms of monetary valuation. Values of attributes were divided into two parts: valuation of boarding time savings and valuation of other attributes. VOBTSs in three different models (pure RP, pure SP and joint RP/SP) reveal that respondents' willingness to pay ranged between 3.6yuan/month (pure RP) and 6.06yuan/month (pure SP) to save their boarding time. The higher VOBTS of the pure SP data indicates that respondents would like to pay more for quicker boarding time to save their whole journey time. In addition to VOBTSs, valuations of other attributes (qualitative) are obtained to examine respondents' willingness to pay for different payment services. It can be seen that with payment attributes becoming better, respondents would like to pay more for using this payment alternative. Among these qualitative attributes, the most perceived attributes by respondents would be 'top-up/purchase methods', 'multifunction', 'geographic areas covered' of smart cards and 'PT service routes covered of travel cards', because categories of these attributes have relatively high monetary valuations, compared to others.
- Thirdly, benefits of smart cards were also discussed by the importance of attributes. The importance of attributes was obtained by analysing the sign and size of estimated coefficients in the logit models. Through the importance of attributes, it can be seen that services covered by travel cards, multifunction of smart cards, geographic areas can be covered by smart cards and top-up/purchase methods of smart cards are the most important factors when respondents choose their fare payment methods. In the mean time, the importance of attribute can feed back to the relevant policy making to enhance the smart card service: the priority of the smart card development should be given to these factors, because these important attributes would have more benefits to PT users than other factors of the smart card payment. For example, the current smart card ticketing in Dalian can only cover the PT services in the urban area. One of future development directions of smart cards would focus on a wider area the smart card scheme can cover.

- In addition, through fare elasticities, the benefits of smart cards were explained by travel cost exclusively, because travel cost may be a primary factor when PT users choose different payment option. In general, all own elasticities of three payment alternatives are greater than 1.0 in three models, indicating that payment demand with respect to changes of their own travel cost is very sensitive. All cross elasticities of cash with respect to travel cost of other two card payment methods are greater than 1.0. It can be implied that the cash user group is very sensitive and readily affected by changes of travel cost of other fare payment methods. However, cross elasticities of smart cards with respect to change of cash cost in three models are all less than 1.0, indicating that relative to change of travel cost of cash, the demand of smart cards tends to be stable. In the pure SP and joint RP/SP models, all cross elasticities between two card payment methods are greater than 1.0. Therefore such high substitutability between travel cards and smart cards can be clearly seen, while in the pure RP model, respondents' actual choice behaviours look more stable between these two card payment methods, because the cross elasticities between two card payment methods are less than 1.0. Therefore, according to the elasticity analysis, it can be seen that changes of travel cost would primarily and directly influence respondents' choice behaviour of fare payment options. So the discounted fare policy in current smart card schemes is one of effective solution to increase the use of smart card payment.
- Finally, the segmentation analysis has an insight into the effect of socio-economic variables on passengers' payment choices. It is helpful to examine whether benefits of a smart card scheme would be the same to all groups of respondents and how such difference would be if any. Three socio-economic variables were used to segment: age, sex and household income. Generally, analysis results reflect the following relationship between respondents' choices and their socio-economic backgrounds:
    - For age factor, the valuations of attributes increase with age, and the younger respondents (16-25 years) have the lowest willingness to pay in these four age groups. Therefore, the relevant fare policy and various smart card products should be introduced to this group of people, such as student smart cards.
    - Sex factor also has the strong influence on the value of attributes. Male respondents would like to pay more for better PT services and payment convenience.
    - People in the lowest income group (less than 1500yuan) would have the lowest willingness to pay. Meanwhile, it can be seen that respondents' choice behaviour would begin to change obviously between the group of less than 1500yuan and 1500-2999yuan. On the other hand, due to the correlation existing between age and income, sex and income to some extent, it could partly explain the

heterogeneity of choices between different age groups, between two sex groups.

In this chapter, the standard logit models are used to analyse choice behaviour and predict demand of payment methods. But there are two main concerns could be taken into account further. The first one is the uncertainty of stated choice response. How certain the attribute-level can vary based on a certain possibility being happen and how certain the individual choices could be, need to be taken into account when modelling human's decision making. Secondly, the standard logit models are based on linear additive utility models. Non-linearity of human's decision making is another potential concern that could influence on demand forecasting ability with discrete choice data. Therefore, in the next chapter, new techniques (FL-fuzzy logic and ANN-artificial neural network methods) for analysing discrete choice data will be explored to identify these two effects by comparing FL, ANN models with logit models and to improve the forecasting ability.

Chapter 8
Modelling Preference Data with Fuzzy Logic and Neural Network Technique

8.1. Introduction

In the previous chapter, data analysis by logit models has been discussed, which includes the pure RP, pure SP and joint RP and SP models. These models provide a framework to explore the trade-off between the attributes of the various alternatives, each of which is associated with a utility. But the principal issue in logit models is that they have the appeal of being stochastic and yet admitting decision variables. That is to say, random utility models assume explicit relations for both the systematic utility and the distribution function of the random utility component present. The systematic utility is determined from the attributes of the fare payment alternative based on the concept of trade-off. However, it would be possible that such explicit relations could not realistically describe human decision mechanisms, nor that compensatory decision mechanisms are applicable in each choice situation (Lotan and Koutsopoulos, 1993a). Particularly when the demand forecast was carried out to identify respondents' behavioural changes based on different payment situations, the limitation of logit model may influence the model performance of the evaluation study. In fact, human decision processes are known to be highly non-linear, stochastic, partially compensatory, and partially lexicographic without any *a priori* assumptions (Hoogendoorn et al, 2000) (such as assumption of error term distribution type, relationship between inputs and outputs, *etc*).

In previous studies, two techniques, fuzzy logic and neural network methods have been widely used in transportation research, some of them in modelling discrete choice problems (such as mode choice, route choice, *etc*). The potential of the fuzzy logic (FL) technique for this research is its ability to simultaneously handle numerical data and linguistic knowledge so as to capture uncertainty and ambiguity of decision making (Zadeh, 1973). It is a nonlinear mapping of an input data (feature) vector into a scalar output, i.e., it maps numbers into numbers. Particularly in preference choice situations, the underlying assumption is that decision makers use a few simple rules that relate their vague perceptions of the various attributes to their preferences towards the available alternatives (Vythoulkas and Koutsopoulos, 2003).

Another technique being discussed in this chapter is the artificial neural network (ANN) method. In recent years, ANN method has received increasing attention on transportation studies, particularly on analysing discrete choice behaviour and forecasting market share based on individual preference data. The principle of ANN model is that it replicates human brain functions and is thus considered as 'intelligent', since it learns and generalises by a designed network structure to find the best fit mapping relationship between inputs and outputs (Reggiani et al., 1997). Therefore, to capture the non-linearity and allow the model for self-learning is the major motivation to employ ANN methods in this research.

Therefore, considering the potential issue of logit models and advantages of FL and ANN techniques discussed above, in this chapter FL and ANN techniques capable of modelling non-linearity and uncertainty of decision making based on the stochastic theory, are proposed as an alternative method of conventional logit models to analyse discrete choice data.

This chapter is organised as follows: first of all, fuzzy logic analysis is introduced in Section 8.2. This includes the determination of inputs and outputs of FL models, membership functions, fuzzy inference system, model calibration and estimation results. Following fuzzy logic analysis, Section 8.3 presents artificial neural network method for analysing discrete choice data, in which input and output, network structure, algorithm of training and learning process, and estimation results are discussed. In order to test the forecasting ability and model performance, comparisons between different models are made in Section 8.4, including standard logit models (discussed in Chapter 7), FL, and ANN models. Finally, Section 8.5 concludes findings through modelling preference data with FL and ANN techniques and compares these two techniques with logit models used in Chapter 7 to identify pros and cons of the three different models on forecasting ability and model performance.

8.2. Fuzzy Logic Analysis

The detailed methodology about fuzzy logic models have been introduced in Chapter 4: Research Methodology. In this section, the following aspects of fuzzy logic methods specific for this research context are discussed: inputs and outputs, membership functions, fuzzy inference system, model calibration and comparisons of estimation results.

8.2.1 Inputs and Outputs

Input variables of the fuzzy logic models are also based on those variables defined in the RP/SP survey, including attributes related to fare payment alternatives. The input data format in FL models is the same as the MNL models. Fare payment attributes and levels are the deterministic factors to influence respondents' decision-making. The output of FL models is the choice probabilities of choosing a given payment method, because we set the output value ranging between 0 and 1. '0' indicates that the relative utility of choosing this payment method is very low, while '1' means a very high probability of choosing the payment alternative. Having the same output measure (i.e., choice probabilities) can also make the FL models and logit models comparable, particularly on forecasting market shares.

In addition to the variables relevant to payment alternatives, in the FL models, the error term (containing those unobserved factors that can influence respondents' decision making) is regarded as another input. In the pure MNL models, we have assumed that the error term in the utility models is Gumbel distributed, and all estimation was based on this assumption to estimate coefficients in the utility model. However, in statistics, the error of a set of data could have any type of distribution (e.g., normal distribution, log-normal distribution, Gumbel distribution, *etc*) before a test is carried out to identify the error distribution type. Therefore, in order to examine effects of different distribution functions on estimating discrete choice data, in the fuzzy logic analysis, random values for some distribution types to represent the error term (unobserved factors), are regarded as one of modelling inputs and will be tested and compared with each other. Two types of error distributions—Gumbel distribution and normal distribution, are selected and compared in this research, because (1) Gumbel distribution is assumed and used in logit models and if the same distribution type can be taken into account in the FL models, the outputs from different two models can be comparable; (2) besides Gumbel distribution, it is necessary to test another distribution type in the FL models so that it can be examined whether different error distribution assumptions influence the forecasting ability of FL models.

So, comparisons are made among the following models in this section, including:

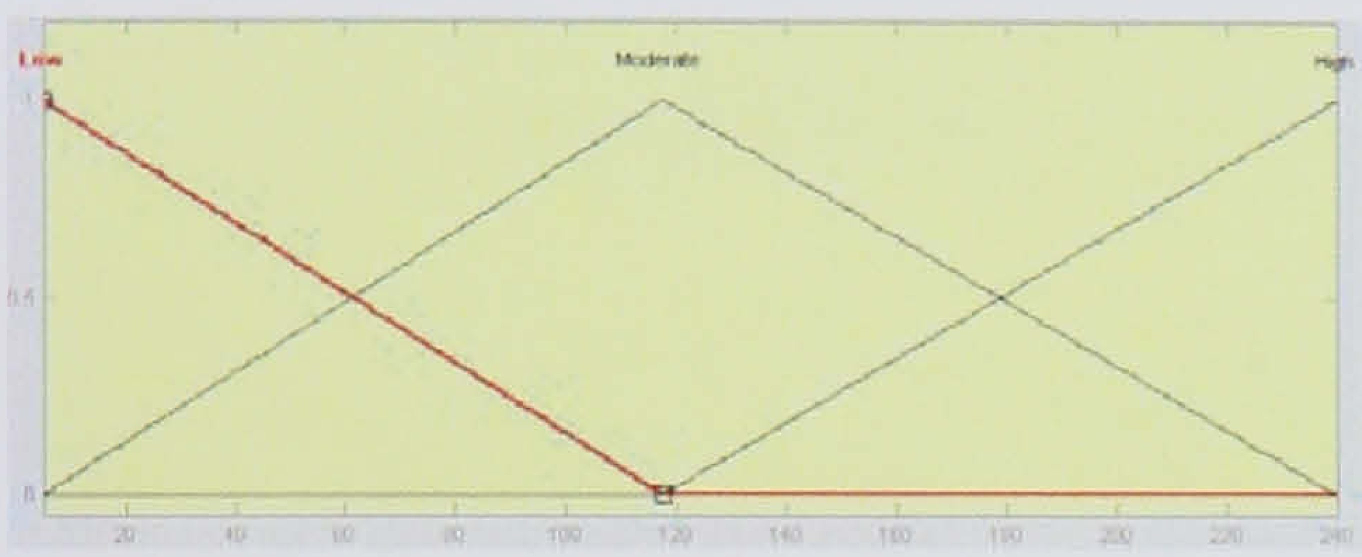
- FL model without any error term;
- FL model with Gumbel distribution error term;
- FL model with normal distribution error term.

8.2.2 Membership Functions

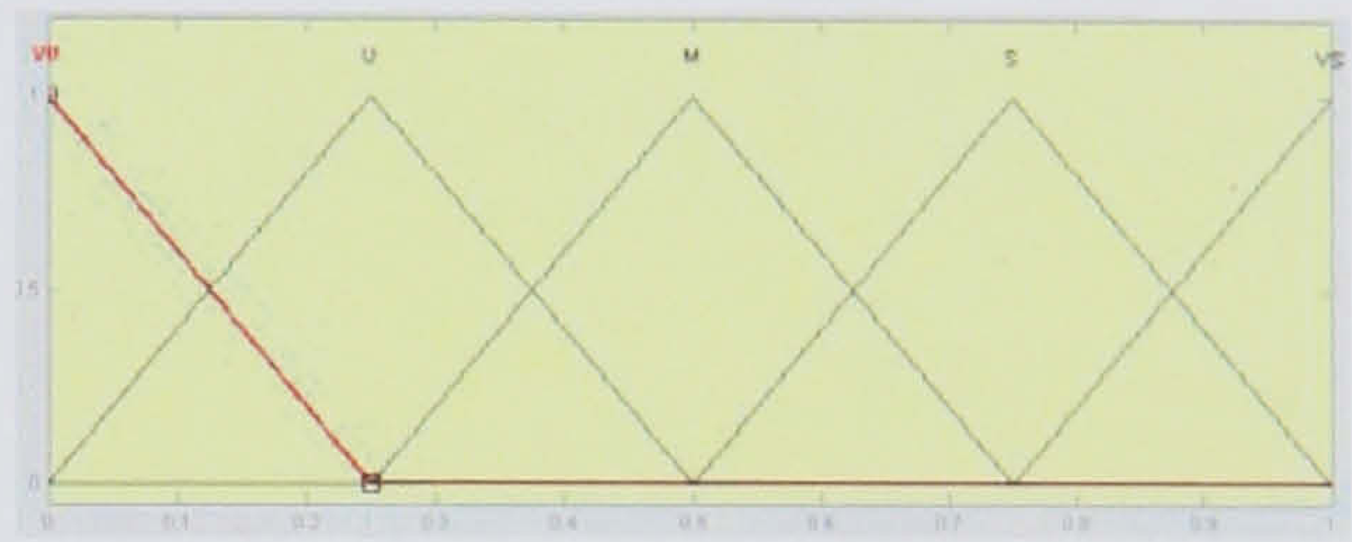
In the Fuzzy Logic Toolbox, MATLAB 6.5, some membership types are available to design membership function for each input variable. In order to simplify computation for this research, first of all, triangular membership type is used. The reason to use triangular membership in the FL models is because of its simplicity compared with other membership types, such as Gaussian, trapezoidal, *etc*. However, in the later model calibration, some membership types different from triangular type will be tested to fine tune the FL models.

The example of membership functions (before the model calibration: the base model) in the RP data set is presented in Figure 8.1-8.4. It can be seen that in Figure 8.1-8.4, membership functions for all variables, including inputs and outputs, are designed by triangular type. Among these figures, Figure 8.1-8.3 are related to input membership functions for three fare payment methods in the RP survey. Figure 8.4 is output membership functions. Meanwhile, in these membership functions, error of choice data distributed by Gumbel distribution type was considered as one of inputs in the FL model. Another kind of distribution type, normal distribution, is also taken into account as an alternative to test the effect of different error distribution types in the model estimation. For normal distribution error, the same membership function type as Gumbel distribution is used, but the different range from Gumbel distribution was allocated because of randomly generated error values.

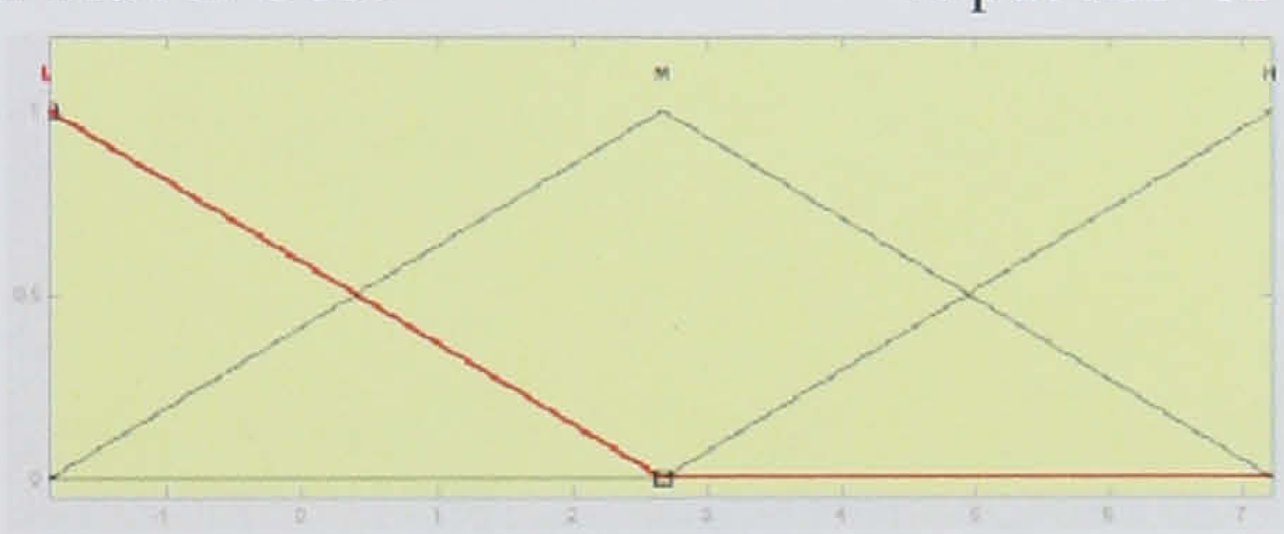
Through these membership functions, linguistic categories for each input variable can be obtained. The details about linguistic categories of variables are introduced in the following section relevant to fuzzy rules. For each output, three categories of probability of choosing a certain payment option are designed as: *Low, Moderate* and *High*.



Input MF of Travel Cost

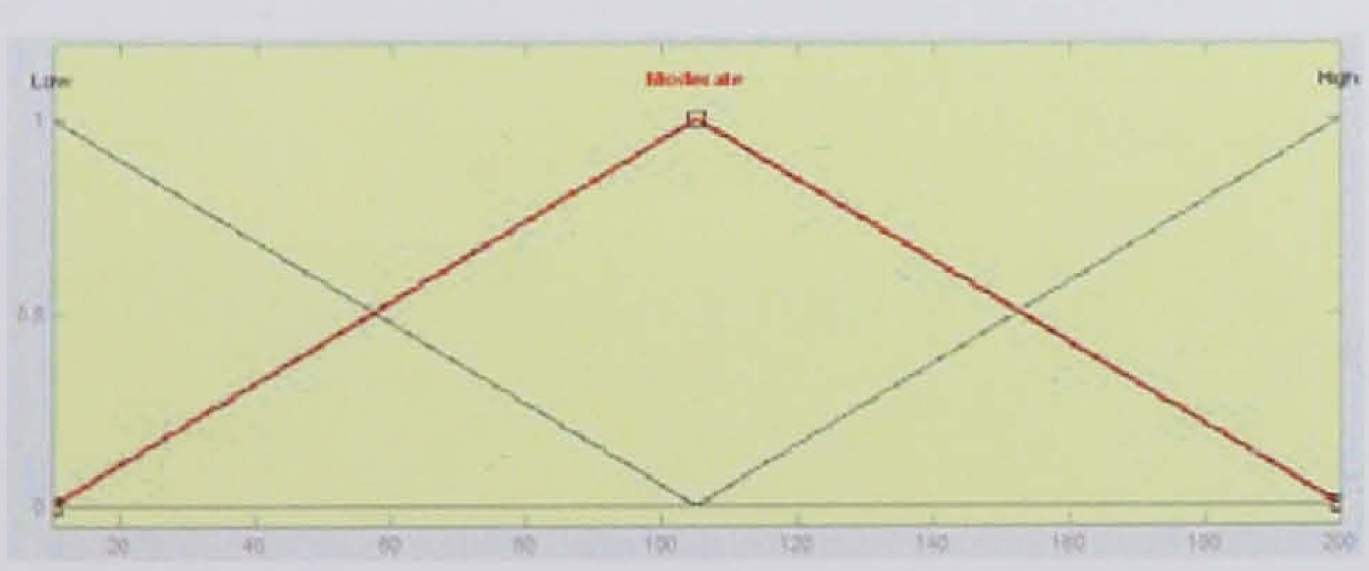


Input MF of Overall Assessment

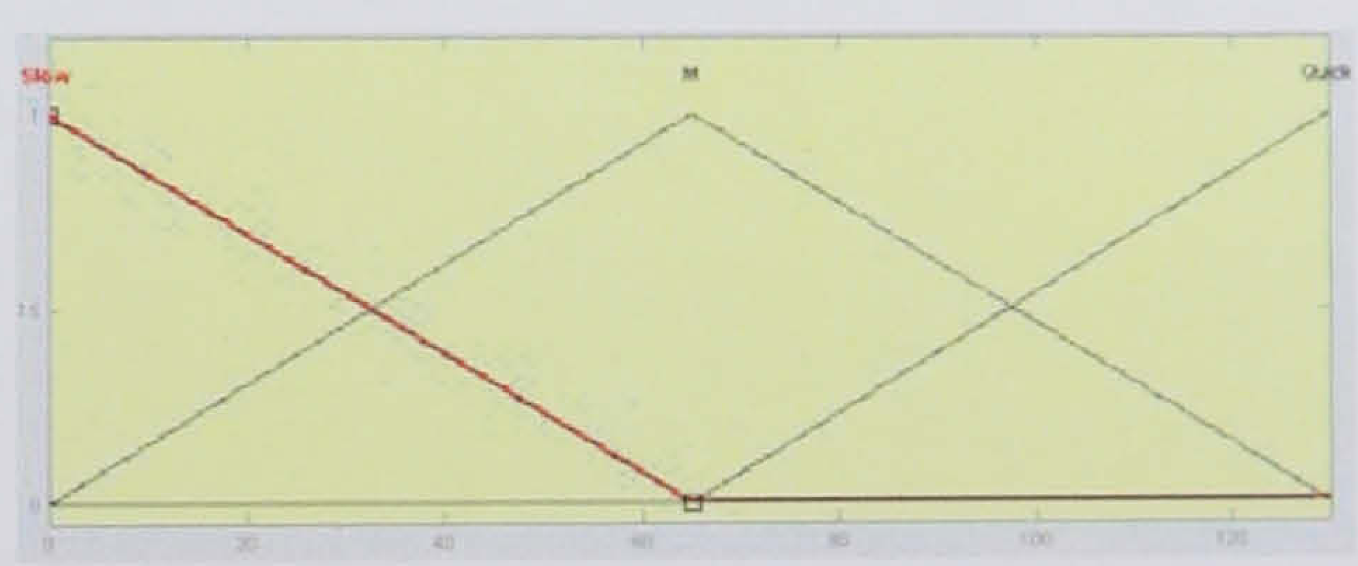


Input MF of Gumbel Error Distribution

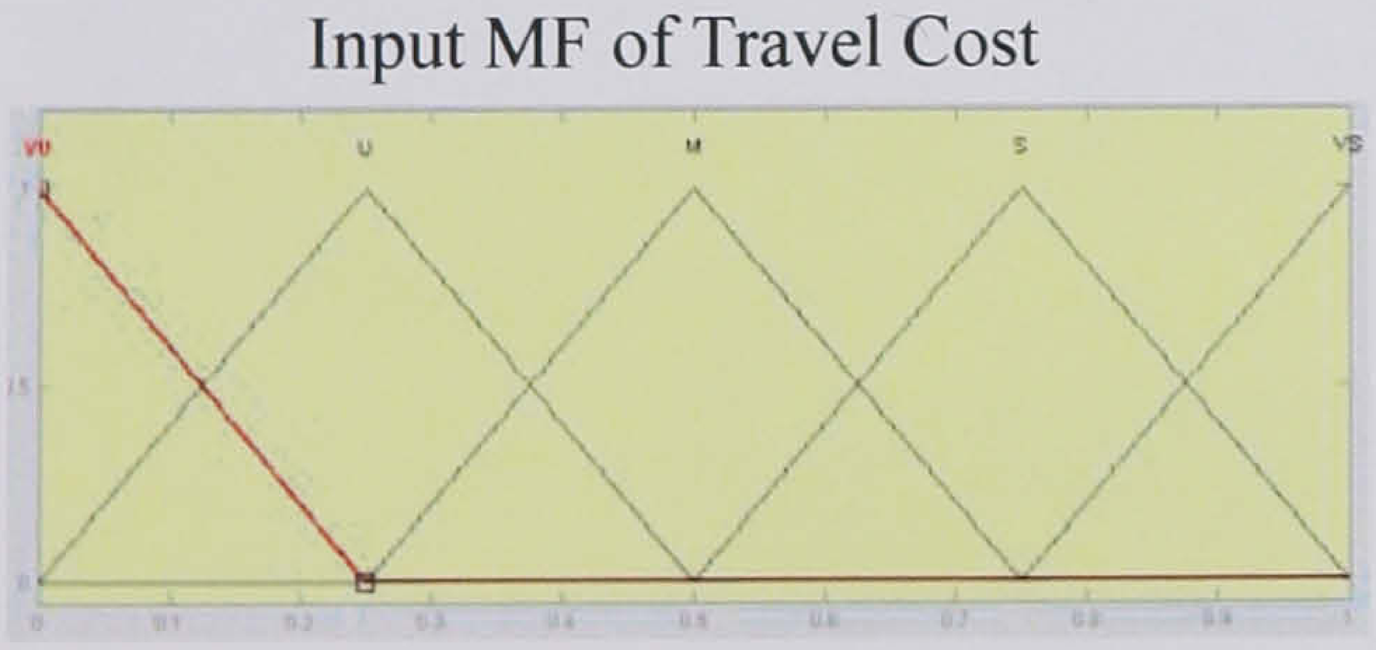
Figure 8.1 Input Membership Functions for Cash Fare Payment in RP



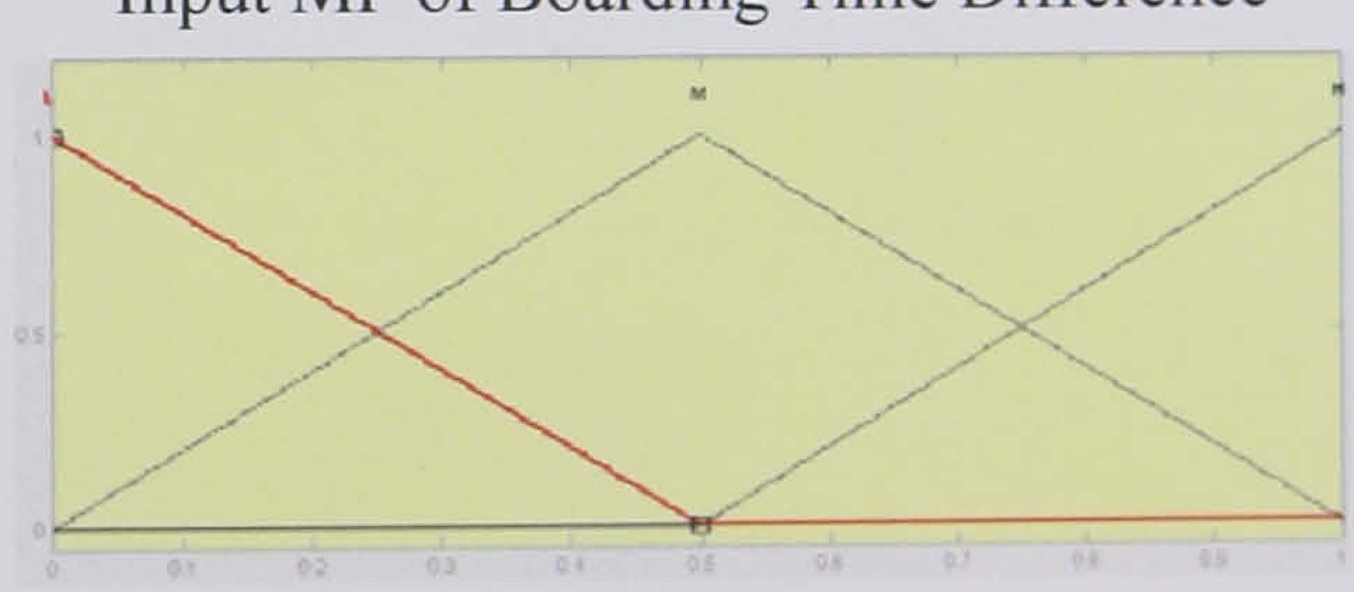
Input MF of Travel Cost



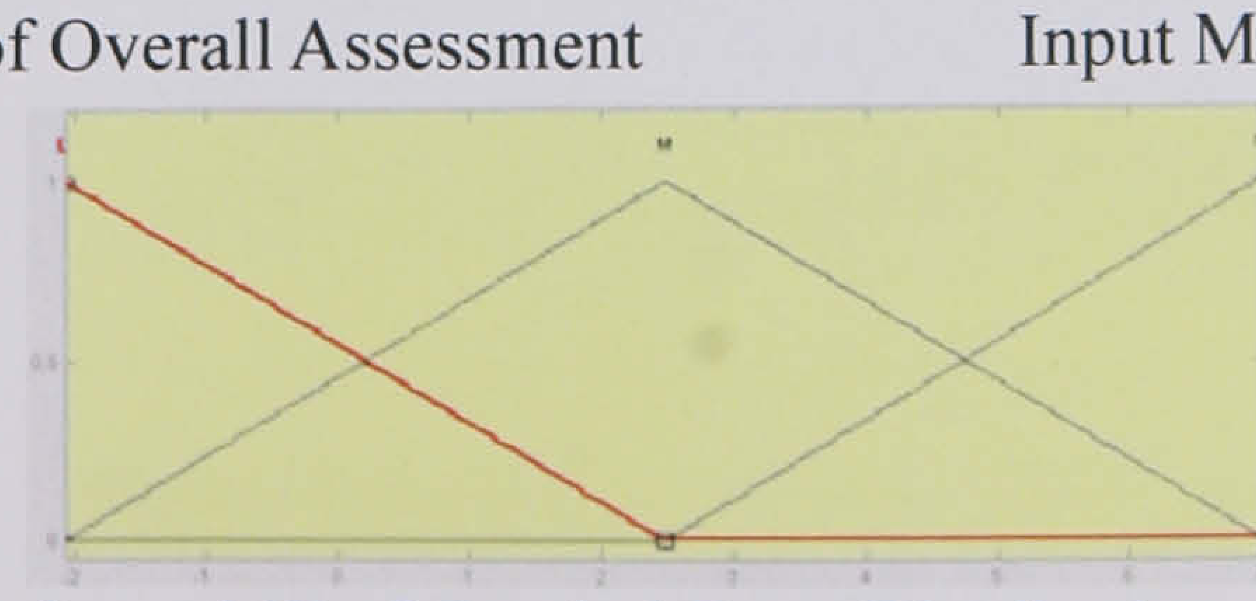
Input MF of Boarding Time Difference



Input MF of Overall Assessment

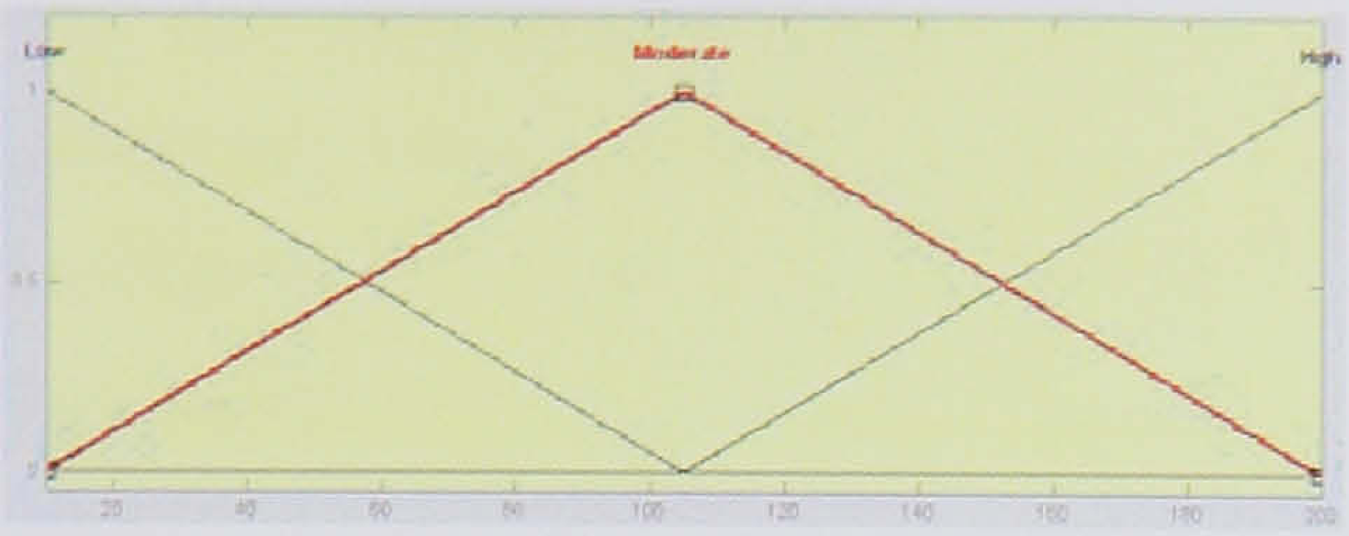


Input MF of Seat Availability

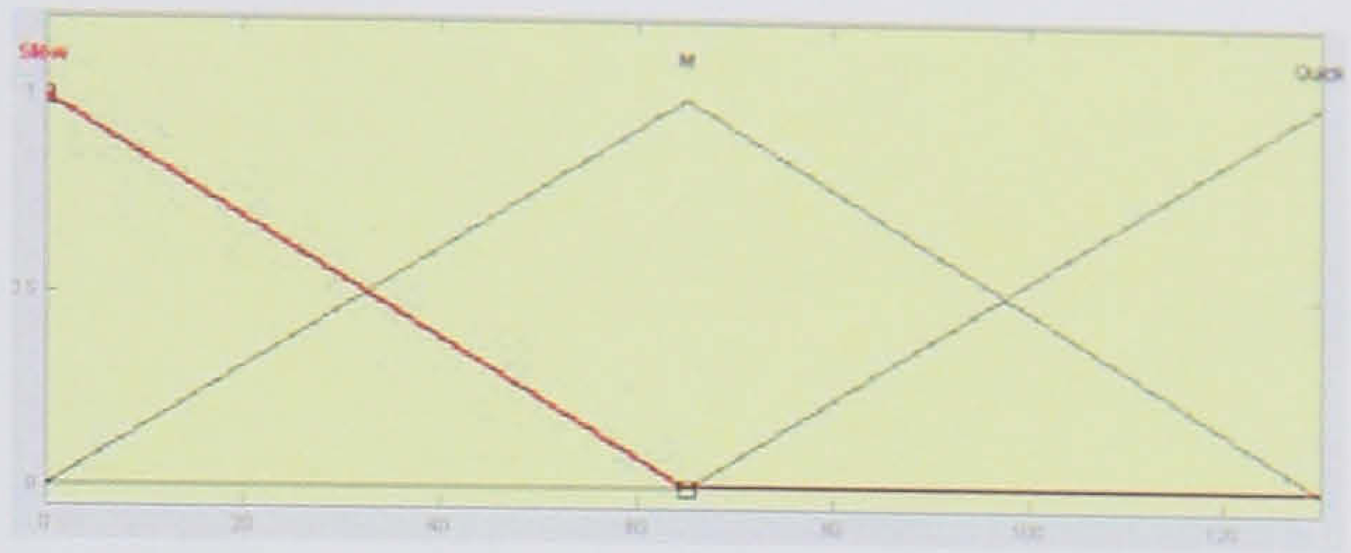


Input MF of Gumbel Error Distribution

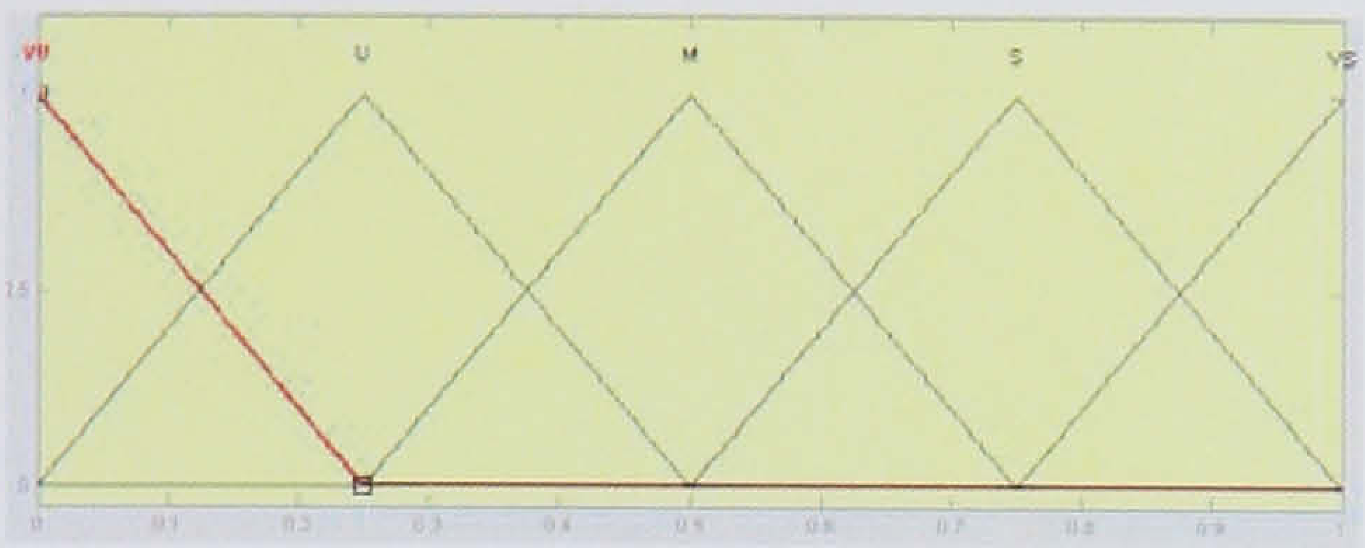
Figure 8.2 Input Membership Functions for Travel Cards in RP



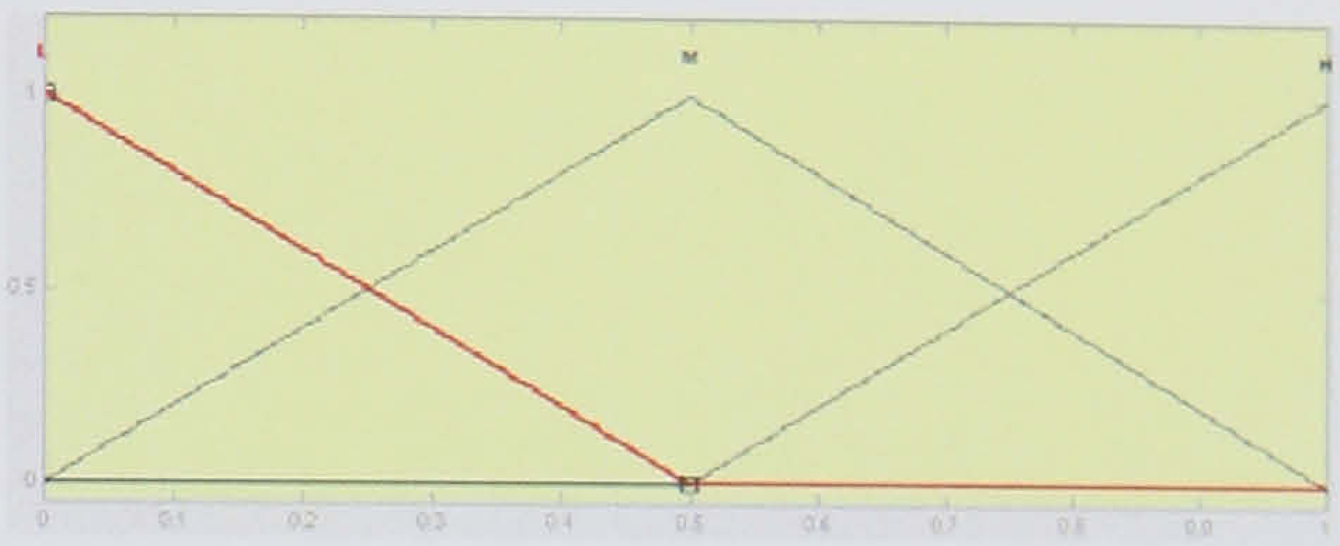
Input MF of Travel Cost



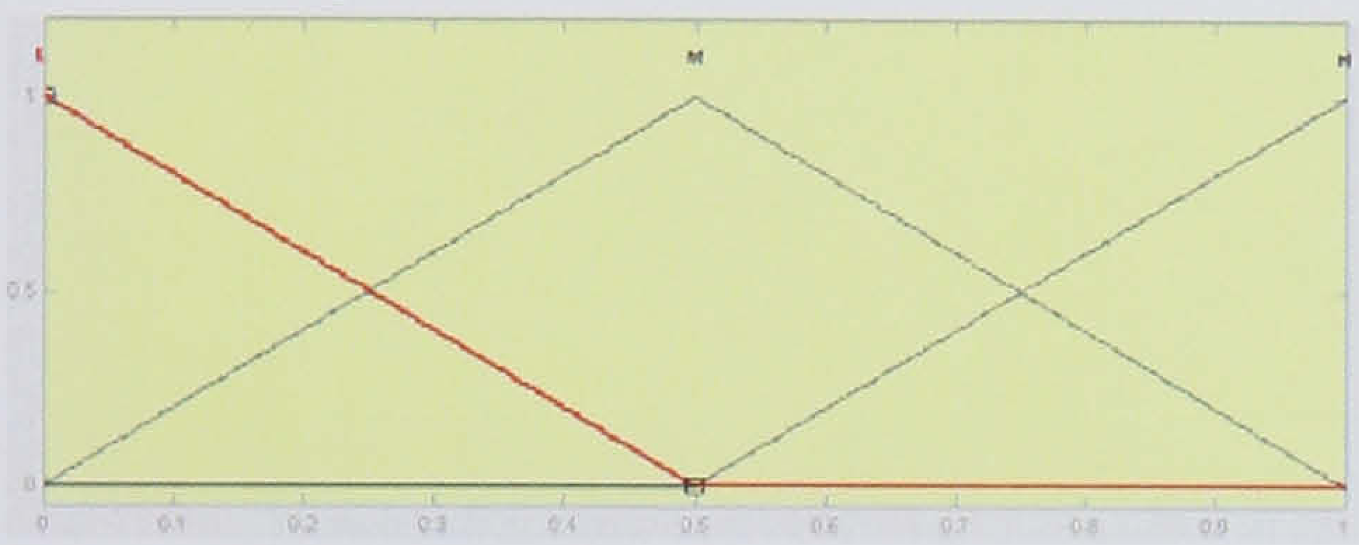
Input MF of Boarding Time Difference



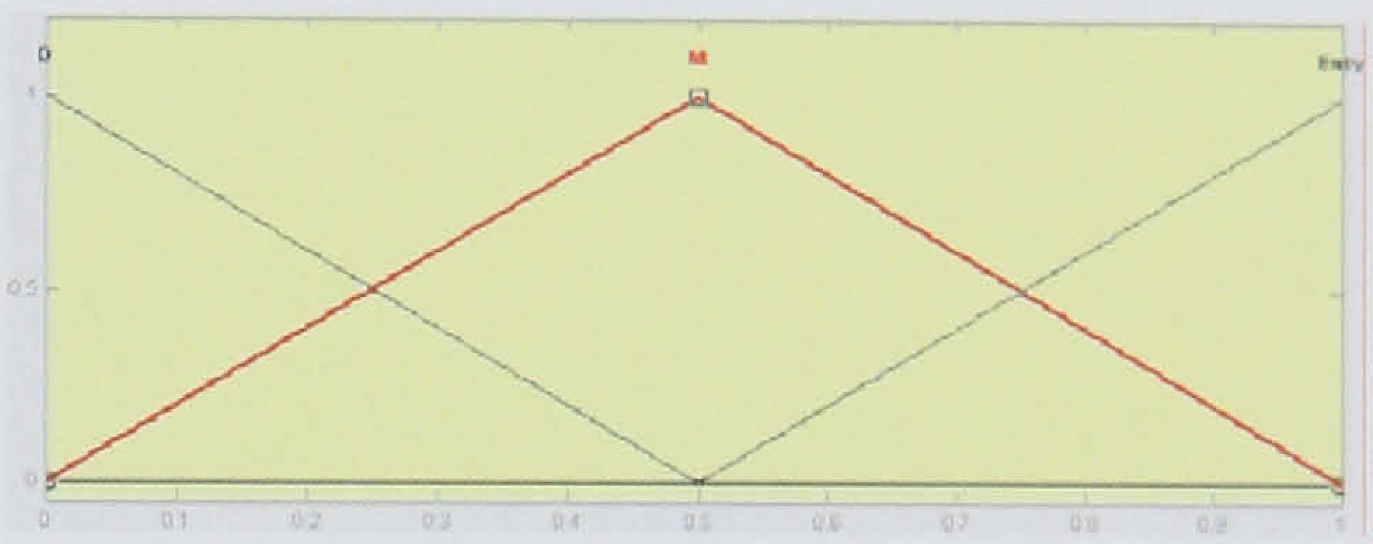
Input MF of Overall Assessment



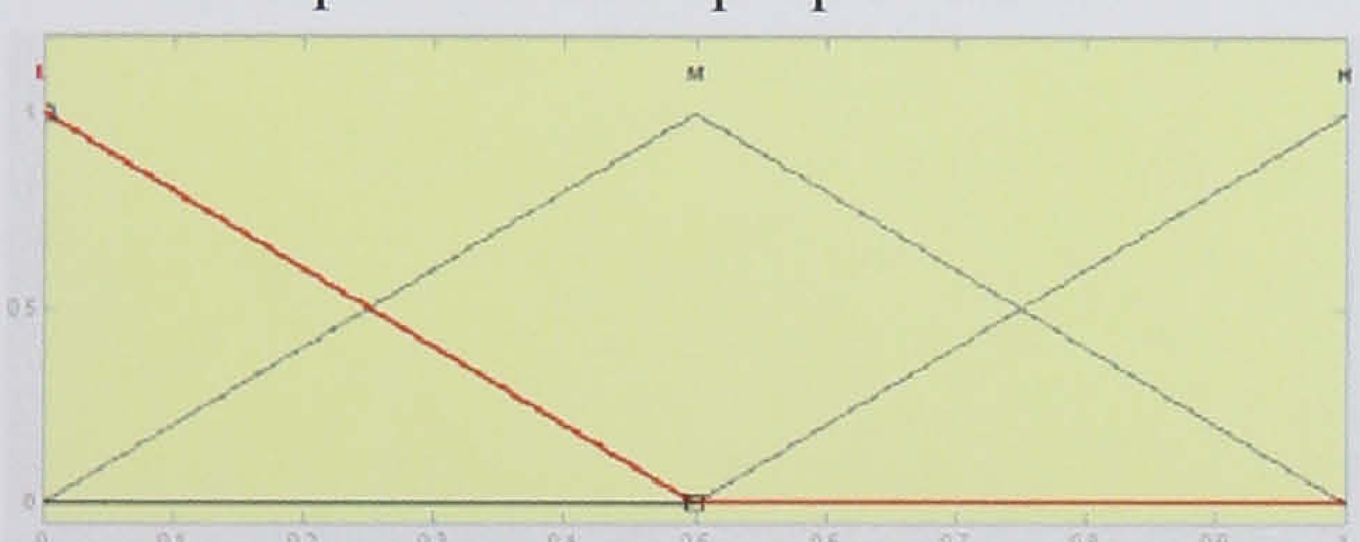
Input MF of Seat Availability



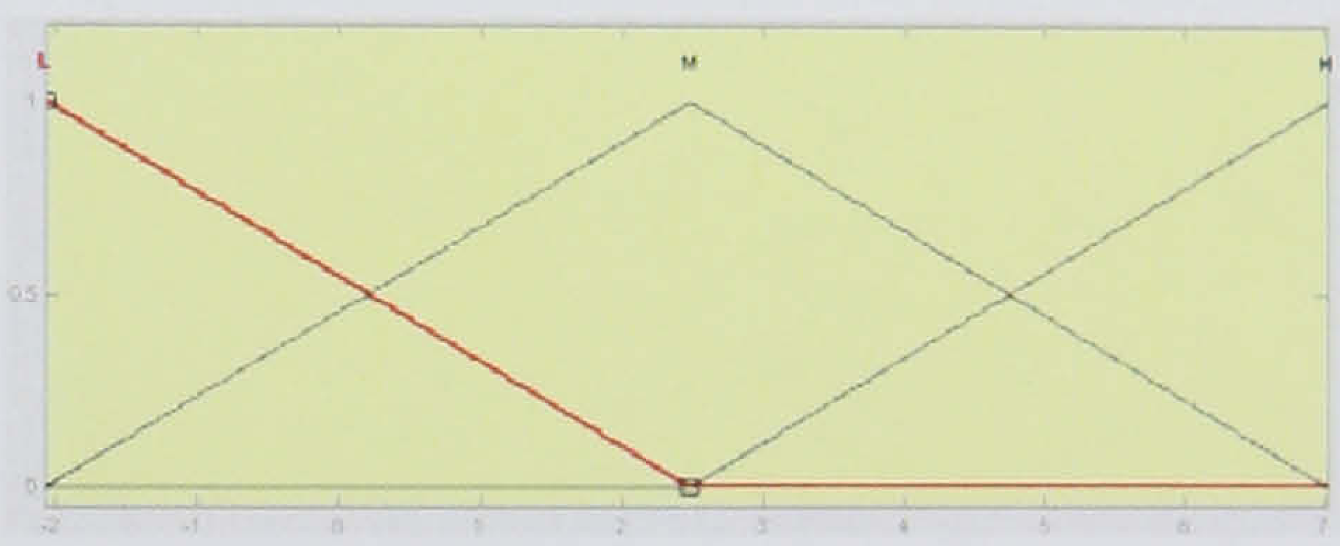
Input MF of Top-up Methods



Input MF of Difficulty of Top-up

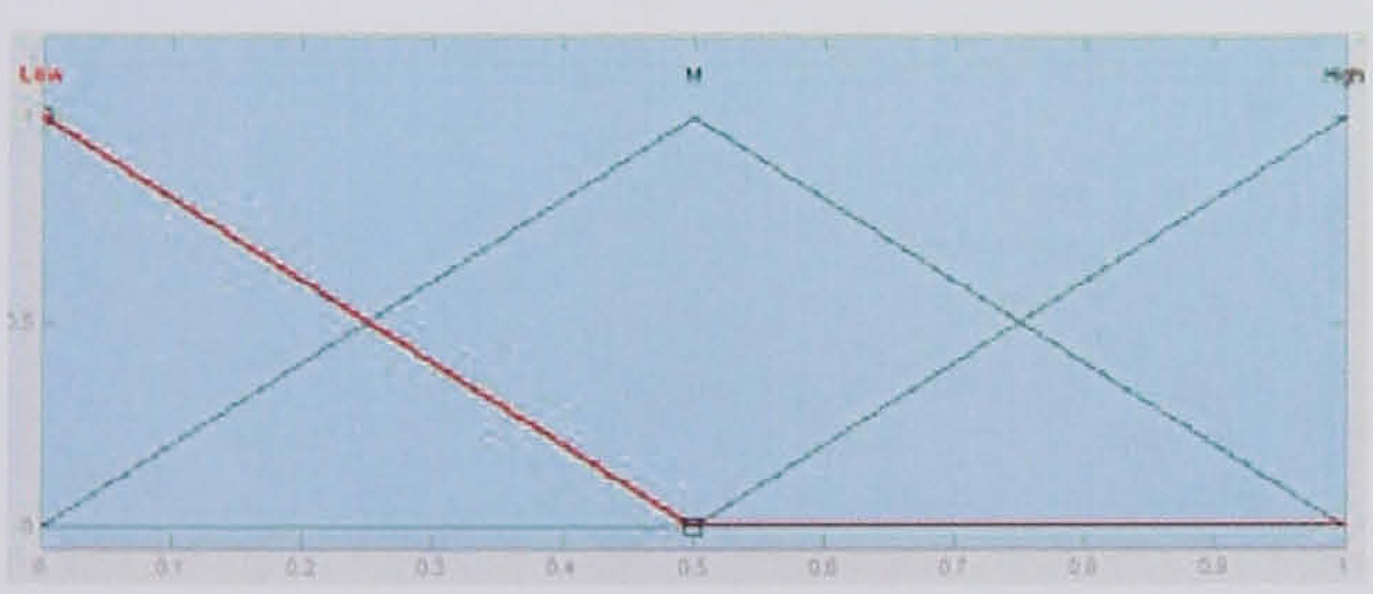


Input MF of Multifunction

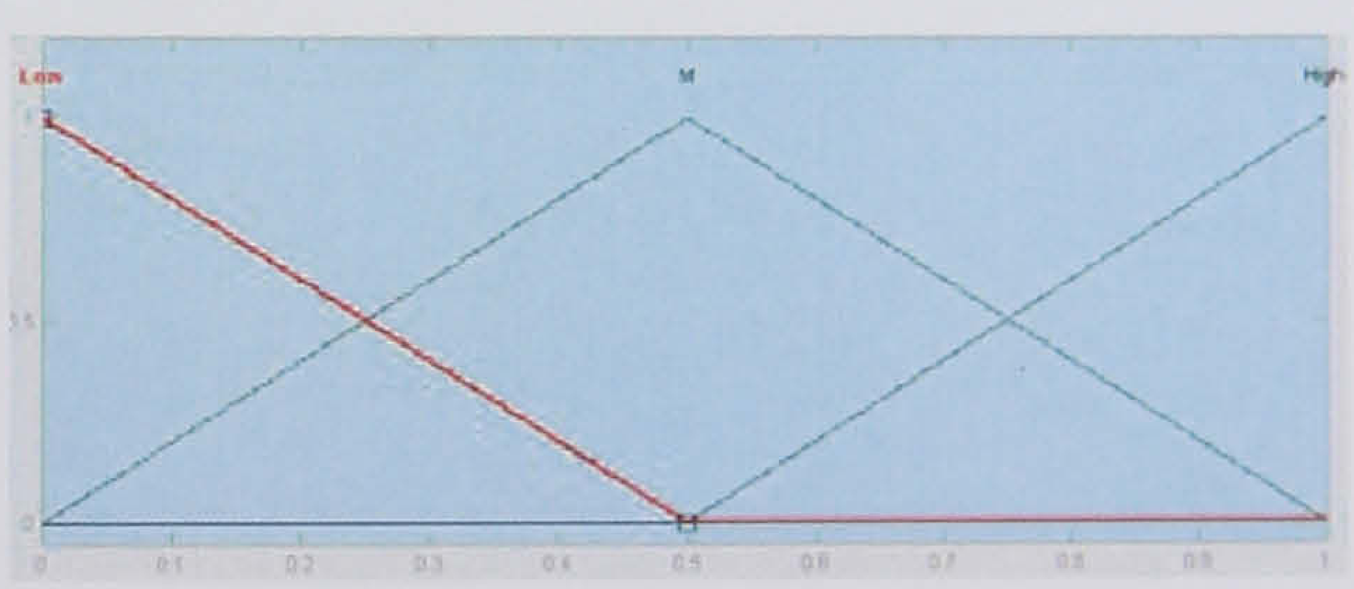


Input MF of Gumbel Error Distribution

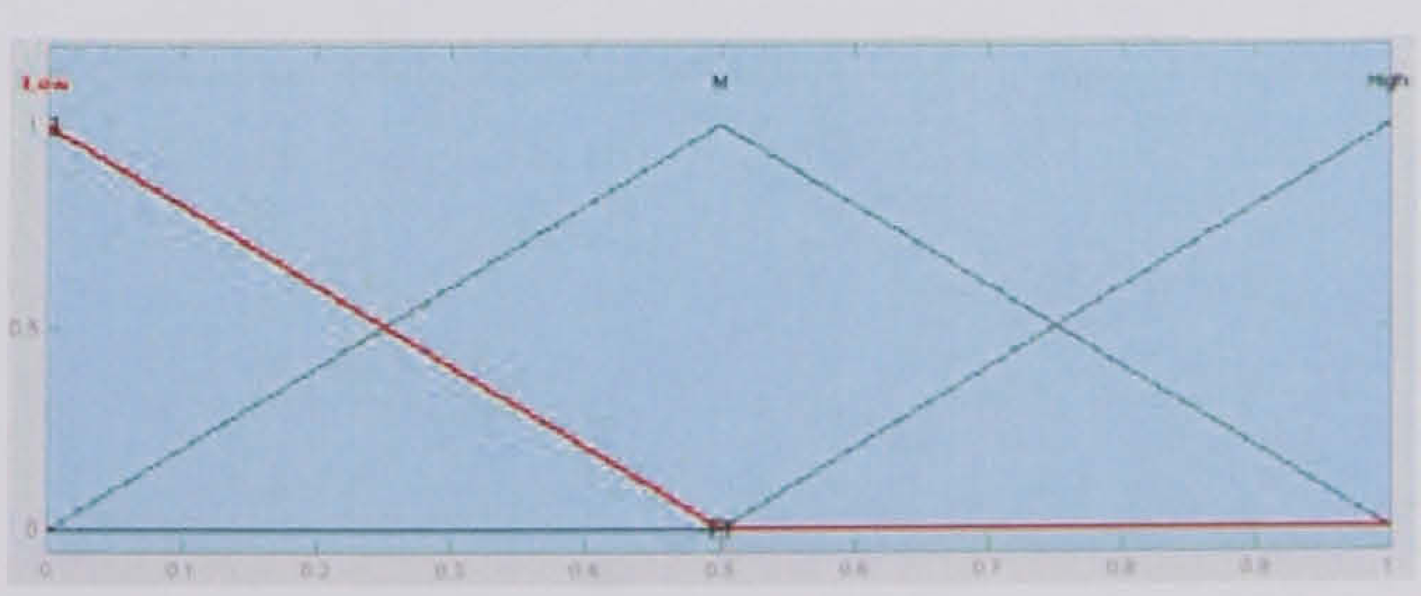
Figure 8.3 Input Membership Functions for Smart Cards in RP



Output MF of Cash Payment



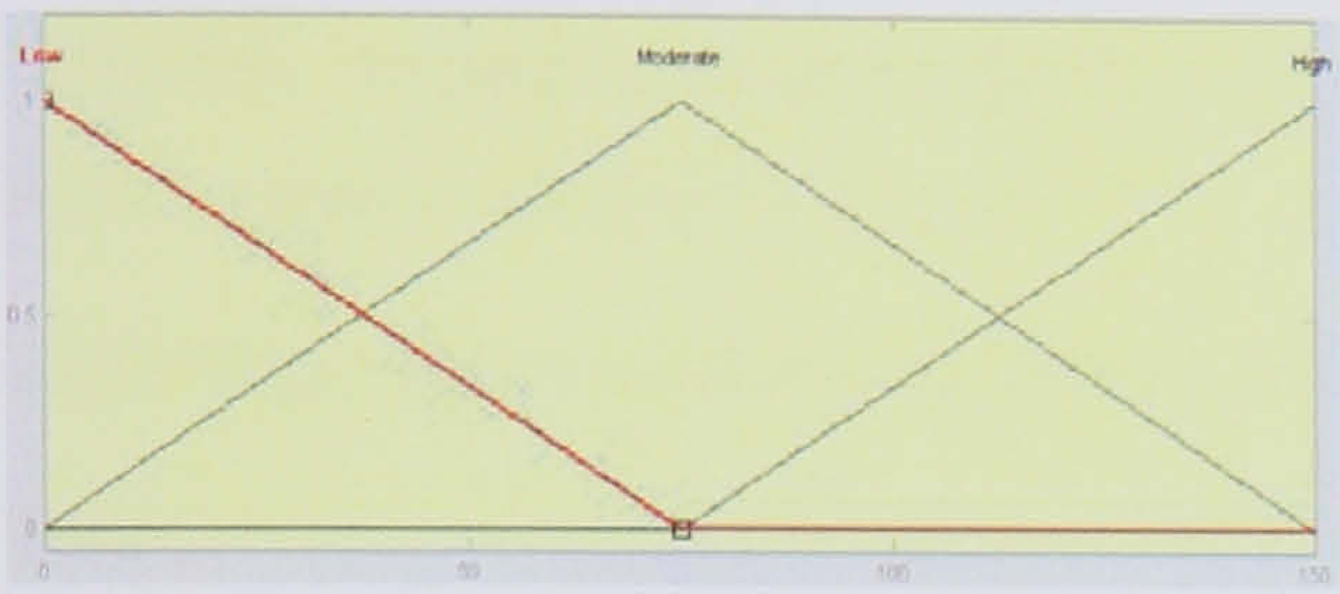
Output MF of Travel Cards



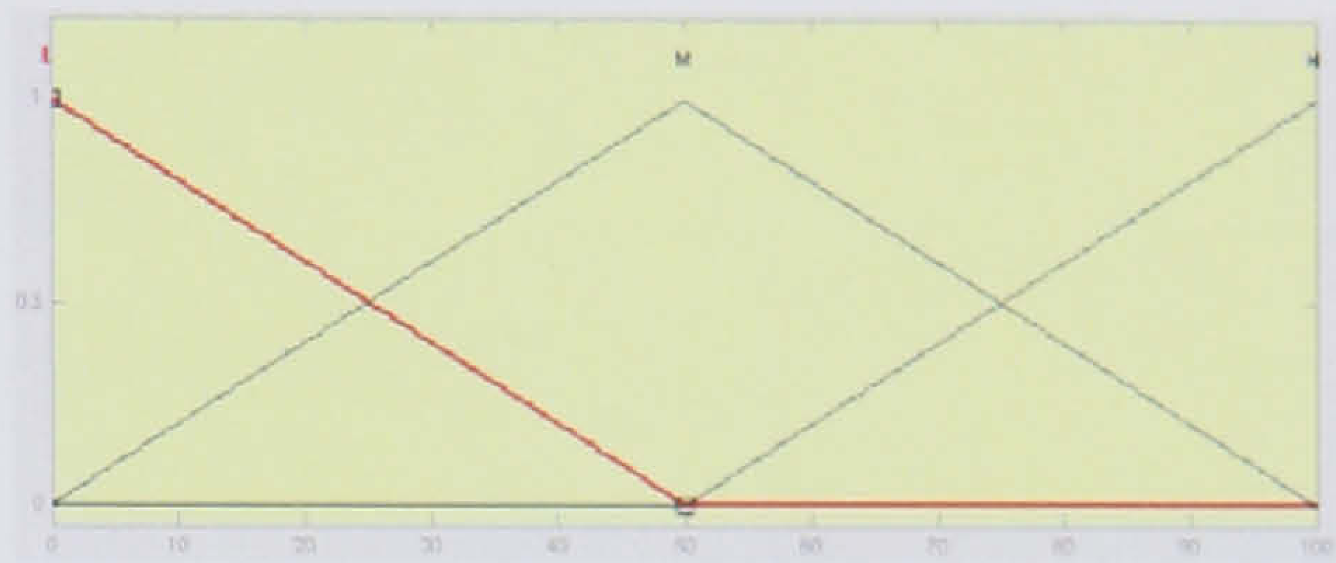
Output MF of Smart Cards

Figure 8.4 Output Membership Functions in RP

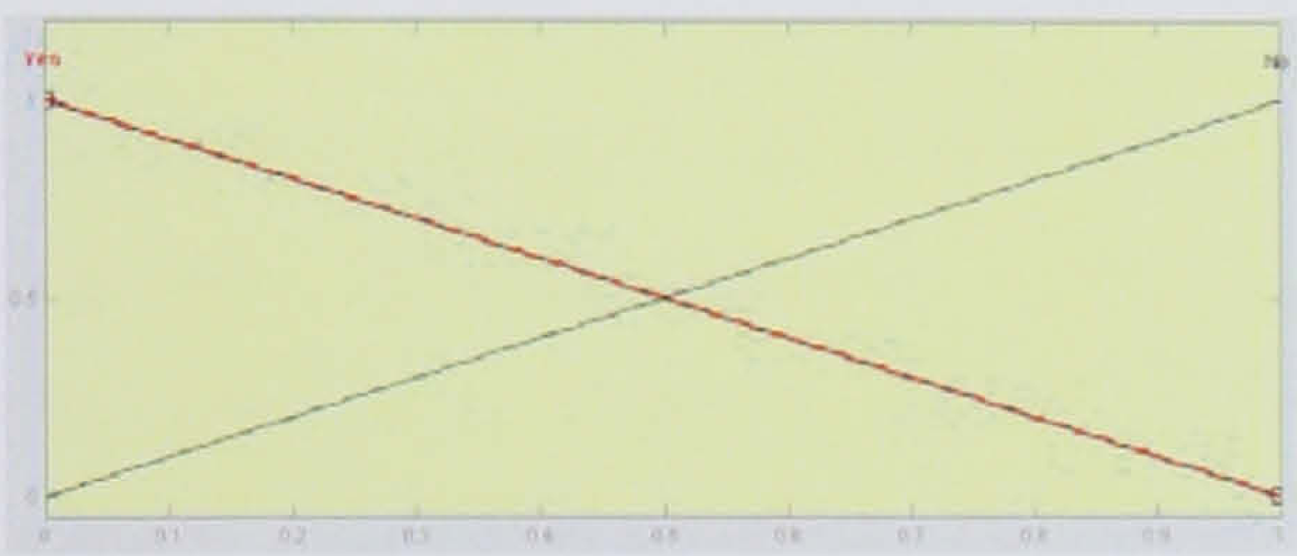
Due to different SP survey designs for four SP games, four different FL models are used in this chapter. A single output is designed to represent the choice probability of using one fare payment alternatives, because the four SP games are based on binary choice situations, if choice probability of one alternative is known (P_i) that means the choice probability of another one is indirectly known ($1-P_i$). As an example, the input and output membership functions for SP-1 are listed in Figure 8.5 and 8.6, respectively.



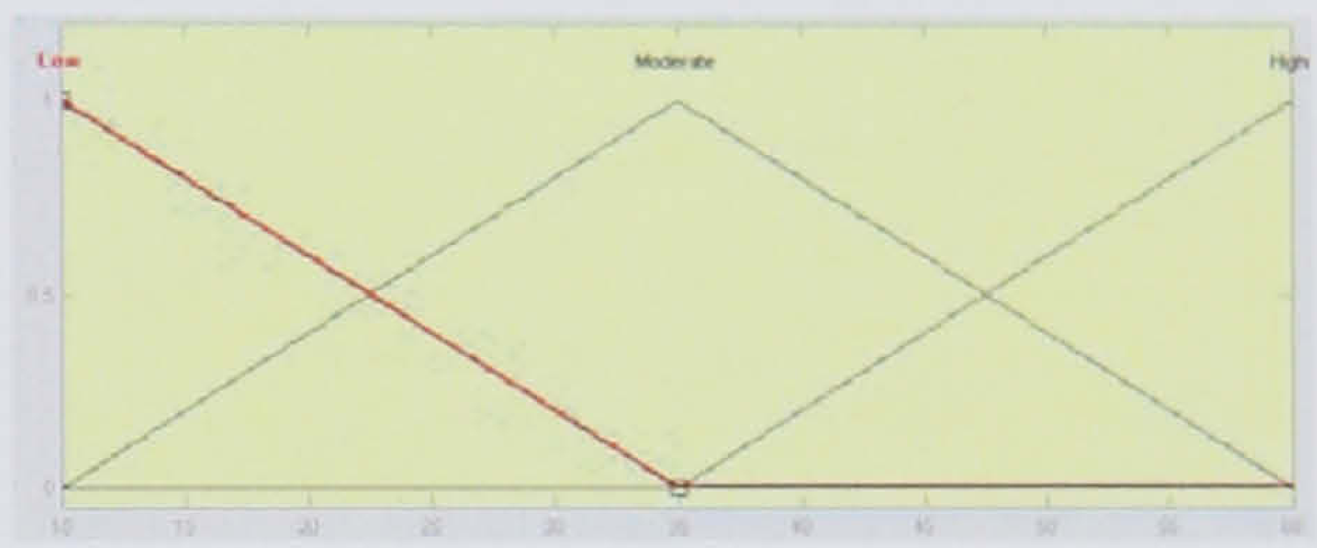
MF of Travel Cost of Cash



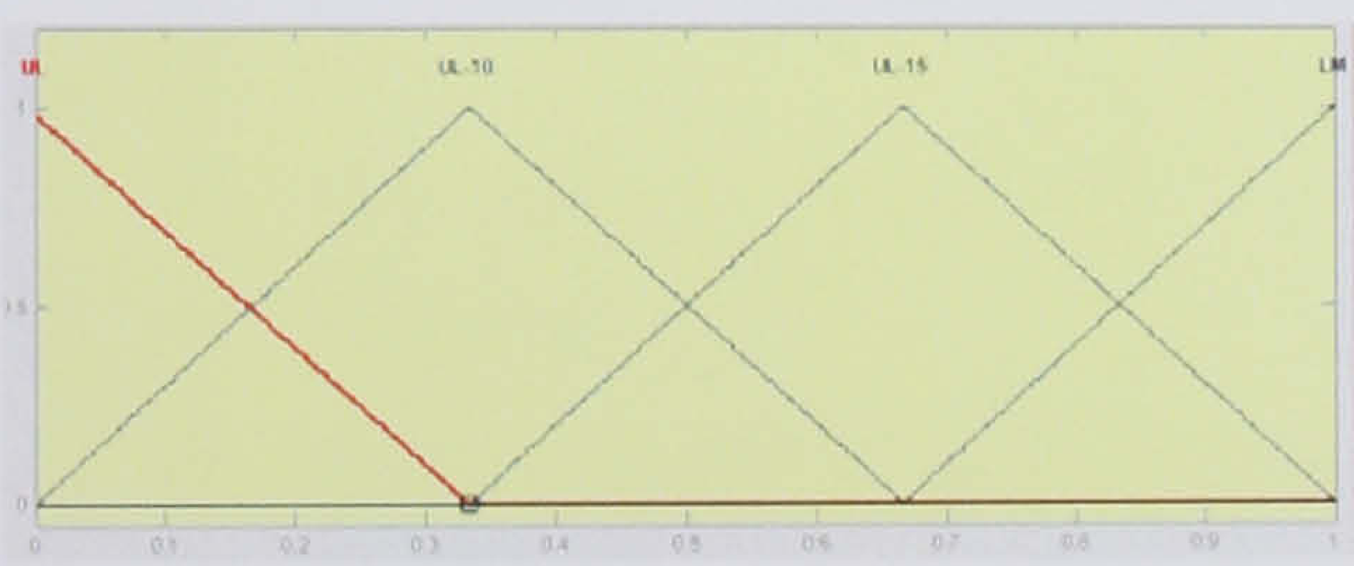
MF of Boarding Time of Cash



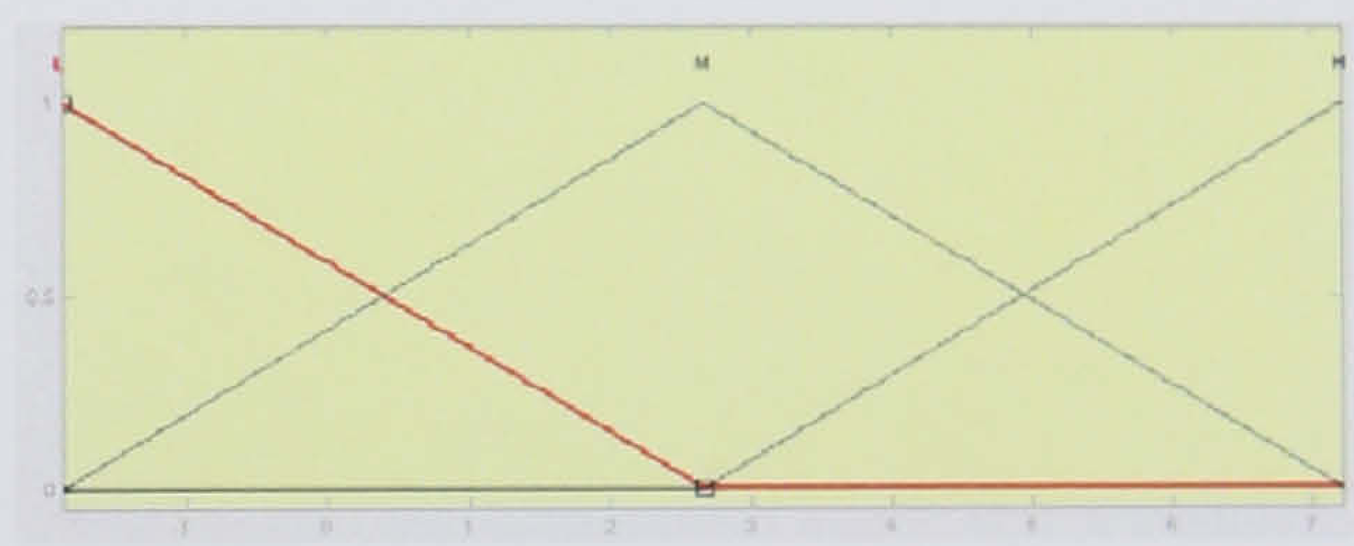
MF of Getting Changes Back of Cash



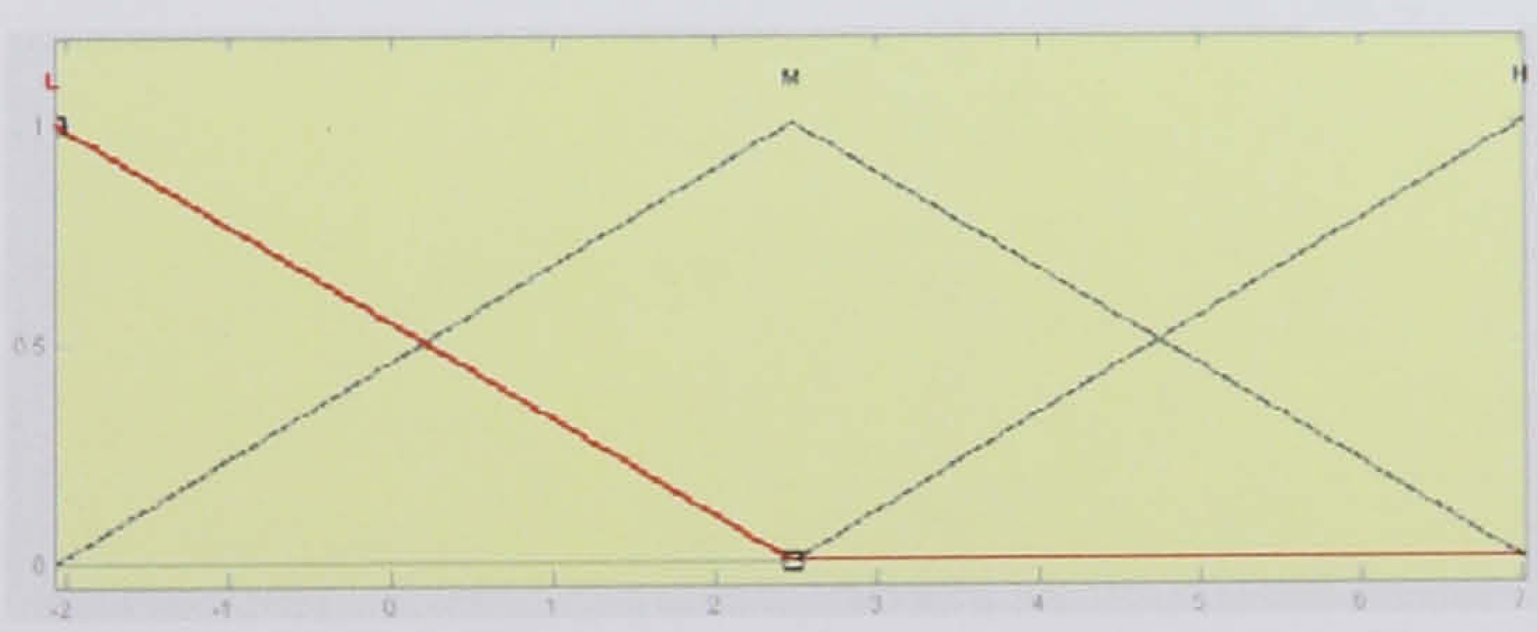
MF of Travel Cost of Travel Cost



MF of Service Routes of Travel Cards



MF of Gumbel Error Distribution of Cash



MF of Gumbel Error Distribution of Travel Cards

Figure 8.5 Input Membership Functions of Cash and Travel Cards in SP 1

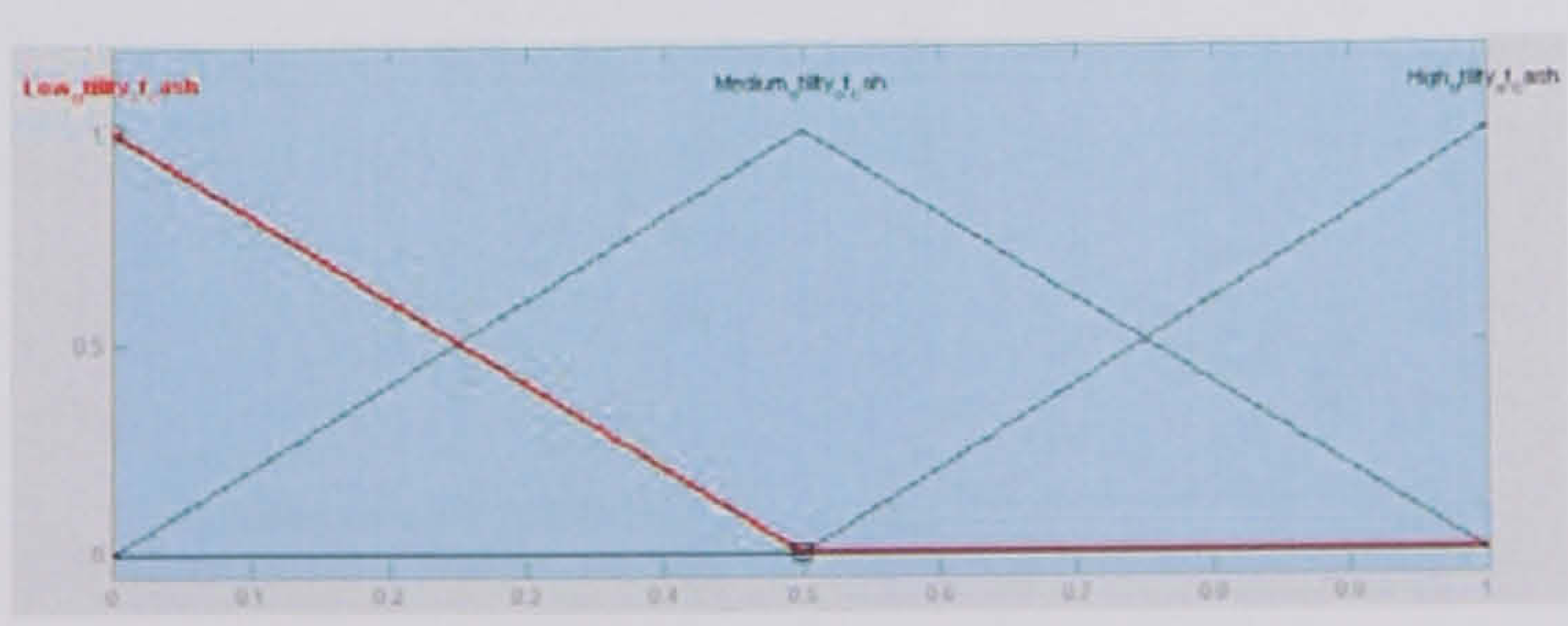


Figure 8.6 Output Membership Function in SP 1

Similar with the RP model, membership type for inputs and output of SP 1 is triangular type in the basic model. It also should be noted that in Figure 8.5, Gumbel error input is included in membership function design for both the alternatives. Normal distribution error term also follows the same membership type as Gumbel distribution in the FL model, except that different ranges (upper and lower boundary values) are allocated because the error numbers were randomly generated by the computer.

In this section, only membership functions of the basic models for the RP data and SP1 are listed. Regarding membership functions of five models (one for RP and four for SP) after the model calibration, please refer to Appendix D.

8.2.3 Rule Base and Fuzzy Inference System (FIS)

Rule

To model decision process, we use a rule-based expert system in the FL models. In this system, first of all, all input variables are categorised by linguistic levels, which can be overlapped (as shown in Figure 8.1-8.6) and assigned a certain degree of membership for a given level. The rule-based system consists of combinations of these linguistic categories of input variables for choosing different fare payment methods. The membership functions of linguistic categories as shown (Figure 8.1-8.6) in Section 8.2.2, present the detailed information. In general, the linguistic labels for input variables in the RP and SP survey include:

- Low, Moderate and High; or
- Yes and No; or
- Totally unsatisfied, Unsatisfied, Neutral, Satisfied, Totally satisfied; or
- Bad, Neutral, Good and Better.

The output labels (choice probabilities) for both RP and SP survey are presented as: Low, Moderate and High.

After determining linguistic labels for inputs and outputs, the next task is to generate decision rules. A rule has a general form: "if A_i then B_i", where the left hand side (LHS) of the rule is represented by the statement A_i, and the right hand side (RHS) of the rule by the statement B_i. The LHS of a rule deals with travel cost, boarding time, and other relevant data associated with fare payment options, expressed as labels of fuzzy sets. Although the RHS is choice related, it does not correspond directly to choice. Rather it serves to model internal representation of (latent) attitudes and preferences which are then used to make a final choice. Hence, it deals with the attractiveness of the various alternatives as a result of the conditions described by the LHS. For example: "IF travel cost of smart cards is *Low* and Boarding time savings is *Moderate*, THEN the probability of choosing smart cards is *High*". Therefore, LHS based on combination of inputs to make decision on RHS is primarily used in the FL model.

Table 8.1 Fuzzy Rules in the RP Survey

	IF																THEN		
	Travel cost of cash	OA of Cash	Travel cost of TC	BT-TC	OA of TC	SA of TC	Travel cost of SC	BT of SC	OA of SC	SA of SC	Top-up of SC	Difficulty of Top-up	Multifunction of SC	Error of Cash	Error of TC	Error of SC	Prob. choosing cash	Prob. choosing TC	Prob. choosing SC
1	L	S	M	L	S	L	N/A	N/A	N/A	N/A	N/A	N/A	N/A	L	M	N/A	M	--	--
2	--	--	L	L	--	--	N/A	N/A	N/A	N/A	N/A	N/A	N/A	--	L	N/A	L	--	--
3	L	TS	H	H	--	--	N/A	N/A	N/A	N/A	N/A	N/A	N/A	--	--	N/A	H	--	--
4	--	--	--	--	TU	L	N/A	N/A	N/A	N/A	N/A	N/A	N/A	L	H	N/A	H	--	--
5	H	--	M	L	--	H	N/A	N/A	N/A	N/A	N/A	N/A	N/A	--	--	N/A	L	--	--
6	M	N	M	M	N	M	N/A	N/A	N/A	N/A	N/A	N/A	N/A	M	M	N/A	M	--	--
7	M	US	L	H	N	L	N/A	N/A	N/A	N/A	N/A	N/A	N/A	L	H	N/A	M	--	--
8	L	S	N/A	N/A	N/A	N/A	M	M	S	M	Good	N	Bad	L	N/A	M	M	--	--
9	H	--	N/A	N/A	N/A	N/A	L	L	--	--	--	--	--	--	N/A	--	L	--	--
10	--	--	N/A	N/A	N/A	N/A	H	H	TUS	L	Bad	Bad	Bad	L	N/A	H	H	--	--
11	M	TUS	N/A	N/A	N/A	N/A	M	L	--	H	N	N	Good	M	N/A	L	L	--	--
12	H	--	N/A	N/A	N/A	N/A	H	M	--	L	Bad	N	N	M	N/A	M	M	--	--
13	M	--	N/A	N/A	N/A	N/A	H	L	--	L	N	Good	N	L	N/A	H	H	--	--
14	M	--	M	--	--	--	L	--	--	--	--	--	--	--	--	--	L	M	H
15	--	TUS	--	--	S	--	--	L	S	--	--	--	--	--	--	--	L	M	H
16	L	S	H	--	US	L	H	--	S	L	N	N	N	L	M	M	H	L	M
17	H	--	L	--	--	--	L	--	--	--	N	N	N	--	--	--	L	M	M
18	--	--	L	L	TS	H	--	--	--	--	--	--	--	H	L	H	L	H	L
19	M	S	M	M	S	M	M	M	S	--	--	--	--	M	M	M	M	M	M
20	M	US	M	L	S	H	M	L	S	H	Good	Good	Good	M	M	L	L	M	H
21	H	US	L	L	TS	M	H	M	N	M	Bad	Bad	Bad	L	L	H	L	H	M
22	L	S	H	M	S	L	M	M	M	M	N	N	N	L	M	M	H	L	M

Based on the above design, the rule base of the RP FL model consists of 22 rules. The detailed rules for fare payment choice decision are listed in Table 8.1. It should be noted that in the RP data set, due to the availability of payment alternatives, some data observations are presented by three payment choices situations, and some are based on binary choice situations, therefore, the rule base is distinguished by two situations in Table 8.1 (i.e., three outputs for cash, TC and SC; two outputs for TC and cash, or SC and cash).

Table 8.2 lists 16 fuzzy rules for fare payment choices in the SP-1 survey. Because four SP experiments were designed, in which different payment alternatives and attributes were allocated, the fuzzy rules also are separated according to four different SP survey designs. Detailed fuzzy rules for rest three SP games are attached in Appendix E.

Table 8.2 Fuzzy Rules in the SP-1

	IF							THEN
	Travel cost_C	Boarding time_C	Changes C	Travel cost_TC	Service TC	Error C	Error TC	Probability choosing cash
1	L	L	Y	L	Bad	L	L	H
2	M	H	Y	H	Bad	L	H	M
3	H	M	N	M	Bad	H	M	M
4	M	L	N	M	Neutral	M	M	M
5	M	L	Y	M	Good	M	H	H
6	L	M	Y	M	Good	M	L	H
7	M	H	N	H	Good	H	M	M
8	H	L	N	H	Better	L	M	L
9	M	M	N	L	Better	M	H	L
10	M	M	Y	H	Good	H	M	H
11	L	H	Y	M	Neutral	H	M	M
12	H	--	--	L	--	--	--	L
13	L	--	--	H	--	--	--	H
14	H	H	--	M	--	--	--	L
15	--	H	N	--	Better	--	--	L
16	L	--	--	--	Better	--	--	L

Fuzzy Inference System and Decision Process

Following determination of membership functions and rules, the next stage is to determine the fuzzy inference system being used in the FL model. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned (MATLAB Handbook). There are two types of fuzzy inference systems that can be implemented in the Fuzzy Logic Toolbox: Mamdani-type and Sugeno-type. Mamdani-type inference expects the output membership functions to be fuzzy sets. After the aggregation process, the fuzzy set for each output variable needs defuzzification. The main difference between Mamdani and Sugeno is that the Sugeno output membership functions are either linear or constant, while the Mamdani output membership function is a fuzzy set. Mamdani-type FL model is used in this case because it is intuitive, most commonly used and well-

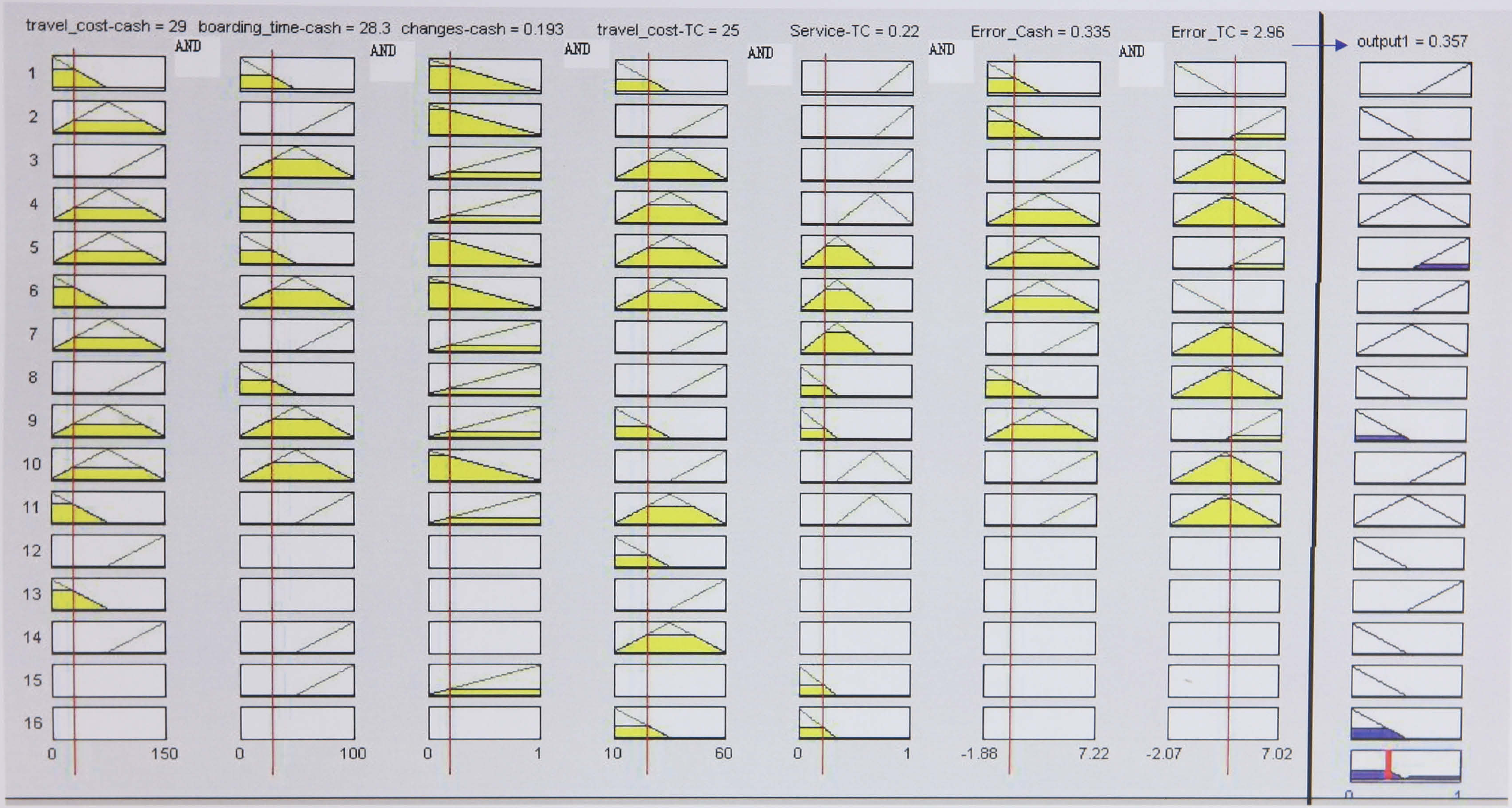


Figure 8.7 Example of Fuzzy Inference Diagram for SP 1

suited to human input. On the other hand, outputting values as estimation results rather a linear function in the FL models is easily comparable with results in the logit models.

Figure 8.7 illustrates an example of fuzzy inference system for SP 1 model. The fuzzy inference diagram is the composite of all the smaller diagrams, containing membership functions defined before. As can be seen from Figure 8.7, there are seven input variables and one output variable. For this specific example, let the input variables take the following values (please see Figure 8.7): travel cost of cash—29 yuan; boarding time – 28.3 seconds slower than travel cards; possibility of getting changes back – 0.193; travel cost of travel cards – 25 yuan; assessment of service by using TC – 0.22; random error of cash – 0.335; random error of TC – 2.96. All these input values are indicated by red line in each input attribute across these 16 rules. A yellow patch of colour under the actual membership function curve is used to make the fuzzy membership value visually apparent. The aggregation occurs down the eighth column, and the resultant aggregate plot is shown in the single plot to be found in the lower right corner of the plot field.

The defuzzified output value is shown by the thick line passing through the aggregate fuzzy set in Figure 8.7. As can be seen, the output value (the probability of choosing cash) is 0.357. That is to say, the probability of choosing TC is 0.643. It should be noted that the fuzzy inference diagram can show only one calculation at a time for one data observation, therefore the output value is an individual choice probability.

8.2.4 Estimation Results

Estimation results of five basic models for the RP and SP data are as follows:

RP Model

The rule base and membership functions for the RP model have been presented in Table 8.1 and Figure 8.1-8.4. In total, 22 rules are used as an expert system. All membership function types are triangular in this basic model.

Table 8.3 illustrates the estimation results by the RP FL model. In Table 8.3, the estimation results are presented by three sub-models: FL without error input; FL with Gumbel error input and FL with normal distribution error input. As aforementioned, the output of FL models is a value ranging between 0 and 1 (binary choice, for instance), which can be regarded as choice probability of using one alternative. In order to compare modelled choices with actual choices, all output values by FL models are finally rounded. A measurement to examine the model performance in the basic model is "percentage of matched estimation". From Table 8.3, we can see that the forecasting ability of FL models with error input is better than the model without error input. But when comparing with logit models, we find that in general, the forecasting performance was not improved by FL models obviously as we expect. In the RP logit model, the overall percentage of matched estimation reached about 70%. However in FL models, except that results of two FL models with error input are very close to that of the logit model, the performance of FL model without error is even worse than logit models. Some reasons could influence the estimation of FL models, such as the selection of membership function type, the membership curve location, and the rule base, *etc*. Therefore, the further work for fine-tuning the model is needed to be carried out in the following stage of model calibration.

Table 8.3 Predicted Share by three FL Models in the RP Survey

	FL (without error)	FL (Gumbel)	FL (Normal)
Cash	261	291	291
Travel Cards	342	283	283
Smart Cards	318	347	347
% of matched estimation	65%	69%	69%

SP 1 Model

Four different SP experiments were designed in the survey design. Since four data sets cannot be combined in one FL model as the logit model did, four different FL models for these four SP data sets are used separately in this section.

Table 8.4 presents the estimation results by three different sub-models for SP 1. The rule base of SP 1 model has been listed in Table 8.2, containing 16 rules. FL model without error has 5 input variables, and FL models with error have 7 inputs (plus error input for cash and travel cards, based on initial five inputs). Like the RP basic model, in order to simplify the estimation, only triangular type is considered as the membership function in the fuzzy inference system of SP 1 model.

The estimation result of SP 1 model has the similar trend with the RP model: "Percentage of matched estimation" in two models with error input is much higher than the model without error input. We can thus conclude that error input with a given distribution type does influence the model performance. FL model with Gumbel error input is slightly better than the model with normal distribution error input. When comparing with the logit model estimation, we can see that the highest % of matched data in FL model (78%) is still lower than in logit model (83.3%). That is to say, SP 1 basic model also should be calibrated in the later stage to improve the model performance.

Table 8.4 Predicted Share by three FL Models in SP 1

	FL (without error)	FL (Gumbel)	FL (Normal)
Cash	225	261	256
Travel Cards	556	520	525
% of matched estimation	69%	78%	76%

SP 2 Model

SP 2 model considers the trade off situations between cash and smart cards. The detailed rule base is presented in Appendix E, in which there are 21 rules to combine 11 input variables. Considering length of this thesis and relatively better performance of FL models with error input, we only list rules and membership functions, involving the error input. Again, triangular type is used for membership function in the basic model.

Table 8.5 Predicted Share by three FL Models for SP 2

	FL (without error)	FL (Gumbel)	FL (Normal)
Cash	597	623	681
Smart Cards	1671	1645	1587
% of matched estimation	68%	73%	76%

Estimation results in Table 8.5 show that the model performance of FL model with normal error is better than the model with Gumbel error, and better than the model without error. Compared with the logit model, it can be seen that FL model with Gumbel input has the same percentage of matched estimation as the logit model (73% for both models), while the model with normal error achieves a better model performance than logit model. This proves the expectation for FL models on improving forecasting ability. However, FL model without error still cannot reach the percentage of matched estimation as high as the logit models.

SP 3 Model

In SP 3 model, we use 10 input variables and 18 rules as the rule base (please refer Appendix E). From Table 8.6, we find that all three sub-models achieve more than 70% accuracy of prediction in fare payment share. But the overall percentage of matched estimation in the logit model reached about 80%. In the model calibration, some work need to be carried out to improve the mode performance, such as changing membership function type, fine-tuning membership curve location in the diagram, etc.

Table 8.6 Predicted Share by three FL Models for SP 3

	FL (without error)	FL (Gumbel)	FL (Normal)
Travel Cards	732	760	775
Smart Cards (Pay as you go)	967	939	924
% of matched estimation	72%	74%	75%

SP 4 Model

In SP 4 model, 9 input variables are considered and the expert decision system consists of 14 rules. In general, three sub-models can achieve a good performance (all above 70%). But when we compare estimation results of FL models with the logit model, we can see that percentage of matched data in the logit model reached about 84%, which is much higher than FL models. The differences of this measurement between the FL models and logit models call for further model calibration to identify whether the FL models can be improved on forecasting ability.

Table 8.7 Predicted Share by three FL Models for SP 4

	FL (without error)	FL (Gumbel)	FL (Normal)
Travel Cards	922	889	889
Smart Cards (Pay monthly)	543	576	576
% of matched estimation	74%	76%	76%

8.2.5 Model Calibration

In Section 8.2.4, estimation results by FL basic models have been presented. As can be seen from the five base FL models, in general the model performance is not satisfactory as expected. Arbitrary selection of components of FL models (e.g., membership functions) may result in such model performance in the base models. Therefore, in this section, FL models need to be calibrated based on these basic models to achieve the satisfactory model outputs. Fuzzy logic model calibration is aimed to fine tune the fuzzy inference system so as to capture the best fit decision-making process. In this section, the model calibration is discussed by the following two aspects:

- Calibration of membership type.
- Calibration of location of membership curve.

Calibration of membership type

In the basic models, triangular type is selected for all input and output variables as the membership function. The reason for choosing triangular type is because of its simplicity, but the key drawback of this kind of membership type should not be overlooked: changes of degree of membership from point to point on the curve is too straight, particularly in the turning point. This can result in significant changes around the turning point when input value changes. Therefore, some other membership functions can be tried in the model calibration to examine whether different membership types can influence the estimation result.

Among membership functions available in fuzzy logic toolbox of MATLAB, Gaussian can overcome the problem in triangular type aforementioned. It has the advantage of being smooth and non-zero at all points. Although some other membership type can also achieve smoothness as Gaussian type, considering its relative simplicity to some extent, in the model calibration, only Gaussian type is tested. After changing to Gaussian type for all variables, we can compare changes of "overall percentage of correct prediction" between the basic model and calibrated model in Table 8.8:

Table 8.8 Comparison between the Basic Model and Calibrated Model-1

		Basic Model	Calibrated Model-1
FL without error	RP	65%	69%
	SP1	69%	69%
	SP2	68%	71%
	SP3	72%	73%
	SP4	74%	76%
FL with Gumbel error	RP	69%	71%
	SP1	78%	80%
	SP2	73%	78%
	SP3	74%	75%
	SP4	76%	77%
FL with Normal error	RP	69%	71%
	SP1	76%	79%
	SP2	76%	79%
	SP3	75%	76%
	SP4	76%	77%

It can be observed that after changing to Gaussian distribution type, "percentage of correct prediction" in five calibrated models become higher than in the basic models. This can prove that different membership types can influence the model performance as assumed beforehand. In the calibrated model-1, "percentage of correct prediction" of RP models with error input and SP2 model with error input are higher than the logit model. But SP1, SP3 and SP4 models still cannot achieve as good results as the logit model. Therefore, another calibration needs to be tried next.

Calibration of location of membership curve

In the basic models and calibrated model-1, all membership functions are arranged by symmetrical layout. For example, three curves of travel cost membership function is evenly distributed and overlapped in the diagram (Please see Figure 8.1). The peak point for membership "medium" is right arranged in the middle point of X-axis. To design such kind of membership layout, one pre-condition is that data are expected to be normally distributed within the range. However, a variable may not be normally distributed, for example, if the range of travel cost is between 10 and 200yuan, the majority of value could be much closer to the left side (e.g., 40-50yuan). In other words, not many travellers spend more than 100yuan per month. Under this circumstance, the range of 40-50 may be regarded as "Medium", rather than the range around 105yuan we set before. Therefore, if a symmetrical membership function was used for such input variable, the output may be biased due to unsuitable degree of membership and then the accuracy of output also would be affected. So in the second stage of model calibration, some asymmetrical membership curve will be used to fine tune the fuzzy inference system.

After testing distribution type for input variables of RP and SP data, we find that travel cost is different from other variable, because it is not normally distributed. Although travel costs in the RP and SP range between 10yuan and 200yuan per month, most cost values are around 40-50yuan. Therefore, the membership curve for travel cost is adjusted closer to the left side of X-axis (please refer to Appendix D). Among membership types in MATLAB, Gaussian-2 membership curve can achieve asymmetrical designs, in the meantime, other advantages of Gaussian type still can be retained. Finally, based on the calibrated model-1, the FL models for the RP and SP data are adjusted as: travel cost variable is designed as asymmetrical Guassian-2 (for the category of "Medium", 50yuan is set as the peak point in the membership curve); other variables still keep the same membership (Gaussian type). The calibrated results in Model-2 are listed in Table 8.9.

Table 8.9 Comparison between the Calibrated Model-1 and Calibrated Model-2

		Calibrated Model-2	Calibrated Model-1
FL without error	RP	72%	69%
	SP1	71%	69%
	SP2	73%	71%
	SP3	75%	73%
	SP4	79%	76%
FL with Gumbel error	RP	74%	71%
	SP1	82%	80%
	SP2	81%	78%
	SP3	82%	75%
	SP4	82%	77%
FL with Normal error	RP	74%	71%
	SP1	82%	79%
	SP2	82%	79%
	SP3	83%	76%
	SP4	82%	77%

From Table 8.9, we can see that the model performance of the calibrated model-2 is significantly improved after we changed the location of membership curve of cost variable. Among three sets of models for five different data sources, FL models with error input still show a better performance than FL models without error input. "Percentages of correct prediction" for two FL models with different error inputs are almost the same. Therefore, regarding this research, we can imply that Gumbel error and normal distribution error have a very similar effect to the estimation results. In addition, calibrated FL models for RP, SP2 and SP3 outperformed the logit model on "percentage of correct prediction". SP1 and SP4 also show the very close outcomes to the logit models. Finally, because of satisfactory outcomes by FL models with error input, the estimation results by calibrated model-2 are considered as the final estimation results for the later analysis.

8.2.6 Discussions on FL Model Performance

First of all, from estimation results in two kinds of FL models (with and without error input), we can find a common feature that FL with error input (Gumbel and normal distribution type) can achieve higher percentage of correct prediction than FL model without error input, not only in the basic model stage but also in the calibration stage. Therefore, it can be concluded that like the error term in the utility function of logit model, the error input in the FL models does influence human decision making besides those attributes we have included as deterministic factors in decision making process. But when comparing two FL models with error inputs, we can see that FL model with Gumbel distribution type and normal distribution type present the same results (including percentage of correct prediction and forecasted market share). Therefore, it can be implied that based upon the fuzzy rule base, both distribution types for the error input have the same effect on the model forecasting ability.

Secondly, when comparing the standard logit models in Chapter 7 with the FL models, we can observe that after model calibration, generally FL models can achieve a satisfactory prediction rate as high as (or very close to) the logit models. However, such improvement on model performance is not as significant as we expected. Some potential factors could influence the final model performance:

- Fuzzy rule base: the fuzzy rule base was pre-determined to simulate human's decision making. Although it is called an expert system, it cannot cover all respondents' decision process. We can hardly include every respondent's decision process in the rule base, though the model performance could be better than models with some representative rules. Therefore, the balance between the number of rules being used and the goodness of fit is one of concerns in FL models and can be one of my further research directions. Therefore decision bias and error still does exist.
- Half-half chance of choosing either payment alternatives: not only for the basic models but also for calibrated models, the result of 50%-50% choice probability for using a payment method can be found. When forecasting market share for each fare payment method, in order to obtain the number of individual choices we rounded choice probabilities equal or greater than 0.5 to 1.0, because 0-1 binary code was used to indicate two alternatives (1 for choosing alternative "A", 0 for choosing alternative "B"). But actually it is very difficult to tell 50% must indicate that the respondent chose alternative "A", not "B". Therefore, such kind of data (0.5, or a range of 0.45-0.55 for instance) can reveal psychologically uncertain situation of respondents when they traded off between different payment methods. In order to capture such choice behaviour, choice probabilities ranging between 0.45 and 0.55 by the FL models are picked up and listed in Table 8.10.

Table 8.10 Uncertainty of Choice Behaviour in FL Models

	RP	SP-1	SP-2	SP-3	SP-4
No. of Uncertain Choices	78	86	273	255	221
% in the total data	9%	11%	12%	15%	15.1%

As can be seen from Table 8.10, percentage of uncertain response in the SP survey is slightly higher than the RP survey. It is not surprising because the RP survey collected respondents' actual choices in reality, while the SP survey was carried out based on hypothetical situations. Therefore, Table 8.10 can explain uncertainty of choice behaviour, particularly for the SP survey.

Meanwhile, in order to compare the model performance on the uncertainty of choice behaviour between different models, Table 8.11 lists the number of uncertain choices in the RP and SP data (individual choice probabilities, ranging 0.45 to 0.55) by the logit models. It should be noted that in Table 8.11 the uncertain choices are based on the validation data, because the main sample had involved in the model estimation. Therefore, it is necessary to use another data set (rather than the data that have been used in the model estimation) to check the model performance of the logit model. It can be seen that besides SP4, the percentages of uncertain choices in RP, SP1, 2 and 3 of FL models are slight lower than in logit models. Admittedly the uncertainty on choice behaviour (i.e., no preference to either alternative) cannot be avoided (but it could be minimised to some extent through the model design), because that is also a rational reaction of respondents to some choice situations.

Table 8.11 Uncertainty of Choice Behaviour in Logit Models

	RP	**SP-1**	**SP-2**	**SP-3**	**SP-4**
No. of Uncertain Choices	8	9	29	27	21
% in the total data	9.2%	11.5%	12.78	15.9%	14.5%

- Different algorithm of FL and logit models: Although the utility function in the logit model was based on a linear expression between deterministic input variables and alternative utility, the maximum value of log-likelihood function can be obtained by the Newton-Raphson algorithm, which uses first and second derivatives to reach the optimum. That is to say the Newton-Raphson algorithm is basically a sort of data-driven approach to get the optimum, because this algorithm takes all data into account and then finds the best fit outcome. However, FL models' approximate reasoning approach, which can be viewed as non-linearity, is based on the rule base and membership functions predefined. The determination of rules and selection of membership function could highly influence the model performance.

Thirdly, the calibration of membership function indicates that shape of membership and location of membership curve can influence the FL model estimation. For some input data, it is necessary to identify the distribution type of input data before we set membership curves in different locations in the X-axis. If an unsuitable membership curve was set, it is highly possible that the outcomes would be biased.

Finally, in addition to comparison of "percentage of correct prediction", another way to compare different FL models is to measure the goodness of fit. In the FL models, measurement of the goodness of fit is the root mean square error (RMSE). Because in three

FL models, models with different error distribution types present almost the same results, it is necessary to use the goodness of fit measure to identify which model performed better. Table 8.12 compares RMSE between these two FL models. As can be seen, RMSEs by FL model with normal distribution are slightly lower than FL model with Gumbel distribution, indicating FL with normal distribution has the relatively better model fit than FL with Gumbel distribution. Therefore, in the future analysis, the estimated results by FL with normal distribution are used in comparisons with other models.

Table 8.12 Comparison of the Goodness of Fit (RMSE) of FL Models

	RP	SP1	SP2	SP3	SP4
FL with Gumbel error	0.412	0.3433	0.3491	0.3511	0.3671
FL with Normal error	0.406	0.3251	0.3212	0.3428	0.3287

8.3. Artificial Neural Network Analysis

In the ANN model analysis, the following aspects are specified for this research, including inputs and outputs, network structure, training and learning process, and model validation.

8.3.1 Inputs and Outputs

The inputs and outputs for the ANN analysis need to be determined in the first instance. In this research, the inputs are attributes and levels of fare payment alternatives, the same as the discrete choice model analysis and the output of the ANN model is defined as the individual choice probabilities of using a certain payment method so that the estimation results from different models (standard MNL model and ANN model) can be comparable.

The input data are structured as follows:

- For those quantitative variables (e.g., travel cost, boarding time), data normalisation is required prior to the modelling analysis, because among input vectors, travel cost and boarding time vary in a wider range (unlike other input vector ranging between 0 and 1). The purpose of normalizing inputs and target outputs is to ensure that the statistical distribution of values for each net input and output is roughly uniform. In addition, the values should be scaled to match the range of the input neurons. Advantages of data normalization include that it can greatly improve a network's performance (good results) as well as significantly fasten the calculation, because normalization reduces the differences between the variation ranges of the different variables. Therefore, before using the ANN model, it is necessary to normalize these two variables to fall in [0 1].
- For those qualitative variables (previously coded as dummy variables in the discrete choice model analysis), they are re-structured as binary-value input (0-1 input). Therefore, the number of inputs for qualitative variables in the ANN model depends on

how many levels those variables have (e.g., if a variable has three levels, then in the ANN model, three separate inputs columns are used by 0-1 value for this variable, in which '1' indicates the presence of this level for a individual, '0' for the absence of the relevant level of the variable).

The output is defined by any values ranging between 0 and 1, so we can regard it as the individual choice probabilities of using different fare payment methods. The number of output neurons is decided by the number of alternatives in the RP/SP survey. In the RP survey, we investigated respondents' actual choice behaviour to the existing three PT payment means: cash, travel cards and smart cards, so the output layer for the RP data has three neurons. In the SP data, because four different binary choice experiments were employed, two output neurons to present binary choice situations are set in the ANN model for the SP data.

8.3.2 Network Structure

The network structure is used to indicate to the connection between input neurons and output neuron(s). There are a great number of ANN structure types available, such as perceptron, backpropagation network, self-organising network, recurrent network, etc. As discussed in Chapter 3 and Chapter 4, BP network structure has been widely used in transportation studies. In this research, considering efficiency of training process and simplicity of data flow (from the left side of network to right side, input→output), three-layer feedforward network structure is finally used.

In addition, the ANN model is defined as supervised network, which means that the model training process is supervised by the target value so that the output of the ANN model can be adjusted gradually until the outcome is satisfied or it reaches the maximum training epoch. The target value of this model is surveyed individual choices of fare payment methods. Regarding the supervised training process, Section 8.3.3 gives the explanation in details.

In this section, another important task is to determine the number of neurons in the hidden layer. As stated in the last section, the number of neurons in the input and output layers is decided by the number of variables and alternatives. Because the number of inputs and outputs were known, we can directly set the relevant number of neurons in the output layer and input layer. But the determination of neurons in the hidden layer requires trial and error in most previous studies. Through comparing the goal error and model performance (running time, training epoch, *etc*), the best fit number of neurons in the hidden layer is determined. Finally, the number of neurons for each layer in five network models (RP, SP1, SP2, SP3 and SP4) is listed in Table 8.13.

Table 8.13 Number of Neurons for Each Layer in the ANN Model

ANN Model	On Input Layer	On Hidden Layer	On Output Layer
RP	57	68	3 (probabilities for three alternatives)
SP1	16	17	2 (choice probability of cash and TC)
SP2	28	31	2 (choice probability of cash and SC)
SP3	18	22	2 (choice probability of TC and SC)
SP4	19	24	2 (choice probability of TC and SC)

The connection between two layers and algorithm of how each layer and neuron simulate and transfer information is introduced in Section 8.3.3.

8.3.3 Training and Learning Process for ANN Model

The training the learning process of ANN model is to simulate and replicate human brain functions and obtain the best fit input-output mapping relationship through learning the mapping relationship between known input data and output (target value in supervised training process). As shown in Figure 4.4, the type of neural networks that are considered in this paper is the feedforward (FF) multi-layer neural network. Multi-layer neural networks are often trained using an algorithm known as "backpropagation" (BP) algorithm. In this section, the discussion about the training and learning process includes the determination of training function, transfer function, learning function and the relevant parameters.

Training Function

The training function in this research is Levenberg-Marquardt (LM) algorithm. LM algorithm uses standard numerical optimisation techniques and is regarded as the fastest training algorithm for moderate sized feedforward neural network. Another advantage of LM algorithm is that it has computer memory reduction feature for use when the training set is large (thousands of data observations in this research study, for example).

The training parameters for LM training algorithm include: training epoch and error goal. Training epoch, as one of training parameters, is used to define the maximum times to train the network. In this model, the training epoch was set 1000.

The concept of error goal is used to measure the network performance. The default performance function for feedforward networks is mean square error (MSE), the average squared error between the outputs a and the target outputs t. The default MSE error goal is 0 in MATLAB software, which can be viewed as the best fit training process. But with the increase of input vectors, the training error the network can achieve also increases. The error goal can be influenced by the number of input variables. Considering more than about 10 input vectors for each data set, the MSE error goal was set to 0.1. Moreover, training epoch and error goal are two measurements to stop the network training process (either reaching the maximum training epoch or error goal).

Transfer Function

The role of transfer function is to produce a scalar output from a neuron based on a weighted input into the neuron. For this ANN model, two kinds of transfer function types are used: tan-sigmoid (for data transferring between input layer and hidden layer) and log-sigmoid (for data transferring between hidden layer and output layer). Basically, they are sigmoid transfer type, which may have any value between plus and minus infinity, and squashes the output into the range 0 to 1. Because in the ANN model, it is desirable to constrain the outputs in the network (such as for the output layer, we define the output ranging between 0 and 1), so that it can be considered as individual choice probability, transfer functions linking different layers are based on sigmoid function. Secondly, sigmoid function is non-linear. Linear networks cannot perform any nonlinear computation. Use of a nonlinear transfer function makes a network capable of storing nonlinear relationships between input and output. Thirdly, sigmoid transfer function is commonly used in backpropagation (BP) networks, in part because it is differentiable, unlike linear transfer function type.

Learning Function

The learning function in the ANN model is used to update the network weight/bias until the best fit output can be obtained. In this feedforward network, gradient descent with momentum (called 'Learngdm' in MATLAB) training function was used, providing faster convergence to update network weight and bias. For "learngdm" function, first of all, the momentum needs to be determined, because without momentum a network may get stuck in a shallow local minimum. Acting like a low pass filter, momentum allows the network to ignore small features in the error surface.

Momentum can be added to backpropagation learning by making weight changes equal to the sum of a fraction of the last weight change and the new change suggested by the backpropagation rule. The magnitude of the effect that the last weight change is allowed to have is mediated by a momentum constant (mc) which can be any number between 0 and 1. When the momentum constant is 0 a weight change is based solely on the gradient. When the momentum constant is 1 the new weight change is set to equal the last weight change and the gradient is simply ignored. In this ANN model, momentum constant is set as default: 0.9.

In the ANN models, another parameter for 'learngdm' function that needs to be set is the learning rate. The changes to the weights and biases of the network are obtained by multiplying the learning rate to the negative of the gradient. The larger the learning rate, the bigger the step. If the learning rate is made too large the algorithm will become unstable. If the learning rate is set too small, the algorithm will take a long time to converge. According to previous studies (Neural Network Toolbox, MathWorks, 2004), the learning rate for this ANN model is set as 0.001. Meanwhile, in order to reduce the running time in the software. in this MATLAB programme file, parameter of ratio of learning rate increase ("lr_inc") was used at the same time. Ratio of the learning rate increase is particularly for adaptive learning process, in which the predefined learning rate can be changeable by a give ratio to increase. The ratio of learning rate increase is 1.05, as suggested in the neural network toolbox (MathWorks, 2004).

Weights and Biases

In the training process, another important task is to determine initial weights and biases for the ANN models. In MATLAB, input weights and biases can be initialised by the syntax of "rands" to generate random values ranging between -1 and 1. But the feedforward network itself can automatically initialise input weights and biases according to the method by Nguyen and Widrow (1990). This method generates initial weight and bias values for a layer so that the active regions of the layer's neurons will be distributed roughly evenly over the input space. It has several advantages over purely random weights and biases: (1) few neurons are wasted (since the active regions of all the neurons are in the input space); (2) training works faster (since each area of the input space has active neuron regions). Therefore, in the ANN models, we let the network itself initialise input weights and biases.

8.3.4 Estimation Results and Model Validation

A problem that occurs during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network, the error is large. The network has memorised the training examples, but it has not learned to generalise to new situations. This issue that exists in the trained network calls for the model generalisation. Generalisation is an attribute of a network whose output for a new input vector tends to be close to outputs for similar input vectors in its training set. The calibration task for the neural network model is to improve the generalisation.

First of all, besides 90% of total data, which were entered in the ANN model as training data (the same data set as in the logit model estimation), the rest 10%, which once were used as the model validation data in the logit model, were retained as test data to examine the generalisation ability in the trained network.

There are two solutions available and widely used for improving the generalisation: regularisation and early stopping with validation. The regularisation method involves modifying the performance function, which is normally chosen to be the sum of squares of the network errors on the training set. Neural network toolbox in MATLAB has included a routine which can automatically set the optimal performance function to achieve the best generalisation.

As another solution for improving generalisation, early stopping is a technique based on dividing the data into three subsets. The first subset is the training set used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error will normally decrease during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set will typically begin to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned.

Except these two solutions, some new training options can also achieve improvement of network generalisation, including:

- Minimising with variations of mean squared error for better generalisation. Such training simplifies the problem of picking the number of hidden neurons and produces good networks that are not over-trained.
- Training with validation to achieve appropriately early stopping. Here the training result is checked against a validation set of input output data to make sure that overtraining has not occurred.
- Stopping training when the error gradient reaches a minimum. This avoids wasting computation time when further training is having little effect.

Regularisation technique is used in the network validation for ANN models in this research because of its easiness of editing in the MATLAB programme file.

Moreover, compared with genetic algorithm which is capable of capturing a global minimum of error during the model estimation, another problem of ANN models is that estimation results would fall in a local minimum. The algorithm of ANN method (such as gradient algorithm) may result in this problem, because the model always stops training as long as it cannot capture a lower error than the current one within a given subset of input data, rather than a full range of input data. Therefore, in order to minimise the influence of this intrinsic problem in ANN technique, in this research the network models were trained several times and the best results (with the lowest training error) were chosen as the final results. Finally, the calibrated ANN model outputs are listed in Table 8.14-8.18 for the RP and SP data respectively. In these five tables, the following measures are listed to assess the performance of ANN models, including, modelled choices, MSE value, and percentage of correct prediction.

RP Model:

Table 8.14 presents the estimation outcomes for the RP ANN model. As can be seen from Table 8.14, both training and test RP data can obtain higher 'percentage correct prediction' than the results in the logit model. It can be proved that in this case non-linear functions can output better results on modelling respondents' choice behaviour and predicting market share of fare payment methods. However, the results of the ANN model using the RP test data presented a lower 'percentage of correct prediction' result than the training data. Therefore, it can be concluded that generalisation ability of network can also be affected by the number of inputs in the network to some extent (in this case, the RP ANN model contained much more inputs than the SP ANN model).

Table 8.14 RP Prediction Results by ANN Model

Observed	Predicted by ANN							
	Training Data				Test Data			
	Cash	**TC**	**SC**	**Total**	**Cash**	**TC**	**SC**	**Total**
Cash	163	21	42	226	19	4	2	25
TC	21	233	29	283	4	24	4	32
SC	12	16	245	273	4	5	21	30
Total	196	270	316	782	27	33	27	87
MSE (error goal: 0.1)	0.137/1000				0.169/1000			
% of correct prediction	82%				73.6%			

Another output of the ANN model is MSE (Mean Square Error), which is used to measure the goodness of fit. MSE in Table 8.14 is final MSE value which was obtained when the maximum training epoch was reached in the RP model. From viewing the final MSE value, the ANN model for the RP data cannot reach the error goal even when the model had reached the maximum training epoch (1000). This is because, as aforementioned, with the increase of number of input vectors, the training error that the network model can achieve also becomes big.

SP1 Model:

When observing outcomes by SP models, we find that in general, ANN SP models achieved satisfied results for predicted market shares and good model fits. All percentages of correct prediction are above 80% not only in training data sets but also in test data sets, higher than the logit models and fuzzy logic models.

For SP 1 model, the percentage of correct prediction for the test data is about 84.6%. In the meantime, we can see that after trained for 423 times, the network stopped and obtained the optimal output with MSE error of 0.091. The MSE value is lower than the error goal we set for the network training process. Therefore, it can be concluded that SP 1 model can achieve a better model performance than the RP ANN model and logit models.

Table 8.15 Prediction Results by ANN Model for SP 1

Observed	Predicted by ANN					
	Training Data			Test Data		
	Cash	TC	Total	Cash	TC	Total
Cash	192	38	230	23	4	27
TC	40	433	473	8	43	51
Total	232	471	703	31	47	78
MSE (error goal: 0.1)	0.087/412			0.091/423		
% of correct prediction	89%			84.6%		

SP 2 Model:

Table 8.16 Prediction Results by ANN Model for SP 2

Observed	Predicted by ANN					
	Training Data			Test Data		
	Cash	SC	Total	Cash	SC	Total
Cash	462	144	606	59	13	72
SC	142	1293	1435	31	124	155
Total	604	1437	2041	90	137	227
MSE (error goal: 0.1)	0.071/398			0.099/422		
% of correct prediction	86%			80.6%		

Table 8.16 lists estimation results by ANN SP 2 model. As discussed for SP-1 model, the model performance of ANN SP-2 is also acceptable as we expected. The measurement of goodness of fit is lower than the error goal when the training process stopped at 422 times.

SP3 & SP4 Model:

Regarding SP3 and SP4 models, Table 8.17 and 8.18 illustrate estimations in detail. We find the similar trend about the outcomes in SP3 and SP4 models with SP1 and SP2 models. Meanwhile, it is not surprising that we find the final MSE values in test data for these four SP models are greater than in training data. Because the test data are not involved in the training process, some mapping relationship existing in the test data might not be captured in the trained network. But the final result still can be satisfactory and be meaningful to this research.

Table 8.17 Prediction Results by ANN Model for SP 3

Observed	Predicted by ANN					
	Training Data			Test Data		
	TC	SC	Total	TC	SC	Total
TC	662	76	738	66	15	81
SC	93	698	791	16	73	89
Total	755	774	1529	82	88	170
MSE (error goal: 0.1)	0.078/356			0.090/401		
% of correct prediction	89%			82.3%		

Table 8.18 Prediction Results by ANN Model for SP 4

Observed	Predicted by ANN					
	Training Data			Test Data		
	TC	SC	Total	TC	SC	Total
TC	665	127	792	67	17	84
SC	45	483	528	5	56	61
Total	710	610	1320	72	73	145
MSE (error goal: 0.1)	0.063/326			0.082/389		
% of correct prediction	87%			84.8%		

Compared with the RP model, all SP models can stop training within the maximum training epoch (most models trained the network for 300-400 times when getting the optimum output) and obtain the lower training error than the error goal. Therefore, the model performance of ANN models for the SP data is better than the RP data. It can also be implied that although neural network models can be capable of modelling non-linear mapping relationship between input and output, with the increase of complexity of network structure (e.g., increasing the number of input vectors), the model performance and generalisation ability may be impacted.

8.4. Comparisons between MNL, FL and ANN Models

This chapter and Chapter 7 have modelled the discrete choice data with different approaches. The final purpose of using different modelling techniques in this research is to explore and compare the model performance and forecasting ability between MNL, FL and ANN models. The comparisons are discussed from the following aspects in this section:

- General model expression and estimation algorithm;
- Model forecasting ability;
- The goodness of fit measurement; and
- Interpretability of outputs.

8.4.1 General Model Expression of MNL, FL and ANN

The main difference between MNL models in Chapter 7 and two techniques (FL and ANN) in this chapter exists in the model expression. For MNL models and other model expression in the logit model family, the basic assumption is the RUT and utility functions, which contain the linear additive expression for deterministic variables of fare payment alternatives. Therefore, MNL models can be transparent and easily interpretable by those variables and coefficients in the utility models.

Compared with MNL models, FL and ANN models do not require a function to indicate the relationship between deterministic variables and decision making (i.e., relationship between V_i and U_i in the utility function). An FL model is based on the expert system (IF-THEN rule) and fuzzy inference system, which are predefined by modellers. It assumes that individuals make their choices based on simple rules relating perceptions (of the attributes of the available alternatives) to preferences (towards them) both of which are modelled using fuzzy sets. The basic motivation of FL technique is to model human's linguistic and uncertain decision making. A fuzzy logic system is a nonlinear system that maps a crisp input vector into a crisp scalar output.

The fuzzy rule base in FL models presents input-output relationship to some extent, therefore, FL models are still based on some *a priori* assumptions between alternative attributes and decision making, although such relationship between input and output can be modelled by non-linearity in FL models. Compared with FL models, ANN models do not have any *a priori* assumptions for the mapping relationship between inputs and outputs. The basic assumption of ANN models is that the input-output relationship is unknown and the neural network can be trained by simulating the learning process of human's brain to capture the input-output mapping relationship, then the best fit relationship can be generated. Therefore, basic components in ANN models include the number of layers, neurons, how layers can be linked (transfer function) and how the network can be trained (learning function).

The differences on the model expression result in the different algorithms for MNL, FL and ANN models. The algorithm of logit models is that the maximum of a log likelihood function can be obtained by Newton-Raphson method, which uses first and second derivatives to reach the optimum. In case of a linear utility function, a unique maximum is guaranteed, if it exists. However, if a linear utility function is used to model a process that is based on an underlying non-linear function, the model will have its source of errors increased considerably. Unfortunately, if the utility function is non-linear in parameters, there is no guarantee that a single maximum exists, as the error surface can be non-convex. This implies that the Newton-Raphson algorithm may not converge on the maximum likelihood solution.

However, in order to obtain the single optimum result, the approximate reasoning algorithm (i.e., max-min gravity method in this research) in FL models does not necessarily require the linear additive expression between input and output, therefore, the algorithm can be capable of modelling any relationship between inputs and outputs, including the non-linearity. Similarly, ANN models rely on the network structure and transfer function (sigmoid function in this case), learning function, rather than the pre-defined linear utility function, therefore, non-linear mapping could be captured by ANN models (because ANN models can be regarded as a sort of 'black box system', it is very difficult to exactly know the model expression between inputs and outputs through ANN models). All in all, FL models and ANN models can be regarded as a sort of data-driven estimation technique.

8.4.2 Model Forecasting Ability

In this section, percentage of correct prediction is used to compare the forecasting ability of MNL, FL and ANN models, because the purpose of introducing FL and ANN techniques in this research is to explore the forecasting ability of new models. In this case, results of MNL models can be viewed as the benchmark when making the comparison.

The percentage of correct prediction for the validation data by MNL models is around 81.4%. Through comparing with FL models, we can see that in general FL model with error inputs can almost achieve better forecast results than the MNL models. However, the percentage of correct prediction of FL model without error input is not better than MNL models and FL models with error input. Therefore, we can conclude that not only for MNL models but also for FL models, the error term, containing unobserved factors, plays an important role when modelling human's decision making.

Percentages of correct prediction by ANN models show that the forecasting ability of ANN models is better than MNL models (overall 81.4% correct prediction in MNL models; 85.9% correct prediction in ANN models). A possible reason to obtain the improved forecast results by ANN models is because there is no *a priori* assumption for the mapping relationship between inputs and outputs, the network can thus learn through the input and output data. During the learning and training process, the non-linearity can be captured by the network.

However, a common feature existing in MNL, FL and ANN models is that a number of 50%-50% choice probabilities (outputs) occurred when calculating individual choice probabilities. Such kind of result indicates no preference between two alternatives. But it is not surprising in the real life that respondents could have an equal preference to any alternatives. For this reason, it can be viewed as future research to investigate whether respondents would have certain preferences to alternatives or not in stated choice surveys.

8.4.3 The Goodness of Fit Measures

Regarding the goodness of fit, MNL model can be measured by Rho-squared value with respect to constant and the likelihood ratio test. Through checking the goodness of fit of MNL models, we find that the good model fits exist in the estimated MNL models, because most ρ^2 values range between 0.2-0.4. However, in order to make the model fit measures comparable with FL and ANN models in this chapter, some other measurements are introduced and compared, including: MSE (mean squared error), VSE (variance of squared error) and MMS (mean of market share). These three measures can be used to test the error during the model estimation (difference between modelled data by MNL, or FL or ANN models and actual data).

The goodness of fit measures for the RP and SP data are listed in Table 8.19 and 8.20. Measures are compared among actual data, MNL model, FL model and ANN model. In addition, because during the data analysis, in order to carry out model validation, the full data set was divided into two sub-sets, training data and test data (i.e., main data and validation data in MNL model), comparisons of the goodness of fit also are made between these two data sets.

Table 8.19 Comparison of Goodness of Fit of MNL, FL and ANN Models: RP

	Training Data			Test Data		
	MSE	MMS	VSE	MSE	MMS	VSE
Actual data	---	0.361	---	---	0.370	---
MNL model	0.2035	0.406	0.322	0.2806	0.433	0.341
FL model	0.139	0.383	0.201	0.1753	0.389	0.229
ANN model	0.137	0.372	0.186	0.169	0.383	0.222

Table 8.20 Comparison of Goodness of Fit of MNL, FL and ANN Models: SP

	Training Data			Test Data		
	MSE	MMS	VSE	MSE	MMS	VSE
Actual data	---	0.322	---	---	0.328	---
MNL model	0.1106	0.341	0.219	0.1522	0.358	0.283
FL model	0.077	0.336	0.191	0.1029	0.34	0.216
ANN model	0.075	0.331	0.192	0.09	0.34	0.203

Through comparing measures of the goodness of fit, we can see that MSE, MMS and VSE in FL and ANN models are lower than in MNL models, indicating that FL and ANN models offer advantages over the conventional logit models on the goodness of fit, not only in the training data but also in the test data. The goodness of fit measures between FL and ANN models indicate that in general the model fit in ANN models is better than FL models.

8.4.4 Interpretability of Outcomes

Due to the different model expressions for MNL, FL and ANN methods as discussed in Section 8.4.1, outputs also are different between these models. Because utility functions in MNL models are transparent and expressed by the utility function, the output in MNL models are also more various and interpretable than FL and ANN models. The direct outputs from the MNL models by using ALOGIT, which have been discussed in the previous chapter, include the following results:

- Coefficient estimates: can be interpreted by the sign and size to show the effect of alternative attribute-level to the relative utility. The choice behaviour and perceptions can be explained by each estimated coefficients, such as parameters of dummy variables.

- T-statistics and standard errors: can be used to test the statistical significance of each estimated parameter in the utility model under a given confidence interval.
- Log-likelihood measures: can be used to test the statistical significance for a set of coefficients of explanatory variables in a utility model.
- Rho-squared value for goodness of fit measurement: plays the same role as r-squared value in regression models. In the logit model, ρ^2 value ranging between 0.2-0.4 is regarded as a good model fit like 0.7-0.9 of R^2 in regression models.

Except direct outputs, behavioural outputs, such as valuation of attribute, demand elasticities, individual choice probabilities and predicted market shares in aggregation can also be obtained based on estimated coefficients in the MNL models. All of these in the MNL models can be used to analyse respondents' choice behaviour and measure the model performance.

Compared with a variety of outputs in the MNL models, the principal output in FL and ANN models for this research is values ranging between 0 and 1, which can be regarded as individual choice probabilities for different alternatives. Outputs in FL and ANN models can be used to compare individual choice probabilities in the MNL models.

In previous studies, artificial neural network models have often been criticised for their lack of interpretability (Bentz, and Merunka, 1998; Carvalho, et al, 1998; Sayde and Razavi, 2000). Unlike logit modelling, the analysis of network parameters does not reveal anything useful about the fitted function, except for very simple networks (e.g. perceptron with no hidden units). However, in the MNL model, through viewing the estimated coefficients (i.e., sign and size of coefficient), effects of attributes on individual choices can be obtained. Although the direct effect of variables on user demand cannot be obtained in ANN models, we still could capture the relationship between inputs and outputs through the simple data-sorting method and it also could partly explain criticism from the previous studies about the lack of interpretability of ANN models. But it should be noted that data-sorting method is suitable for continuous (quantitative) variables in this research to capture effects of input on individual choices, such as travel cost, because coefficients of qualitative variables in the logit models, coded by dummy variables, only reflect the relative effects between the dummies and the base. The data-sorting result in ANN models cannot show the relativities of dummy variables as the logit models have done. Therefore, the data-sorting method can only explain some of variables in this research.

Figure 8.8 illustrates effects of travel cost of the three payment methods on individual choices. We can observe that the effect of travel cost for three payment methods on individual choices are all negative (the slope of the fitted lines), corresponding to the sign of the cost variable in the logit models. However, R squared values in Figure 8.8 are low. Therefore, although the data-sorting method can partly explain the relationship between inputs and outputs, the interpretability of the ANN models are still not as satisfactory as the logit models.

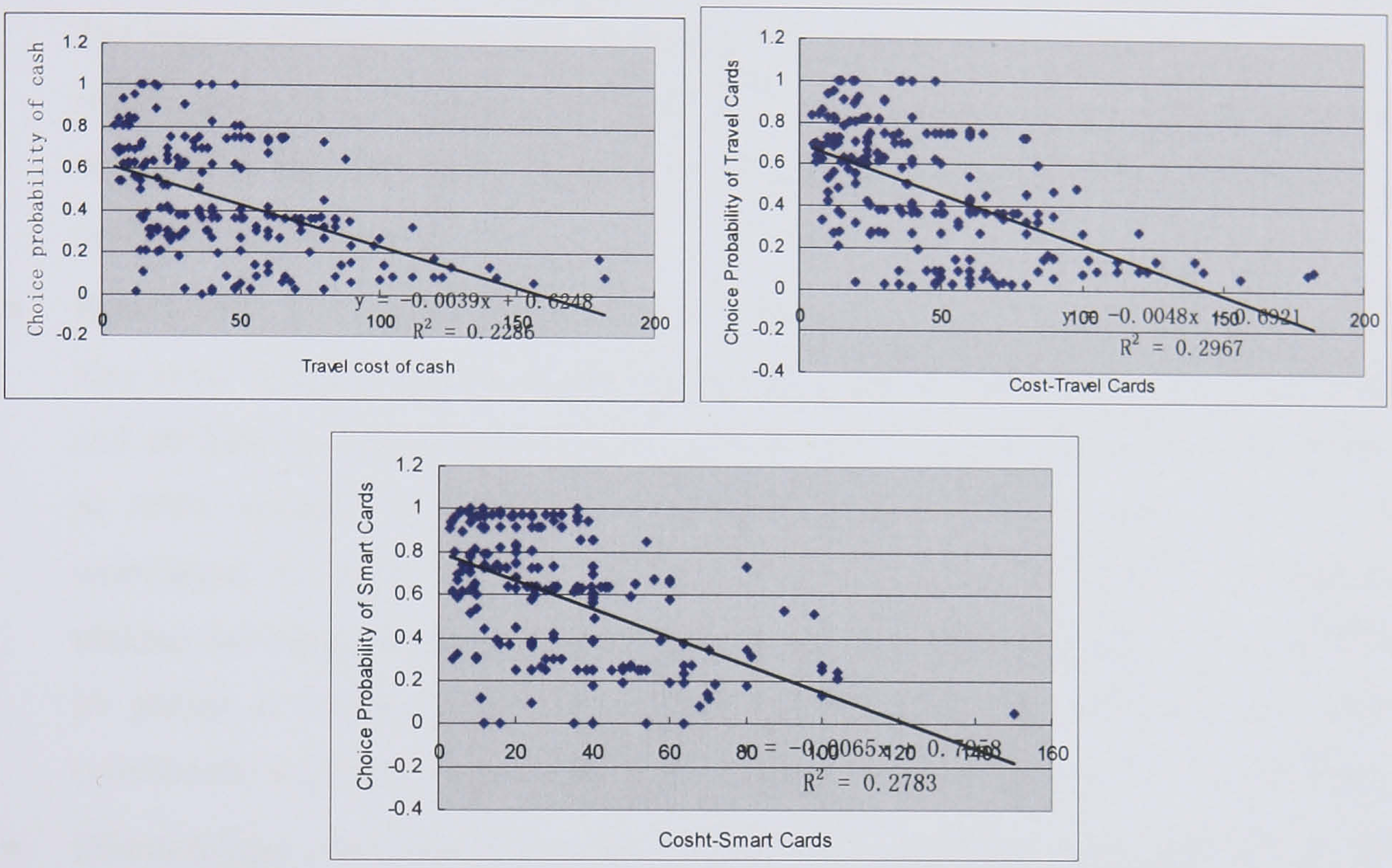


Figure 8.8 Effects of Travel Cost on Choices

Because in the logit model analysis, individual relative utility (*Vi*) can be calculated, according to the definition of elasticities, cost elasticities can be obtained in the logit models through derivatives. However, in FL and ANN models, the relationship between inputs and outputs cannot be expressed like the logit models, therefore, derivative values of the model expression cannot be calculated by the FL and ANN techniques. So in this research cost elasticities are only calculated in the logit models.

Except choice probabilities in FL and ANN models, error for checking the goodness of fit can be obtained in FL and ANN models, such as MSE. However, these measures cannot play the role as important as Rho-squared value in the MNL models, because ρ^2 can be measured by the empirical criterion of [0.2-0.4] to determine the model fit. In FL and ANN models, there is no a criterion to evaluate whether the well-estimated FL model or trained network has a good model fit or not. Comparisons must be made to reveal the model performance of FL or ANN models, for example, comparing the goodness of fit in different models (MNL, FL and ANN) as discussed in Table 8.19 and 8.20.

8.5. Findings and Conclusions

In this chapter, fuzzy logic and neural network techniques are explored as an alternative method to model discrete choice data. It is expected to provide additional explanation of benefits of smart cards to PT users. Through the analysis, the following findings and conclusions can be summarised as follows:

- Advantages of ANN models: through comparing forecasted market shares by three different models, we can conclude that the ANN technique shows the best performance. The biggest difference of ANN technique from MNL and FL models is that it does not require any *a priori* assumptions about the mapping relationship between input and output data. The network is capable of capturing such mapping relationship based on the high non-linearity.
- Results from FL models: compared with the MNL models, FL models in this research also show improvement on forecasting ability depending on the model specification and attribute-level taken into account. But results in FL models are not as good as those in ANN models. However, the advantages of FL technique still should not be overlooked. FL technique is suitable for modelling human's uncertainty of decision making and vague concepts. Secondly, another attractiveness of the FL technique lies in its ability to model the decision process as a non-linear combination of various considerations (rules), each of which deals with a different aspect of the overall choice.
- Disadvantages of FL and ANN models: the main disadvantage in FL and ANN models is the lack of interpretability of some parameters. Unlike MNL models with a variety of outputs, the outcomes of FL and ANN models are relatively simple, because only individual choice probabilities can be output (under the specific definition of outputs in this research). Therefore, it does not seem that FL and ANN could provide the detailed explanation about the perception of respondents to each attributes through estimated coefficients as MNL models (i.e., the sign and size of estimates). Indeed, information in a neural network is processed in a complete delocalised way. Furthermore, degrees of freedom are often large enough to allow the network to fit the same function with different combinations of parameters. This is probably the reason why neural networks have been called 'black boxes', capable of mimicking relationships between a set of variables but incapable of explaining the nature of these relationships. However, in this research, through the simple data-sorting method, a general relationship between inputs and outputs could be partly captured. Moreover, it intends to explain effects of inputs on outputs, which could be indirectly interpretable like estimated coefficients in the logit models, although the limitation of this method is that it may be suitable for continuous (quantitative) variables.
- In this research, it is found that during the model calibration for FL models. the shape of membership functions and location of membership curve has influenced the model performance. Smooth curve (e.g., Gaussian type) may perform better than straight line

membership (e.g., triangular type). In addition, the distribution of input data sometimes decides the proper location of membership curve. The interval for the majority of input data needs to be captured so as to determine the peak point of membership curve.

- Determination of network structure is another potential issue in ANN technique, particularly for selecting the number of neurons in the hidden layer, because there is no an effective solution available to this problem, except trial and error method. Therefore, the drawback is that it is somewhat time-consuming to optimise the configuration of the network.
- Outcomes of demand forecast from FL/ANN models can support the analysis result in Chapter 7, related to the evaluation of benefits and effectiveness of smart cards. From the forecast results, it can be seen that for the RP data market shares of travel cards and smart cards in the ANN model are slightly higher than in the FL model. For the SP data, the use of card payment options is much more than the cash fare payment, while shares between travel cards and smart cards are very close to each other. Therefore, based on the model assumption of ANN model (non-linearity between inputs and outputs and self-learning process), we can obtain that users' perceptions toward smart cards would be more than toward other two fare payment options.

All in all, FL and ANN techniques present a new direction to model discrete choice data, particularly when studying complex transportation systems that are potentially highly non-linear and it is very hard to develop a mathematical model. But the determination and calibration of fuzzy rules, and configuration of network structure are the two difficulties that are worth noting when we use these two techniques. Moreover, in order to obtain detailed explanation about choice behaviour, such as perceptions to specified attributes, or levels, the MNL model, which can be regarded as a 'white box' method, should be preferred firstly.

Finally, for further work, to investigate 'no preference' response behaviour can be regarded as one direction. In addition, socio-economic variables can also be introduced in the FL and ANN models to examine the effect of these variables on the choice behaviour.

Chapter 9
Conclusions

9.1. Introduction

This chapter describes the principal findings and contributions that have been made in the methodology and analysis about the evaluation of benefits and effectiveness of smart cards with the RP and SP data on choice perceptions of different fare payment means. According to the research objectives in Chapter 1, Chapter 2 – Chapter 8 have answered relevant questions including, the current use of smart cards and other fare payment applications; preference changes (market share) based on changes of payment features; importance of payment attributes of smart cards.

The conclusions can be divided into two sections in this chapter: first of all, the summary of achievements and key findings of this research are presented in Section 9.2. The achievements are discussed from the following aspects including:

- Benefit evaluation by results of fare payment choice behaviour with RP and SP data (including forecasted market share of different payment methods, valuation of attributes, travel cost elasticities, *etc*);
- Findings of the evaluation methodology; and
- Contributions to relevant policy-making to improve fare payment service quality of public transport.

The second part of this chapter is the recommendations for future work, which is presented in Section 9.3.

9.2. Summary of Achievements and Findings

9.2.1. Evaluation of Benefits of Electronic Fare Payment

Regarding the evaluation of benefits of smart cards and respondents' choice behaviour towards different payment methods, the achievements and findings of this research can be summarised as follows:

Market Share Forecast of Fare Payment Methods

The most straightforward measurement to evaluate the benefits of the smart card application is to forecast the market share of fare payment methods through modelling individual preference data, so as to show whether smart cards are preferred by respondents and how payment features of smart cards would influence respondents' choice behaviour (users' demand). However, in general the previous studies have done little on this aspect.

especially on the demand analysis with individual preference data. The study on PT users' demand of fare payment choices is the primary originality in this research.

In particular, as two alternatives to the smart cards, payment situations of traditional payment methods: cash and travel cards, may directly influence the choice of smart cards. That was the reason why the RP and SP surveys were used to investigate preferences toward different payment options, by presenting alternatives' attributes and levels. The previous studies have not explained respondents' choice behaviours by pooling RP and SP data when they evaluate benefits of smart cards. Therefore, the joint analysis of the RP and SP data to identify choice behaviours is also the originality of this research.

Two situations were considered in the demand forecast: demand changes by considering single factor and multiple factors. The purpose of considering two situations in the demand forecast analysis is to obtain changes of users' demand when payment factor(s) change(s). Eight smart card payment attributes were considered in the single factor scenario analysis. In each scenario, only one factor varied and the market shares of three payment options were obtained. Meanwhile, two 'multiple changes' of smart card attributes were employed (i.e., the best level of all smart card features and the medium level of all smart card features).

In the single factor analysis, for cost variable, market share of smart cards would vary from 84% to 0% when cost ranging between 20yuan and 120yuan. For other attributes, when the best levels presented to respondents, the forecast shares of smart cards in the market place are about 50% (e.g., 53.2% when deposit is 0yuan; 50.1% when unlimited PT routes can be covered by smart cards; 56.2% when smart cards could cover whole Liaoning Province, the widest area in the SP survey).

In the 'multiple change' analysis, the results indicate that when the best levels of all features of smart card payment are presented, 78.2% of PT users would choose smart cards, which is dominant over the two traditional payment options. Even when all attributes were applied by medium levels in the SP design, the market share of smart cards would exceed 50%. The results of the 'multiple changes' of smart card attributes indicate that most PT users would like to choose smart cards when the facilities of smart cards were kept in the best or the medium level. That is to say, the magnitude of switching between smart cards and traditional payment options in the two scenarios of the 'multiple changes' analysis is also great. From the forecast results, it also can be concluded that these attributes of smart cards would be highly perceived by PT users. Relevant improvement and enhancement on the fare payment applications would trigger changes of user demand.

In addition to the demand forecast, respondents' perceptions of payment attributes when trading off can also be used to identify the importance of smart card features. Through the single factor analysis, it was found that 'travel cost', 'multifunction' and 'geographic areas covered' would be the three most important attributes among these smart card features to cause a large change of smart cards use, because the variations of market share of smart cards for these three features are relatively more than other features (shares of smart cards ranging between 0-84% when 'travel cost' changing; 37.1% and 58.3% for 'multifunction' changing; 36.6% and 56.2% for 'geographic areas covered' changing). For travel cost, because discounted fare policy is most applied in the smart card schemes, this would be the key benefits to card users. Regarding 'multifunction' and 'geographic areas covered', the main reason for using smart cards is its convenience to users when a variety of add-in services and wider areas can be covered by one card.

Valuation of Attributes

In this research, benefits of smart cards were measured through measuring value of attribute. The originality to introduce valuation of attributes is to explain PT users' willingness to pay for different payment services in terms of monetary value, which has not been measured in the previous studies. Through valuations of attributes, importance of attributes also was explained in this research.

Valuation of attribute is divided into two parts: valuation of boarding time savings (VOBTS) and valuation of other attributes. Firstly, VOBTSs in three models (pure RP, pure SP and joint RP/SP) were compared. The VOBTS of the pure RP model (3.6yuan/min) is lower than that of the pure SP model (6.06yuan/min) and the joint RP/SP (4.45yuan/min) model, which indicates that in the SP model, more respondents perceived the better services and are willing to pay more than the RP survey for these payment services (features of payment).

Secondly, valuations of other attributes (qualitative) are obtained to examine respondents' willingness to pay. It can be seen that with payment attributes becoming better, respondents would like to pay more for using this payment alternative, for example, valuations of three multifunction levels (from the least multifunction to the most) are 2.32yuan, 6.37yuan and 8.9yuan. Among these qualitative attributes, the most perceived attributes by respondents would be 'top-up/purchase methods', 'multifunction', 'geographic areas covered' of smart cards and 'PT service routes covered by travel cards', because categories of these attributes have relatively high monetary valuations, compared to others. The reason for respondents' higher willingness to pay for these perceived attributes would be that these attributes can bring the most convenience to card users in day-to-day life.

Moreover, monetary valuation of attributes can be used to measure un-quantified benefits, called 'soft benefits'. The concept of 'soft benefit' has been raised in the previous study (Mulley, et al, 2004). 'Soft benefit' usually means intangible benefit, unlikely benefit or difficult to measure in financial term (Schmidt, 2006). The soft benefits of smart cards in this research is the payment convenience and improvement of personal life quality for PT users due to these add-on features in the smart card ticketing, such as multifunction, wider geographic areas covered, various top-up options, etc.

In this research, soft benefits of smart cards can be measured by individual utilities of choosing different payment options (smart cards and traditional payment methods), in which features of payment applications were included, such as these factors that cannot be directly measured but regarded as indicator of 'soft benefit'. All qualitative attributes in the smart card application can be viewed as 'soft benefit' indicators, because convenience for PT users when using smart card with these add-on features is not easily quantified, including, PT routes covered by smart cards, multifunction of smart cards; overdraft, geographic areas covered by smart cards and top-up options for smart cards. Different values of attributes (dummy variables) can be used to explain the 'soft benefits' of smart cards. For example, values of 'geographic areas covered by smart cards' for three dummy variables are 3.42yuan, 5.9yuan and 8.05yuan respectively (See Table 7.20). Therefore, when respondents chose smart cards, which can be used in the widest areas (the whole province), their willingness to pay is 8.05yuan. It also means that the benefit of smart cards with this level is equivalent to 8.05yuan in card users' perception, relative to the base of the attribute (i.e., the smart card can only cover Dalian urban area). So another purpose of monetary valuation in this research is to measure the 'soft benefits' of smart cards to individual users.

Segmentation Analysis

Moreover, benefits of smart cards were explained by segmentation analysis in this research. Also, the segmentation analysis gives us an insight into the effect of socio-economic variables on passengers' payment choices. It is helpful to examine the heterogeneity and homogeneity of individual preferences in the survey (i.e., whether different groups of respondents would have the same perceptions toward the smart card ticketing or not). Three socio-economic variables were used to segment the whole data set, including age, sex and household income. Generally, analysis results can reflect the following relationship between respondents' choices and their socio-economic backgrounds:

- For age factor, the valuations of attributes increase with age (e.g., 3.96yuan/min for aged 16-25, 4.31yuan/min for aged 26-35, 4.59yuan/min for aged 36-45 and 4.72yuan/min for over 45), and the younger respondents have the lowest willingness to pay in these four age groups, while for people aged over 45, they would like to pay more for the convenience of payment services.
- Sex factor also has the strong influence on the value of attributes. Male respondents would like to pay more for better PT services and payment convenience than female

respondents, for example, the valuation of multifunction with the most level for males is about 9.22yuan but females' is about 8.73yuan.

- People in the lowest income group would have the lowest willingness to pay. Meanwhile, it can be seen that respondents' choice behaviour would clearly begin to change between the group of less than 1500yuan and 1500-2999yuan, for example, for PT service covered, multifunction, overdraft, geographic areas covered and top-up options, valuations of these attributes for the group of less than 1500yuan were lower than the average values of non-segmentation, while valuations of attributes for the group of 1500-2999yuan were greater than the average values. But for VOBTS, it seems that the group of 3000-3999yuan and over 4000yuan would have greater willingness to pay (3.78yuan/min and 4.38yuan/min) than the average value (4.45yuan/min). On the other hand, due to the correlation existing between age and income, sex and income to some extent, it could partly explain the heterogeneity of choices between different age groups, between the two sex groups.

Fare Elasticities

Besides studying users' demand changes and valuation of attributes to measure benefits of smart cards, in this research fare elasticities also were taken into account as an original work compared to previous evaluation studies on smart cards.

Fare elasticities in this research are of two types, own and cross elasticities, to produce different measurements to explain the sensitivity of demand of payment options toward travel cost. It can be found that all absolute values of own elasticities of alternatives with respect to their own travel cost are greater than 1.0 (for example, in the RP/SP model, fare elasticities for cash, travel cards and smart cards are -1.9621, -1.4162 and -1.4637 respectively), indicating that the change of travel cost of payment alternatives is very sensitive to the demand for using this payment method. Secondly, cross elasticities greater than 1.0 of cash payment (e.g., 1.2606 for cash with respect to travel cards; 1.6091 for cash with respect to smart cards in the RP/SP model) tell us that cash payment users are also subject to changes of travel cost of the two card payment methods except changes of the own travel cost of cash. On the contrary, demand for cashless payment methods is relatively stable, because all cross elasticites of cashless payment methods with respect to change of travel cost of cash are less than 1.0. Thirdly, because two card payment methods were similar in some features and strongly substitutable to each other, cross elasticites (greater than 1.0) also show that card users would readily switch from one card payment method to another.

In order to explain the competition among three payment options and reveal the influence on utility of choosing different payment options, it is necessary to compare the public transport demand elasticities in China. When comparing with the public transport demand elasticities in China, we find that the elasticities of fare payment demand are relatively higher than PT demand elasticities when allowing for mode switching (PT demand elasticities varying -0.2~ -0.5 in Chinese cities, such as -0.45 in Beijing and -0.30 in Nanjing (Beijing Transport Bureau, 2005; Nanjing Public Transport Development Report, 2006)). Therefore, it can be concluded that compared with the competition between different transport modes, the competition between different fare payment options are much greater. PT users' perceptions are more sensitive to changes of travel cost than to switches between public transport and other modes.

9.2.2. Findings of Evaluation Methodology

In this research, the discrete choice model for analysing preference data so as to obtain the evaluation results was primarily used. Meanwhile, as an alternative to the discrete choice model, two techniques, fuzzy logic and neural network, also were explored to examine the model performance when carrying out the behavioural analysis. Through the comparison of two different modelling mechanisms (logit model and FL/ANN), the following findings and implications are summarised:

About Preference data and Discrete Choice Models

- A variety of measurements of benefits by discrete choice models

Discrete choice models combined all deterministic factors related to fare payment applications into one model, therefore the benefit (in terms of individual 'utility') of smart cards can be viewed as a comprehensive result, decided by a set of attributes, rather than a single factor, and the interaction between payment features when respondents make decision. Eventually individual utilities (benefits) of choosing different payment options were aggregated as 'market shares'.

Moreover, as can be seen in Chapter 7, the benefits of smart cards were measured by market share forecast, value of boarding time savings and other attributes, fare elasticities. In the meantime, the segmentation analysis explained respondents' perceptions and willingness to pay for different groups.

- Users' preference data

When measuring benefits of smart cards, the data sources decide what kinds of outputs we could obtain. One of objectives of this research is to measure benefits from demand side (PT users), therefore, users' perceptions from the survey may be a direct data source for this purpose. One of data sets used in this research is SP data. When we forecasted users' demand, it can be found that the model results have given a sufficient explanation about users' demand based on changes of payment applications. The forecast result directly reflected benefits of different payment options to users when some attributes (payment features) changed.

Therefore, for the objectives of this research, the outcomes from the discrete choice models with preference data have explained benefits of smart cards to PT users.

About FL and ANN Techniques

Fuzzy Logic Model

In the FL models, two different FL models are used and compared with MNL models: FL without error input and FL with error input. For FL with error input, we consider two kinds of error distribution types in this thesis, Gumbel distribution and normal distribution.

Findings through FL models are:

- FL models with error input (Gumbel and normal distribution type) can achieve higher percentage of correct prediction than FL model without error input. Therefore, it can be concluded that like the error term in the utility function, the error input in the FL models does influence human's decision making besides those attributes we have included as deterministic factors in decision making process.
- Between the FL model with Gumbel distribution type and normal distribution type, they present almost the same results (including the percentage of correct prediction and forecasted market share). Therefore, it can be implied that based upon the fuzzy rule base, both distribution types for the error input have the same effect on the model forecasting ability.
- By comparing the measurement of model fit (RMSE: root mean square error), it can be found that RMSEs by FL model with normal distribution are slightly lower than FL model with Gumbel distribution, indicating FL with normal distribution has the relatively better model fit than FL with Gumbel distribution.
- Compared with estimated results from MNL models, in general the calibrated FL model with error input for the RP and SP data obtains better results (percentage of correct prediction). FL models without error input for the RP and SP data cannot achieve the better model performance on forecasting market share than MNL models.
- Through calibrating outputs of FL models, we find that shape of membership function and location of membership curve has influenced the model performance. So when designing a FL model, suitable membership functions and proper locations of membership curve must be carefully determined.

Artificial Neural Network Model

In general, ANN models achieve satisfactory results for predicted market shares and model performance in the both RP and SP data, particularly in the SP data (including training and test data) all percentages of correct prediction are above 80%, higher than the logit models and fuzzy logic models. In the RP data, the training data obtain 82% correct prediction, much higher than the results in the logit model. It also can be proved that in this case non-linear functions can output better results on modelling respondents' choice behaviour and predicting market share of fare payment methods. However, the results of the ANN model using the RP test data present almost the same results as the logit model. Therefore, it can be concluded that the generalisation ability of network can also be affected by the number of inputs in the network (in this research, the RP ANN model contains much more inputs than the SP ANN model).

Meanwhile, the goodness of fit measures, including MSE, MMS and VSE, between MNL, FL and ANN models are discussed and compared. Through comparing measures of the goodness of fit, we can see that MSE, MMS and VSE in FL and ANN models are lower than in MNL models, indicating that FL and ANN models offer advantages over the conventional logit models on the goodness of fit, not only in the training data but also in the test data. The goodness of fit measures between FL and ANN models indicate that in general the model fit in ANN models is better than FL models.

Differences between MNL and FL/ANN Forecasts

Through comparing results from logit models and FL/ANN models, we find that two differences on the demand forecast between MNL and FL/ANN models are:

- Different estimation mechanisms: MNL models assume that the error term is Gumbel distributed. And the model estimation is based on the utility model, in which the systematic term and error term are presented respectively. The maximum likelihood method is applied during the coefficient estimation. However, a model expression was not given in the FL and ANN models. In the FL models, the pre-defined fuzzy inference system controls the whole estimation process. In the ANN models, the pre-defined network is trained by the input and target values to capture the mapping relationship between inputs and outputs. The estimation mechanism of FL/ANN technique is capable of capturing the non-linearity and uncertainty of decision-making.
- Different model outputs: Due to the transparent model expression of MNL models, the forecast results in MNL models are more various than the FL and ANN models, such as elasticities, valuation of attributes, coefficients of attributes, etc., while the FL and ANN models in this research only output individual choice probabilities.

Discussions of Wider Evaluation Methodology

Discrete choice models have been primarily used to analyse users' demand and evaluate benefits of smart cards to PT users. As an exploration, FL/ANN also could present a future direction to model discrete choice data. But the literature review of this research have provided some potential evaluation techniques that would be useful for measuring benefits of smart cards, for example, analysis of smart card operational data. Therefore, a discussion of wider evaluation methodology is summarised so as to understand the suitability of different methods for different purposes.

Analysis on Operational Data

As discussed in Bagchi and White (2004) and Bryan and Blythe (2007)'s work, operational data for smart card use can be used to track PT users' day-to-day travel behaviour and their boarding and alighting points during their journey, therefore, the PT users' demand can be obtained based on users' O-D data. Because of the discount policy in smart card schemes, such demand analysis result can be used to measure PT users' preference and perceptions toward different payment applications in different PT routes, such as what the proportions of using smart cards, cash and travel cards in a given bus route would be and why. However, the analysis would be based on a sufficient historical data requirement. When the data become available in Dalian, the result from analysing operational data also can verify the result of this research (i.e., demand forecast).

Cost-Benefit Analysis:

The evaluation of benefits of smart cards in this research focuses on users' demand forecast so as to explain respondents' perceptions toward different fare payment applications. Meanwhile, such choice behavioural forecast also considers situations (payment features) that would happen in the future. However, benefits of smart cards also could be examined by consumer surplus of CBA method.

General applications of CBA have been discussed in Chapter 3: literature review. CBA method is suitable for appraising a project from the society's point of view and taking account of costs and benefits whether or not they pass through the market.

For this research, the CBA method could focus on two aspects:

- Cost/benefits to PT operators, such as incremental revenue (additional rides, unused, residual value); improved cash flow (admin./labour cost saving, etc.); travel time savings;
- Cost/benefit to PT users, i.e., consumer surplus due to discounted fare, convenience, ticket purchasing time, travel time, etc.

Consumer surplus (CS), representing the difference between what consumers are willing to pay and what they actually pay. According to Nash (2000), this is usually based on the generalised cost of travel, which in turn is related to the concept of utility. However the main issue of CBA is to translate costs and benefits of smart cards into quantified monetary terms. It would be difficult to convert qualitative effects, such as payment convenience, environmental effects, to quantitative values for this research.

To sum up, considering the objectives of this research, discrete choice models may better fit the purpose of measuring benefits of smart cards from different angles (demand forecast, valuation of attributes and fare elasticities, etc.) than FL and ANN models. Various outputs of logit models can have an insight into each single payment attributes and respondents' perceptions towards changes of these attributes. Evaluation methods, such as analysis of operational data, CBA, could be viewed as further directions based on the current work to monitor behavioural changes of PT users, to explain consumer surplus.

9.2.3. Contributions to Smart Card Applications from Policy Angle

Through evaluating the benefits of smart cards to users with the users' preference data, the final aim of this research is to make suggestions for the reform and enhancement of PT fare payment services. Particularly for smart cards, the following contributions from the policy angle may be applied to the future development of the smart card ticketing.

New services being suggested (or enhanced)

As discussed in Chapter 7, through forecasting the market share of fare payment methods and measuring the valuation of attributes, we have determined the importance of attributes (levels) of smart card payment. Those attributes in the survey (particularly for the SP survey, some new features or levels based on the current applications were used), which are most preferred by respondents, may be regarded as the priority to enhance the PT payment service. These attributes include: multifunction, geographic areas covered, various top-up/purchase methods.

According to the SP survey design for "multifunction", the future development may focus on integration of banking, parking fee payment, highway tolling and small value consumption. For the banking function, it may be enhanced by the existing fare collection system, because in China, the final clearance for the PT fare collection under the smart card ticketing is conducted by local banks, rather than the public transport operators and smart cards companies. Therefore this would be a new mode to enhance the financial function of smart cards through add-in applications of local banks based on the current clearance system. For this issue, the London Oyster card has presented a very successful example. Since January 2007, the London Oyster card has cooperated with Barclays bank to enhance the payment function.

The feature of 'Geographic areas covered by smart cards' also can be regarded as a focus of the future smart card development. Through the preference study, it can be found that respondents much perceived the geographic coverage of smart cards in PT services.

Currently, most smart card applications in China can only cover a very limited geographic area, such as the urban area of a city. Therefore, the future development of smart cards on this point is to widen areas where smart cards can be used, furthermore, to achieve interoperability among different authorities or public transport operators. The final objective of improving this feature of smart cards is to provide seamless PT services and the most convenience to PT users when travelling across different cities. To achieve this objective, Chinese authorities may refer the successful applications in other countries. The successful example of the integrated smart card payment application is the Rhein-Ruhr smart card in Germany. PT users can travel by smart cards between different cities around the Rhein-Ruhr area, such as Essen, Dortmund, Bochum, Gelsenkirchen, etc.

Last but not least, with the increasing use of smart cards, the convenience of topping up smart cards has gradually become one of concerns among card users. Various top-up/purchase methods also can increase the attraction of smart cards and improve the service quality of public transport system as a whole. In the evaluation analysis, we find that the valuation of attribute of top-up/purchase methods is relatively higher than other attributes, which means that as well as the convenience of fare payment, most smart card users would like to see a variety of top-up/purchase options so that the convenience of smart card payment functions can be improved and the accessibility of smart cards can also be enhanced. However, the current application of smart cards in Dalian only has two topping up options: banks and ticket offices. The future development on this aspect can refer to other successful smart cards applications (e.g., automatic adding value machine for the Hong Kong Octopus card, mobile phone topping up for the Seoul 'T-Money' card, on line topping up for the London Oyster card, *etc*).

It is clear that smart cards will play a significant role in public transport ticketing for years to come. China is at the cusp of the large-scale deployment of smart cards for transport ticketing as well as a range of other services, both public authority-led as well as private-sector-led. Therefore, the key issue to achieve the objectives above is the interoperability of the smart card system across different social services and authorities. The local government should play an important role for promoting these new smart card payment services.

Fare policy to encourage the use of smart card payment

The second implication from the evaluation result in this research is the reform of fare policy in the current PT system of China. The demand forecast and fare elasticity analysis indicated that PT fare is the primary factor for all PT users to choose their payment methods and assess the relevant benefits to their own. To meet different users' needs, it would be of interest that PT operators and local authorities put emphasis on two aspects of fare policy reform. First of all, discounted fare policy is an effective way to attract passengers to use smart cards. One of examples for the discounted fare is the current discount fare for smart card users in Beijing public transport services. In Beijing, since January 2007, a new fare policy has been introduced for smart card users: 60%-off ticket for smart card users.

The second aspect on reforming fare policy is to introduce the differentiation between peak and off-peak fare. Fare differentiation is a solution to re-allocate revenue for operators, improve the crowding situation on board, in the mean time, to those passengers, who do not have departure time restriction (peak and off peak), fare differentiation may save their travel cost to some extent. In China, PT fare differentiation has not been implemented. Therefore, fare differentiation based on the current discounted fare policy could be one of major concerns for the smart card development in China.

In addition, with the promotion of the new pricing structure for smart cards and potential behavioural changes on fare payment habits, reform of traditional payment methods has also taken place in some cities. For example, in Beijing, China, since 1st January 2007, paper-based travel cards have been fully stopped being used in the urban public transport system. Alternatively, only two payment means are now available for PT users in Beijing: cash and smart cards.

Increasing users accessibility to PT services by different card products

As an issue of smart cards in China today, the simplicity of smart card products have impacted the accessibility of passengers to use PT services, to some extent. From the preference survey (RP and SP) of this study, it can be found that variety of smart cards may influence people's choices of payment methods and their travel behaviour and frequency, such as daily card, weekly card, the elderly card, student card, etc.

For example, smart cards for short period users would be suitable for visitors to the city. Monthly card may be suitable for frequent travellers, while the elderly and students may get more beneficial smart cards than other passengers due to the relevant subsidies from the local governments.

A possible direction to develop different card products in China is that "pay as you go" smart cards can be regarded as a basic product for any PT users. Based on this, daily/weekly/monthly cards may be introduced with unlimited travel frequency and service routes. If any travel zone restriction applied, like London, smart cards for any zones without any restriction could be the most expensive among all smart card products.

9.3. Recommendations for Future Work

Although considerable progress has been made in this research, some further improvements are needed in evaluation of the benefits of smart cards. The recommendation for future work can be discussed from the following aspects:

- Outlook of Smart Card Development

For the future development of smart cards, inter-operation between different social services and PT fare payment, between different authorities in different geographic areas would be one of direction in the coming years. However, as to benefits and effectiveness of interoperation of smart cards, this evaluation study has not given a detailed picture. Particularly for investment and effort to achieve such inter-operability, PT operators and local governments would more concern about if the investment could bring the improvement of service quality of public transport. Therefore, in order to measure this new application in smart cards, cost-benefit analysis (CBA) would be suitable.

- More Explanation of Benefits and Effectiveness from Demand Side

Based on the modelling analysis in this research, it would be worth monitoring changes of users' demand and payment behaviours when smart cards/ card products were introduced. The following aspects would be considered to have a much closer insight into such demand changes: PT users' O-D; travel frequency; departure time; changes of pre-payment behaviour (e.g., frequency, amount of payment, where and how to prepay, etc.); journey time savings and PT mode choices.

Secondly, in this research benefits and effectiveness of smart cards were mainly measured from the demand side, precisely only including PT users. However, with the improvement of smart card payment service (increasing payment convenience, discounted fare, and saved journey time, etc), whether non-PT users' perceptions towards public transport services would change and whether some modal shifting (between public transport and private transport) would happen, also could be one of aspects to enhance the evaluation analysis for smart cards.

- Evaluation from supply/operation side

In this research, the evaluation of smart cards was carried out based on the user demand. However, the smart card ticketing is not a stand-alone system. When some new applications, which may bring benefits to PT users, were proposed, whether such new applications would impact operators' performance (operational efficiency, revenue allocation, workload of PT staff, service frequency, changes on dwell time, etc.) and how, also need to be evaluated as supporting evidences of smart card evaluation. Therefore, another future work that can be done is to take into account the supply/operation side, so as to capture any supply changes of PT operators when some payment features altered. The evaluation analysis based on the demand side as well as the supply side can be regarded as an overall evaluation study of PT fare payment choices.

- "No preference" choice behaviour

Current discrete choice data, particularly for binary choice SP data we analysed, have not considered "no preference" situations, because in the data set, only data with preference (e.g., choosing 'A', otherwise 'B', if under binary choice situations) were included to obtain users' preference as input data to evaluate user benefits of smart cards. However, when observing the modelled choice probabilities, we find that there were a proportion of outputs having 50-50% predicted choice probabilities, quite different from their actual choices (e.g., very certain preference to a payment alternative). Therefore, as an indicator of measuring benefits of smart cards, market share forecasts might be biased due to the presence of 'half-half' responses on individual preferences. For respondents, sometimes "no preference to either alternative" is rational choice behaviour and such responses can also be modelled to identify respondents' preference. Therefore, some future work can be concerned about investigating "no preference" response.

- Evaluation Techniques

Two folds of future work on evaluation techniques would be of interest to enhance the results based on the logit model analysis: one is to look at operational database to obtain individual travel behaviour by using smart cards. The purpose of employing operational data in benefit evaluation is to verify the demand forecast by individual preference data in the PT passenger survey.

Because some evaluation work on supply side may be carried out in further study, another technique would focus on how to comprehensively take into account both the supply and demand side to measure benefits of smart cards. According to discussion in previous sections, CBA would be suitable for this purpose. CBA technique tries to quantify benefits and costs in terms of monetary value, meanwhile, in different designed scenarios the relationship between costs and benefits can be identified by some measurements, such as NPV, IRR, etc. The advantage of using evaluation techniques to measure not only demand but also supply side is to consider the interactive relationship between PT passengers and operators, because for any side (demand and supply) benefits of smart cards may be interacted by each other.

Besides some potential techniques that would be used in the future work, another further work would be the model validation. In Chapter 7, the model validation process mainly focused on the discussion of demand forecast under the stated preference situations. However, how the models can predict actual changes that occur (e.g., either introduction of smart cards or changes to them), could be furthered in the future work. In the mean time, the model validation also should cover FL and ANN models. The aim of using FL/ANN models is to see whether FL/ANN models would improve the forecast ability (compared with MNL models), therefore, in general the forecast results were only compared across different models (Logit, FL and ANN models), which can be regarded the external validation analysis. The internal validation of FL and ANN models could be discussed in future work.

9.4. Conclusions

This research has made a significant contribution to understanding the benefits and effectiveness of smart cards for public transport. The demand forecasting result may suggest the relevant policy making to enhance the public transport services. According to the research result and relevant discussion, the benefits and effectiveness of smart cards can be explained by the following aspects:

Benefits and effectiveness of smart cards were measured from users' demand side. The forecast market share, as one of benefit indicators, revealed that with improvement of fare payment features of smart cards, the use of smart cards would increase. Moreover, the forecast market share also indicated the competition between smart cards and traditional fare payment options. Cash fare users would be the primary potential non-smart card users to switch from cash to smart card payment when the smart card payment was enhanced.

Secondly, through the modelling analysis, importance of attributes of payment applications, which would be most preferred by respondents, was identified. Meanwhile the future direction for enhancing payment services can be identified through the importance of fare payment attributes. The most important features of smart cards, which would benefit PT users most, include multifunction, geographic areas covered and top-up options.

Value of boarding time savings and other attributes indicated the monetary benefits for each single attribute and respondents' willingness to pay. The segmentation analysis results specified these valuations by different groups of people. As can be seen, young respondents would have the lowest willingness to pay for add-on applications in smart cards. With the increase of household income, respondents' perceptions toward smart card services also would increase due to relatively high willingness to pay. Valuation of qualitative attributes in the smart card ticketing also were indirectly used to measure benefits that are not easily quantified, called 'soft benefits', because the convenience due to the improvement of smart card add-on applications (e.g., quicker boarding time, multifunction, a variety of top-up options, etc.) were transferred to respondents' monetary valuation.

In addition, benefits of smart cards were identified by fare elasticities. Fare elasticities revealed the sensitivity of user demand with respect to changes of travel cost of its own and other alternatives. Own elasticities were all greater than 1.0, which means user demand of three payment options with respect to changes of their own fare were elastic. Also, cross elasticities of cash fare with respect to smart cards were greater than 1.0, indicating that cash users would be more sensitive to changes of fare of smart cards and they would likely switch from their current payment method (cash) to smart cards when travel costs by smart cards decreased.

Finally, the evaluation of benefits of smart cards also gave some implications on relevant policy making to enhance the service quality of public transport systems and the future development of smart card applications. Two aspects in the smart card ticketing may become the focus in the future: interoperability and multiple applications of smart cards. These two aspects would provide conveniences for card users across different geographic areas and different social services.

Reference

Agrawal, D. and Schorling, C. (1996) Market Share Forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial logit Model, *Journal of Retailing*, Volume 72(4), pp. 383-407.

Akiyama, T., Tsuboi, H. and Matsuura, T. (1997), Description of Route Choice Behaviour by Fuzzy Neural Network, *Proceedings of Mini Euro Conference: Fuzzy Sets in Traffic and Transport Systems*, 1997.

Ampelas, A. (1998), Contactless Smartcards for Ticketing, Electronic Money and City Services, Conference: Automatic Fare Collection, UITP 1998, Bologna.

Ampt, E. (1990), Selected readings in transport survey methodology: edited proceedings of the 3rd International Conference on Survey Methods in Transportation, Washington D.C., January 5-7, 1990 / edited by Elizabeth S. Ampt, Anthony J. Richardson, Arnim H. Meyburg.

Andrle, S. (1997), Multipurpose Fare Media: Developments and Issues, Research Result Digest, TCRP, TRB, Washington D.C. 1997.

Asia Pacific Smart Card Association (2004), URL: http://www.apsca.org/, [Seen at Aug 2004].

Bagchi, M, and White, P.R. (2004), What Role for Smart Card Data From Bus Systems?, Municipal Engineer 157 Issue MEI, pp.39-46.

Barrett, G. (2003), A Few Notes on New York City's MetroCard Pricing, URL: http://www.worldnewyork.net/ [Seen at 20 July 2004].

Beaman, D (2004), SmartRider - a smartcard ticketing initiative at the cost of 10% of a new bus, Proceeding of European Transport Conference, Oct 2004, Strasbourg, France.

Beijing Statistical Yearbook (2005), compiled by Beijing Municipal Bureau of Statistics, China Statistics Press, Beijing, 2006.

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

Ben-Akiva, M. and Morikawa, T. (1990), Estimation of Mode Switching Models from Revealed Preferences and Stated Intentions, *Transportation Research* A, Special Issue, 24 (6), pp.485-495.

Ben-Akiva, M. (1993), Combining Revealed and Stated Preferences Data, *Marketing Letters*, 5:4 (1994), pp.335-350.

Bentz, Y. and Merunka, D. (1998) Neural Networks and the Multinomial Logit for Brand Choice Modelling: a Hybrid Approach, *Journal of Forecasting*, Vol.19, pp.177-200.

Berardinis, L.A. (1992) Untangling the mystery of neural networks. *Machine Design*, June 25, pp. 55-59.

Bergkvist, E. (2000) Estimating value of time and forecasting transport choice in road freight with a nonlinear profit specification: The logit model versus neural networks, Umea University, Sweden, 2000.

Bierlaire, M. (1998). Discrete choice models, in M. Labbé, G. et al (eds), Operations Research and Decision Aid Methodologies in Traffic and Transportation Management, Vol. 166 of NATO ASI Series, Series F: Computer and Systems Sciences, Springer Verlag, pp. 203-227.

Bishop, C. M. (1995) Neural Networks for Pattern Recognition, Oxford: Oxford University Press, 1995.

Bissell, A.J. and Ferguson, R.A. (1975). The Jackknife-Toy, Tool or Two-edged Weapon?, *The Statistician*, Vol. 24, No. 2, pp.79-100.

Bradley, M.A. (1988), Realism and Adaptation in Designing Hypothetical Travel Choice Concepts, Journal of Transport Economics and Policy, Vol. XXII, No.1, pp.121-137.

Bradley, M. A., and Daly, A. J. (1991). Estimation of Logit Choice Models Using Mixed Stated Preference and Revealed Preference Information. Paper presented to the6th International Conference on Travel Behaviour, Quebec, May 22–24, 1991.

Bradley, M.A. and A.J. Daly (1997) Estimation of logit choice models using mixed stated-preference and revealed-preference Information, in Stopher, P. and M.L. Gosselin (ed.)Understanding Travel Behaviour in an Era of Chance, 209-231, Pergamon, Oxford.

Blythe, P.T. (2000), Blythe, ITS Application in Public Transport: Improving the Service to the Transport System, *Journal of Advanced Transportation*, Vol.34, No.3, pp. 325-345.

Blythe, P. T. (2004), Improving Public Transport Ticketing through Smartcards, *Municipal Engineer* 157 Issue MEI, pp.47-54.

Borsch-Supan, A. (1986). Econometric Analysis of Discrete Choice. Berlin: Springer-Verlaag.

Bridle, J. (1990) Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition, in Neuro-Computing: Algorithms, Architectures and Applications, New York: Springer.

Byran, H. and Blyghe, P. T. (2007), Smartcard Data Analysis, Proceedings of 39th Annual UTSG Conference, January 2007, Leeds, UK.

Business Wire (2002), Cubic Receives $26 Million Contract to Deliver Smart Card-Based Public Transit Ticketing System for San Diego County, Sept 17, 2002.

Cantarella, G. E. and Fedele, V. (2003) Fuzzy Utility Theory for Analysing Discrete Choice Behaviour, *IEEE Proceedings of the Fourth International Symposium on Uncertainty Modelling and Analysis* (ISUMA'03).

Carvalho, M.C.M., Dougherty, M.S., Fowkes, A.S. and Wardman, M.R. (1998), Forecasting Travel Demand: A Comparison of Logit and Artificial Neural Network Methods, *Journal of the Operational Research Society*, Vol.49, No.7, pp.717-722.

Chambers, B. (1998), The Octopus Hong Kong's Contactless Smartcard Project, 2/98, *Public Transport International*, pp.14-19.

Chambers, B. (2001), Smartcard Ticketing Systems for Integrated Transport, UITP Asia Pacific Assembly, 2001 Kuala Lumpur, 2001.

Chan, P. M. L (2002), Smart Card—Service Providers' Perspective, AFC paper 2002, UITP.

Chen H. and Mao B. (2003), Policy Analysis of Urban Public Transport in Developing Countries: A Case Study of China, *Journal of Systems Science and Information*, Vol. 1, No. 3, pp. 383-399.

Cheung, F. (2003), Tripprtpas Smart Card Project: Lesson From the Netherlands, Transportation Research Record: Journal of the Transportation Research Board, No. 1887, TRB, National Research Council, Washington, D.C., 2004, pp. 147–152.

Cheung, F. (2005), Cost-Benefit Analysis of a Nationwide Smartcard for Public Transport in The Netherlands, the 85th TRB 2006 Annual Meeting, January 2006, Washington DC, USA.

Chira-Chavala, T. and Cofiman, B. (1996), Impacts of Smart Cards on Transit Operators: Evaluation of I-110 Corridor Smart Card Demonstration Project, UCB-ITS-PRR-96-17, California PATH Research Report, University of California, Berkeley.

Cirillo, C., Daly, A. and Lindveld, K. (1996). Eliminating Bias due to the Repeated Measurements Problem in SP data, Proceedings of 24th European Transport Forum (PTRC), Association for European Transport, Brunel University, England.

Dalian Statistical Yearbook (2005), compiled by Dalian Municipal Bureau of Statistics, China Statistics Press, Beijing, 2006.

Dalian Evening Post, (2006), "24 Standard PT Vehicle Unit per 10,000 residents in Dalian City: Ranked the Top Level in China", 16/12/2006.

Dalian Statistics Bureau (2003), Dalian Statistics Annual 2003, Dalian Press, Dalian.

Deakin, E. (2002a) Mainstreaming Intelligent Transportation Systems: I. Findings from a Survey of California Leaders. University of California Transportation Research Centre, Berkeley, California.

Deakin, E. (2002b) Mainstreaming Intelligent Transportation Systems: III. Interviews with National Experts. University of California Transportation Research Centre, Berkeley, California.

Demuth, H. and Beale, M. (1998), Neural Network Toolbox for Use with MATLAB, the MathWorks, Inc.

Department for Transport (2004), Validation of the National Car Ownership Models. URL: http://www.dft.gov.uk/pgr/economics/ntm/valiationofthenetionalcarow3039. [seen at Nov 2006].

Dougherty, M. and Joint, M. (1992) A behavioural model of driver route choice using neural networks. In: Ritchie SG & Hendrickson C Artificial Intelligence Applications in Transportation Engineering, pp.91-110.

Dublin Transportation Office (2006). DTO strategy (2000-2016). URL: http://www.dto.ie/

Dubois D. and Prade H. (1983). Ranking Fuzzy Numbers in the Setting of Possibility Theory. Info Sciences 30, pp.183-224.

eEurope Smartcards (2004), Open Smartcard Infrastructure for Europe, Volume 6, Part 3, http://www.eurosmart.com/Update/Download/February04/SINCE_survey.pdf, [Seen at June 2004]

Fleishman, D. et al (1996), Fare Policies, Structures and Technologies, TRB, National Academy Press, Washington DC.

Foote, P. J. and Darwin, G. S. (2002) Testing Customer Acceptance of SmartCards at the Chicago Transit Authority, the 81st Transportation Research Board Annual Meeting, Washington, DC. 13-17 January 2002.

Faghri, A. and Hua, J. (1992) Roadway seasonal classification using neural networks. In: Ritchie SG & Hendrickson C (eds) Artificial Intelligence Applications in Transportation Engineering, pp. 363-381.

Franklin, T. (1999), G-100U Smart Cards, Business Communications Company, Inc. Norwalk, 1999.

Fuzzy Logic Toolbox: for use with MATLAB, Version 2 (2002), the MathWorks, Inc.

Gerland, H. E. (1996), Intelligent on Board—Modern Approach to Transit Fleet Management, Proceedings of the 3rd World Congress on Intelligent Transport Systems, Orlando, USA, Oct, 1996.

Glazebrook, G. (1999), Intelligent Public Transport Systems—Recent Developments in Australia, Asia—Pacific Regional Meeting, International Union (Association) of Public Transport, Sydney, Australia.

Goulcher, A. and Ashmore, D. (2004), Serving the Casual User in User Friendly Smartcard System: An Australian Prospective, 7th UITP International Conference of Automatic Fare Collection, Bologna.

Giuliano, G. and Moore, J.E. and Golob, J. (2000), Integrated Smart Card Fare System, Transportation Research Record 1735, pp.138-146.

Hague Consulting Group (2002), ALOGIT 4.0EC.

Hao, X., Chen, H. and Wardman, M. (2004), A Review on the Development of Smartcards in Public Transport. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 4, No. 4, pp.32-40.

Hensher, D. A. (1986). Sequential and Full Information Maximum Likelihood Estimation of a Nested Logit Model, *Review of Economics and Statistics* LXVIII (4), 657–667.

Hensher, D. A. (1991). Efficient Estimation of Hierarchical Logit Mode Choice Models, Proc. of Japanese Society of Civil Engineers, No. 425/IV-14, January. 17–28.

Hensher, D.A. and Bradley, M.A. (1993) Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models, *Marketing Letters* 4:2 (1993), pp.139-151.

Hruschka, H., Fettes, W., Probst, M. and Mies, C. (2000) A Flexible Brand Choice Model Based on Neural Net Methodology: A Comparison to the Linear Utility Multinomial Logit Model and Its Latent Class Extension, *OR Spectrum*, Vol. 24: pp.127–143.

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Networks 3: pp.359–366

Hoogendoorn-Lanser and Hoogendoorn (2000) A Fuzzy Genetic Approach to Travel Choice Behaviour in Public Transport Networks, CD-ROM Preprints Transportation Research Board 79th Annual Conference, Washington.

ITSO (2006), Smartcards for Midlands franchises, November 2006.

Kamakura, W. A. and Russel, G. J. (1989) A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26: 379–390

Kim, J. (2001), China Road Safety: Past, Present and Future, Asian Development Bank 2001 (in Chinese).

Klir, G. J., Folger, T. A., (1988). Fuzzy Sets, Uncertainty and Information. Prentice-Hall, Englewood Cliffs, NJ.

Kocur, G., Adler, T., Hyman, W. and Aunet, B. (1982), Guide to Forecasting Travel Demand with Direct Utility Assessment, US DOT, Urban Mass Transportation Administration, Report UMTA-NH-11-0001-82-1, Washington DC.

Kores E. P. and Sheldon R. J. (1988), Stated preference methods: an introduction. *Journal of Transport Economics and Policy*. Volume XXII No. 1: pp.11-25.

Kosko, B., (1992). Neural Network and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, Prentice-Hall International Editions.

Laconte, P. (1998), Smartcards within ITS Systems and Services Development, 3rd International Conference on Automatic Fare Payment and Smart Ticketing, *UITP* 1998, Bologna.

Laezza, M. (2004), Experience from the San Francisco Bay Area Translink Project, Conference: Automatic Fare Collection, UITP 2004, Bologna.

Lee, B., Fujiwara, A., Zhang, J and Sugie, Y. (2003) A Hybrid Discrete Choice with Fuzziness, *Proceedings of European Transport Conference* 2003.

Local Transport Today (2006), Oyster Card Breakthrough as Barclaycard Visa Signs Up, 14 December 2006.

Lotan, T., and Koutsopoulos, H.N. (1993a), Models for Route Choice Behaviour in the Presence of Information Using Concepts from Fuzzy Set Theory and Approximate Reasoning, *Transportation* 20, pp. 129-155.

Lotan, T. and Koutsopoulos, H.N., (1993b). Models for route choice behaviour in the presence of information using concepts form fuzzy set theory and approximate reasoning. *Transportation* 20, pp.129–156.

Louviere, J.J., Hensher, D.A. and Swait,, J.D. (2000), *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, England.

MacDonald, N. (2000), Multipurpose Smart Cards in Transportation: Benefits and Barriers to Use, ITS Mainstreaming, The University California Transportation Centre.

Manski, C., (1973). The analysis of quantitative choice, Ph.D. dissertation, Department of Economics, MIT, Cambridge, USA.

Maxey, C. and Benjamin, P. (2001), Washington Metropolitan Transportation Authority – SmarTrip, Proceedings of APTA 2001 Rail Transit Conference, Washington DC, USA.

Meads, R.(1998), Customer and Business Issues of Smartcards, Conference: Automatic Fare Collection, UITP 1998, Bologna.

McFadden, D. (1981), Econometric Models of Probabilistic Choice, *Structural Analysis of Discrete Data*, C.F. Manski and D. McFadden (Eds.). Cambridge, MA: MIT Press, pp.198–272.

Mizutani, K. and Akiyama, T. (2000a) A Descriptive Hybrid Model of Modal Choice with using Fuzzy Reasoning, *Proceedings of the 4th Asian Fuzzy System Symposium*, Vol.1, pp. 593-598.

Mizutani, K. and Akiyama, T. (2000b) A Logit Model For Modal Choice With a Fuzzy Logic Utility Function, *Proceedings of the 2nd International Conference on Traffic and Transportation Studies*, pp. 311-318.

Mizutani, K. and Akiyama, T. (2001) Construction of Modal Choice Model with a Descriptive Utility Function using Fuzzy Reasoning, *Proceedings of the 9th IFSA Congress and 20th NAFIPS International Congress*, CD-ROM, No.149.

Mulley, C., Edwards, S., Blythe, P. and Lewis, P. (2004) Assessment Framework for the Evaluation of Soft Benefits of Smart Card Schemes. National Smart Card Project, Transport Operation Research Group, University of Newcastle upon Tyne.

MVA and PA (2003), Prestige/ITSO Convergence Route Map, Sponsored jointly by Department for Transport, U.K., Transport for London, and ATOC.

National Centre of ITS Engineering and Technology, URL: http://www.itsc.com.cn/188/readnews.asp?newsID=630, [26/02/2004] (in Chinese).

Nelson, J.D. et al., (2001). Intelligent Transport Systems Solutions in Transitional Countries: the Case of Korea, *Transport Review*, 2001, Vol.21, No.1, 51-74.

Nguyen, D., and B. Widrow (1990), "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," *Proceedings of the International Joint Conference on Neural Networks*, vol 3, pp. 21-26.

Nijkamp, P., Reggiani, A. and Tritapepe, T. (1997) Spatial Choice Behaviour: Logit Models and Neural Network Analysis, *The Annual of Regional Science*, Springer-Verlag 1997, pp.411-429.

Nottingham City Transport (2007), URL: http://www.nctx.co.uk/EasyRider/Easyrider.htm, [Seen at 12 December 2007].

Octopus for Tourists (2007), Octopus Cards Limited, URL: http://www.octopuscards.com/consumer/help/tourists/en/index.jsp [Seen as 10 December 2007].

Oorni, S., Korte, S., Koski, K. and Holm, C. (1997), Smart Card Based System in Finland. Mobility for Europe, 4th World Congress on Intelligent Transport Systems, Berlin, Oct, 1997.

Opurum, C.N. (2005), Evaluation of the Impact of Automated Fare Collection System on Rail Rapid Transit: The Case of New York, PhD thesis, University of Leeds, Leeds, UK.

Ortuzar, J.D. and Willumsen, L.G. (2001), *Modelling Transport*, 3rd Ed., Wiley, New York.

Pas, E. and Huber, J.C.(1992), Market segmentation analysis of potential inter-city rail travellers, *Transportation* Vol.19, pp.177-196, 1992.

Paynter, J. and Law, P. (2003), An Arm's Length Evaluation of Octopus, Working Paper, Department of Management Science and Information Systems, Auckland University, New Zealand, 2003.

Pearmain, D., Swanson, J., Kroes, E. and Bradley, M. (1991), *Stated Preference Techniques: A Guide to Practice*, Steer Davis Gleave and Hague Consulting Group, Hague.

Pedrycz, W., (1989). Fuzzy Control and Fuzzy Systems. John Wiley, New York.

Reggiani A, Romanelli R, Tritapepe T, Nijkamp, P. (1997) Neural Networks: an Overview and Applications in the Space-Economy, in: HimanenV, Nijkamp P, Reggiani A (eds) Neural Networks in Transport Systems. Avebury, Aldershot.

Residency/ Community, Gangnam Government, South Korea, URL: http://english.gangnam.go.kr/residency/trans/card.htm, [Seen at 12/07/2004]

Ritchie, S.G., Cheu, R.L. and Recker, W.W. (1992) Freeway incident detection using artificial neural networks In: Ritchie SG & Hendrickson C (eds) Artificial Intelligence Applications in Transportation Engineering, pp. 215-234.

Salonen, J. (2006), A Study of Smart Card Solutions, Challenge of eCitizen Promoting eGovernment Actions in European Cities, Tampere, Finland, May 2006.

Santamaria, G. (2003) The Artificial Neural Network in the Discrete Modal Choice, Catholic University of Leuven, Belgium.

Savage, T.J. (2001), Automated Fare Collection, New York City Transit, 2001.

Sayde, T. and Razavi, A. (2000), Comparison of Neural and Conventional Approaches to Mode Choice Analysis, *Journal of Computing in Civil Engineering*, January 2000, pp.23-30.

Segaran, C. and Sim, L. (2004), Moving from Magnetic to Smart Card: Singapore's Experience, 7th UITP International Conference on Automatic Fare Collection, Bologna, 4-6 February, 2004

Sugeno, M., (1985). Industrial Applications of Fuzzy Control. Elsevier, Amsterdam.

Swait, J., Louviere, J. and Williams, M. (1994) A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice. *Transportation*, 21:pp.135-152.

Teodorvic, D. (1999) Fuzzy logic systems for transportation engineering: the state of the art, *Transportation Research* Part A, Vol. 33 (1990), pp.337-364.

TransLink (2007), Smart Card Smart Travel, URL: http://www.translink.org/whatsTranslink.do, [Seen at 10 December 2007]

Transport for London, www.oystercard.com, [Seen at 21/06/2004]

Transport for London (2007), "New Deal Brings Oyster and Barclaycard Visa Onto on Card", URL: http://www.tfl.gov.uk/corporate/media/newscentre/3746.aspx, [Seen at Jan 2007].

Turksen and Wilson (1994), Consumer Preference Models: Fuzzy Theory Approach, Proceedings SPIE Vol. 2061, pp.203-211, Applications of Fuzzy Logic Technology, Bruno Bosacchi and James, C. Bezdek (Eds.).

US DOT (2003), ORAGNES Electronic Payment Systems Field Operational Test Evaluation, Volpe National Transportation Systems Centre, January 2003.

Vaughn, K.M., Abdel-Aty, M.A., Kitamura, R., Jovanis, P.P. and Yang, H. (1993) Experimental analysis and modelling of sequential route choice behaviour under ATIS in a simplistic traffic network. Transportation Research Board.

Vythoulkas,P.C. and Koutsopoulos, H.N. (2003) Modelling Discrete Choice Behaviour Using Concepts from Fuzzy Set Theory, Approximate Reasoning and Neural Networks, *Transportation Research* Part C, Vol. 11, pp.51–73.

Wang, M. et al. (2003), A Study of Smart Card Policy Implementation in Taipei, Working Paper, Shih Hsin University, 2003. (in Chinese).

Wardman, M. (1988), A Comparison of Revealed Preference and Stated Preference Methods of Travel Behaviour, *Journal of Transport Economics and Policy*, vol. XXII, No. 1, pp.71-92.

Whelan, G. A. (2003) Modelling Car Ownership in Great Britain, PhD Thesis, University of Leeds, Leeds.

Wikipedia (2006), URL: http://en.wikipedia.org/wiki/questionnaire, [Seen at Dec 2006]

Wilks, S.S. (1962) Mathematical Statistics, New York: Wiley, 1962.

Xinhua Daily Telegraph (2006), Peak Time Congestion of Public Transport in Beijing, 30/07/2006.(in Chinese)

Yang, H., Kitamura, R., Jovanis, P., Vaughn, K.M. and Abdel-Aty, M.A (1993) Exploration of Route Choice Behaviour with Advanced Traveller Information Using Neural Network Concepts, *Transportation*, Vol.20, pp.199-223.

Yang, H., Akiyama, T. and Sasaki, T. (1992) A neural network approach to the identification of real-time origin-destination flows from traffic counts. In: Ritchie SG & Hendrickson C (eds) Artificial Intelligence Applications in Transportation Engineering, pp. 253-269.

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3). pp.338-353.

Zadeh, L.A. (1973), Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. Systems*, Man, and Cybernetics, vol. 3, pp. 28-44, January 1973.

Appendix A: Comparisons of Different Features of Main Smart Card Applications in the World

Project	Card Type		Transport Mode	Card Volume or Transaction	Fare Policy and Structure	Discount Policy	Adding Value Method	Security	Multi-function
	Time based	User based							
Go-Card, USA	N/A	A, S, SC	Bus, rail	14k cards delivered and additional 50k by January, 1999	Flat fare, distance and zone-based fare structure	$20 discount for students	Vending machine	N/A	Parking, food, photocopies
TransLink, USA	N/A	SC, S, DI	Bus, rail, light rail, ferry	700k card users expected till 2004	Overdraft is permitted. Free transfer	20% off	Add value machine, ticket office, telephone, mail and Internet	Lost/stolen cards would be blocked through balance protection	Parking
Oyster cards, UK	W, M, Y, PP	A, S	Bus, tube, rail, tram	80k issued in 2002. By 2003 1.4 m Oyster cards in use	Zonal fare	70p instead of £1 adult single cash fare	Ticket office, telephone, Internet	Within 24hours cards can be stopped if stolen/lost	N/A
Smartcards in Milton Keynes UK	N/A	N/A	Bus	15k cardholders in 1990 and 40k in 1996	Zonal fare	N/A	Ticket office or by direct debit services	N/A	N/A
FirstCard Bradford, UK	W, M, Y	A, S, SC	Buses of First bus company	1.5k trial users in 1997 and 40k cardholders in 2003	Zonal fare, peak/off-peak fare, and free transfer	Distance-based: 100 miles for £1 bonus	On buses, travel centre, ticket office	Lost cards reported are replaced within 3 working days	N/A
Onelink, Australia	M, Y	S, DI, long term riders	Bus, tram, light rail, private bus	N/A	Time and zone based	Discounts applied	Add value machine, retail stores, Internet, telephone	Value can be transferred if lost or stolen	Shopping, telecommunication, parking, *etc*
Helsinki smartcards, Finland	14 days up to 2 years	N/A	Bus, coach, tram, train, taxi	1991-1992: 20k 2000: 300k	Zonal fare	50% off	Add value machine, bank, telephone and Internet	N/A	Multi-service
Gothenburg, Sweden	W, M,Y	A, S, SC	Bus, tram, train, ferry	N/A	Zonal fare, peak/off-peak fare, free transfer	Offer a reduction fare	Ticket offices, banks and Internet	Lost and stolen cards can be blacklisted	Parking, purchasing, *etc*
Calypso, France	D, W, M, Y, PP	A, S	Bus, light rail, railway, tube	40k cards in 1993	Zonal fare	Discount policy is applied	Add value machine, ticket office	A call system for emergency service	Retail stores, phones, parking

Railway Smartcards, Holland	W, M, PP	N/A	Bus, tram, underground, train	N/A	Zonal fare and distance-based fare	N/A	Add value machine, ticket office	N/A	Bicycle shed, parking
PayCard, Germany	PP	N/A	Bus, tram, tube, train	Monthly transactions- 97: 350k; 98: 1,100k; 99: 1,700k. Monthly active cards: 97: 200k; 98: 410k; 99: 490k	Zonal fare	50% off	Vending machine, ticket office	N/A	Parking, tolling, telephone and shopping
Dash Cards, Ireland	W, M	N/A	Bus	1, 540 cards were issued in trial scheme in 1994, 10,014 daily transactions	Zonal fare and distance-based fare	N/A	Add value machine, ticket office terminal, and bank.	N/A	Parking, tolling, telephone
T-Money, Korea	N/A	A, S, C	Bus, tube, tram, rail	100k in 1995, 2 million daily transactions in 1997	Free transfer with given period, zone- and distance-based fare	20% off for the age of 13-18; 100 Won saved per trip	Add value machine, kiosk, mobile-phone, Internet	N/A	Credit cards, telephone, ID and health insurance
Automated Fare Collection, Singapore	M, PP	A, S, SC	Bus, train, light rail	5m cards by 2002; >= 3k cards issued daily; 1.3m monthly transactions in 2002	Zonal fare, peak/off-peak fare differential, and free transfer	A small discount over cash fare	Add value machine, ticket office, Internet	N/A	National ID , transport access, retail store, *etc*.
Octopus, Hong Kong	PP	A, S, C, SC, P, T	Bus, light rail, train, tube, ferry	1997 about 3m cards, 5.8m in 1999, 6m in 2000, 10m in 2003; Daily transaction is more than 7.2 m now	Free trips for tourist cards. Discounted transfer fare	Discounted transfer fee. HK$0.5 discount for child & senior citizen	Self-service add value machine, transport service counter, some retail stores	Only for personalised cards. handling fee is required	Transport, banking, shopping, telephone, food and *etc*.

Remarks A: adult, C: child, S: student, SC: senior citizen, T: tourist, DI: disabled, P: personalised, PP: pre pay, D: daily, W: weekly, M: monthly, Y: yearly, N/A: not available

Appendix B: RP and SP Questionnaires

Institute for Transport Studies

University of Leeds

Questionnaire for Public Transport Fare Payment Choices

Dear Survey Participants,

This questionnaire is a part of a research project being conducted by a PhD student at Institute for Transport Studies, University of Leeds. The purpose of the questionnaire is to reveal public transport users' choices on different fare payment methods: Cash, Travel cards and Smart cards.

Your participation is critical to the success of the study. Would you please take just 10-15 minutes to complete the questionnaire and submit to the surveyors? Your co-operation would be greatly appreciated.

Any information provided will be dealt with in the strictest confidence and used for research purpose only, in accordance with the Data Protection Act 1998, UK.

Three questionnaire papers are for different respondents: Cash, Travel Cards and Smart Cards. Before you start filling in the questionnaire, please ensure that you have got the correct questionnaire paper proper for you.

Thank you!

Xu Hao
Research student
Institute for Transport Studies
University of Leeds
Leeds, U.K.
LS2 9JT
xhao@its.leeds.ac.uk

Smart Cards Fare Payment Users

Section A:

Please answer all questions in this section according to your smart cards fare payment behaviour at **LAST MONTH.** (*Please tick "√" one answer from the following options*)

1. What type of smart card did you use last month?
   ☐ A. "Pay as you go" card
   ☐ B. Electronic travel card (a minimum payment required per month)
   ☐ C. Student smart card
   ☐ D. Elder smart card
2. Please estimate your travel cost at **LAST MONTH** by using smart cards:_______ yuan.
3. How much quicker than cash on boarding time when you used smart cards?
   Roughly ( ) seconds
4. Compared with cash, do you think you make more trips because of smart cards fare payment?
   ☐ Yes. ☐ No.
5. Compared with Cash, please indicate seat availability when using smart cards
   ☐ No difference ☐ Slightly better ☐ Better ☐ Much better
6. How did you buy/ top-up your smart cards (please tick all that apply)?
   ☐ At ticket offices
   ☐ By banks
   ☐ By agencies
7. How easy to top up/purchase your smart cards?
   ☐ Very Difficult ☐ Difficult ☐ Neutral ☐ Easy ☐ Very Easy
8. Did you use other functions of smart cards in **LAST MONTH**? (please tick all that apply)
   ☐ Banking
   ☐ Parking/tolling fee payment
   ☐ Shopping
   ☐ None. Only used for public transport
9. Please give your overall assessment for smart cards fare payment method
   ☐ Totally unsatisfied
   ☐ Unsatisfied
   ☐ Neutral
   ☐ Satisfied
   ☐ Totally satisfied
10. Except that you mainly used smart cards in **LAST MONTH**, how much extra cash did you pay?
   ☐ No need to use cash any more
   ☐ Need to pay by cash: ( ) yuan

Please go to Section B. Thank you!

Section B:

The aim of this section is to identify what would be like if you had travel cards.

Would it be possible for you to use Travel Cards at **LAST MONTH**?

☐ Yes. ☐ No.

*If "**Yes**", please complete this section.*

*If "**No**", please go to **Section C**.*

1. What type of travel card could you use last month?

☐ A. Monthly cards with limited bus route

☐ B. Quarterly cards with limited bus route

☐ C. Monthly cards with unlimited bus route

☐ D. Quarterly cards with unlimited bus route

2. How much would you cost by using travel cards at **LAST MOTH**? ______ yuan

3. How much quicker than cash on boarding time when you would use travel cards?
Roughly (   ) seconds

4. Compared with cash, do you think you would make more trips because of travel cards?

☐ Yes. ☐ No.

5. Compared with Cash, please indicate seat availability when using travel cards fare payment

☐ No difference ☐ Slightly better ☐ Better ☐ Much better

6. How would you buy/renew you travel cards? (please tick all that apply)

☐ At ticket office

☐ At agencies

7. How easy to renew/purchase your travel cards?

☐ Very Difficult ☐ Difficult ☐ Neutral ☐ Easy ☐ Very Easy

8. Please give your overall assessment for travel cards fare payment method.

☐ Totally unsatisfied

☐ Unsatisfied

☐ Neutral

☐ Satisfied

☐ Totally satisfied

9. If you mainly would use travel cards in **LAST MONTH**, how much extra cash would you pay?

☐ No need to use cash any more

☐ Need to pay by cash: (   ) yuan

10. If you mainly would use travel cards in **LAST MONTH**, how much extra money on smart cards would you pay?

☐ No need to use smart cards any more

☐ Need to pay by smart cards: (   ) yuan

Please go to Section C. Thank you!

Section C:

If you would primarily use cash fare payment method last month, please imagine what could happen and answer the following questions.

1. What type of ticket could you buy primarily?
   ☐ Flat fare ☐ Zonal Fare

2. How much would you cost by using cash fare payment at **LAST MONTH** ______ yuan

3. Please give your overall assessment for cash fare payment method.
   ☐ Totally unsatisfied
   ☐ Unsatisfied
   ☐ Neutral
   ☐ Satisfied
   ☐ Totally satisfied

4. If you mainly would use cash fare payment method in **LAST MONTH**, how much extra money would you pay by smart cards?
   ☐ I would not consider using smart cards.
   ☐ Need to pay by smart cards: _______ yuan

Please go to Section D. Thank you!

Section D:

All respondents must complete this section

1. How old are you?
☐A. 16-25 ☐B. 26-35 ☐ C. 36-45 ☐D. 46-60 ☐E. Over 60

2. Your gender? ☐A. Male ☐B. Female

3. Your educational level?
☐A. High school or less
☐B. Undergraduate student
☐C. College graduate
☐D. Postgraduate or equivalent

4. What is employment status?
☐A. Employed full-time
☐B. Employed part-time
☐C. Unemployed
☐D. Student, working full or part time
☐E. Student, not working
☐F. Homemaker
☐G. Retired

5. Household income per month?
☐A. <￥1500
☐B.￥1500-￥2999
☐C. ￥3000-￥3999
☐D. ￥4000-￥5999
☐E. ≥￥6000

6. Do you have a personal vehicle (e.g., car, truck, motorcycle) available for transportation when you want it?
☐A. Always
☐B. Most of the time
☐C. Sometimes
☐D. Rarely
☐E. Never or no personal vehicle

7. Would you like to pre pay your public transport fare?
☐ Yes, I would pre pay weekly.
☐ Yes, I would pre pay monthly.
☐ Yes, I would pre pay quarterly.
☐ No, I would not.

The End

Thank you for your co-operation!

ITS
Institute for Transport Studies

University of Leeds

Public Transport Fare Payment Survey

Dear Passengers,

The purpose of the questionnaire is to identify public transport users' choices on different fare payment methods: Cash, Travel cards and Smart cards. Your participation is critical to the success of the study. Would you please take just 10-15 minutes to complete the questionnaire and submit to the surveyors? Your co-operation would be greatly appreciated.

Any information provided will be dealt with in the strictest confidence and used for research purpose only, in accordance with the Data Protection Act 1998, UK.

Thank you!

Xu Hao
Research student
Institute for Transport Studies
University of Leeds
Leeds, U.K.
LS2 9JT
xhao@its.leeds.ac.uk

Before you start the game, please indicate your fare payment type by using tick "√".

☐ **If you are a cash fare payment user primarily, please answer the questionnaires related to cash: Cash vs. Travel Cards; Cash vs. Smart Cards (Pay as you go cards).**

☐ **If you are a travel card user primarily, please answer the questionnaires related to travel cards: Cash vs. Travel Cards; Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

☐ **If you are a smart card user primarily, please answer the questionnaires related to smart cards: Cash vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

Please ensure that you have got the correct questionnaires before you start the game.

Cash vs. Travel Cards

(*For cash or travel card users*)

1. **Did you pay your fare by using cash?**
   ☐Yes. ☐No (if answer NO, please go to Q4)
2. **What kind of ticket did you buy primarily?**
   ☐ Flat fare ☐ Zonal Fare
3. **How many trips did you make by using cash in one month? About (   )**
4. **Did you pay your fare by using travel cards in last month?**
   ☐Yes. ☐No. (if answer NO, please skip Q5 and Q6)
5. **What type of travel cards did you use in last month?**

☐ Monthly cards with limited bus route
☐ Quarterly cards with limited bus route
☐ Monthly cards with unlimited bus route
☐ Quarterly cards with unlimited bus route

6. How many trips did you make by using travel cards in one month? About (　　　)

Suppose that you could only choose fare payment methods from Cash and Travel Cards for your ONE MONTH trips according to those factors as described below. Then which one would you prefer?

Cash: The most traditional fare payment which can be used in any public transport services. Normally cash fare is collected by bus drivers, fare boxes or conductors.

Travel Cards: One kind of cashless fare payment with limited or unlimited services depending on what kind of card type you buy. Normally card users have to renew their cards monthly.

An example is shown below.

EXAMPLE

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1yuan per ride	Average 20seconds slower than travel cards	Any	Yes	
Travel Cards	29yuan monthly	Straight getting on	Limited route: Only one bus or light rail route service. But you still can pay by cash to take other services	N/A	√

Now please make your 8 choices as though you were paying your public transport fare.

SITUATION 1

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	0.8yuan per ride	Average 20seconds slower than travel cards	Any	No	
Travel Cards	29yuan monthly	Straight getting on	Limited route: Only one bus or light rail route service. But you still can pay by cash to take other services	N/A	

SITUATION 2

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1yuan per ride	Average 60seconds slower than travel cards	Any	No	
Travel Cards	53yuan monthly	Straight getting on	Limited route: Only one bus or light rail route service. But you still can pay by cash to take other services	N/A	

SITUATION 3

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1.4yuan per ride	Average 40seconds slower than travel cards	Any	Yes	
Travel Cards	45yuan monthly	Straight getting on	Limited route: Only one bus or light rail route service. But you still can pay by cash to take other services	N/A	

SITUATION 4

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1yuan per ride	Average 20seconds slower than travel cards	Any	Yes	
Travel Cards	37yuan monthly	Straight getting on	Unlimited routes and public transport modes with extra charge: 15% more based on this value	N/A	

SITUATION 5

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1.4yuan per ride	Average 80seconds slower than travel cards	Any	No	
Travel Cards	29yuan monthly	Straight getting on	Unlimited routes and public transport modes with extra charge: 15% more based on this value	N/A	

SITUATION 6

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	0.8 yuan per ride	Average 80seconds slower than travel cards	Any	Yes	
Travel Cards	53yuan monthly	Straight getting on	Unlimited routes and public transport modes without any extra charges	N/A	

SITUATION 7

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	1.2 yuan per ride	Average 20seconds slower than travel cards	Any	No	
Travel Cards	45yuan monthly	Straight getting on	Unlimited routes and public transport modes without any extra charges	N/A	

SITUATION 8

Fare Payment Methods	Travel Cost	Boarding Time	Public Transport Services Covered	Whether passengers can get change	Choice
Cash	0.8 yuan per ride	Average 40seconds slower than travel cards	Any	No	
Travel Cards	37yuan monthly	Straight getting on	Unlimited routes and public transport modes with extra charge: 10% more based on this value	N/A	

Background Survey

1. How old are you?

☐ A. 16-25 ☐ B. 26-35 ☐ C. 36-45 ☐ D. 45-60 ☐ E. Over 60

2. Your gender? ☐ A. Male ☐ B. Female

3. Your educational level?

☐ A. High school or less
☐ B. Undergraduate student
☐ C. College graduate
☐ D. Postgraduate or equivalent

4. What is employment status?

☐ A. Employed full-time
☐ B. Employed part-time
☐ C. Unemployed
☐ D. Student, working full or part time
☐ E. Student, not working
☐ F. Homemaker
☐ G. Retired

5. Household income per month?

☐ A.<￥1500
☐ B. ￥1500-￥2999
☐ C. ￥3000-￥3999
☐ D. ￥4000-￥5999
☐ E. >￥6000

6. Do you have a personal vehicle (e.g., car, truck, motorcycle) available for transportation when you want it?

☐ A. Always
☐ B. Most of the time
☐ C. Sometimes
☐ D. Rarely
☐ E. Never or no personal vehicle

7. Would you like to pre pay your public transport fare?

☐ Yes, I would pre pay weekly.
☐ Yes, I would pre pay monthly.
☐ Yes, I would pre pay quarterly.
☐ No, I would not.

The End

Thank you for your co-operation!

ITS
Institute for Transport Studies

University of Leeds

Public Transport Fare Payment Survey

Dear Passengers,

The purpose of the questionnaire is to identify public transport users' choices on different fare payment methods: Cash, Travel cards and Smart cards. Your participation is critical to the success of the study. Would you please take just 10-15 minutes to complete the questionnaire and submit to the surveyors? Your co-operation would be greatly appreciated.

Any information provided will be dealt with in the strictest confidence and used for research purpose only, in accordance with the Data Protection Act 1998, UK.

Thank you!

Xu Hao
Research student
Institute for Transport Studies
University of Leeds
Leeds, U.K.
LS2 9JT
xhao@its.leeds.ac.uk

Before you start the game, please indicate your fare payment type by using tick "√".

☐ **If you are a cash fare payment user primarily, please answer the questionnaires related to cash: Cash vs. Travel Cards; Cash vs. Smart Cards (Pay as you go cards).**

☐ **If you are a travel card user primarily, please answer the questionnaires related to travel cards: Cash vs. Travel Cards; Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

☐ **If you are a smart card user primarily, please answer the questionnaires related to smart cards: Cash vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

Please ensure that you have got the correct questionnaires before you start the game.

Cash vs. Smart Cards

(*For cash or smart card users*)

1. **Did you pay your fare by using cash?**
   ☐Yes. ☐No (if answer NO, please go to Q4)
2. **What kind of ticket did you buy primarily?**
   ☐ Flat fare ☐ Zonal Fare
3. **How many trips did you make by using cash in one month? About (   )**
4. **Did you pay your fare by using smart cards before?**
   Yes. No. (if answer NO, please skip Q5 and Q6)
5. **What type of smart cards did you use?**

☐ "Pay as you go" card
☐ Electronic travel card (a minimum payment required per month)
☐ Student smart card
☐ Elder smart card

6. **How many trips did you make by using smart cards in one month? About (   )**

Suppose that you could only choose fare payment methods from Cash and Smart Cards for your ONE MONTH trips according to those factors as described below. Then which one would you prefer?

Cash: The most traditional fare payment which can be used in any public transport services. Normally cash fare is collected by bus drivers, fare boxes or conductors.

Smart Cards (pay as you go cards): One kind of cashless fare payment combined with some new features, such as multifunction, overdraft, much more top-up means than travel cards and various ticket packages and so forth. Smart cards can provide discounted single fare. Deposit is required.

An example is shown below.

EXAMPLE

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1yuan per ride	Average 20seconds slower than smart cards	No	N/A	None	
Smart Cards	0.8 yuan per ride	Straight getting on	N/A	0yuan	None. Only for public transport	√

Now please make your 8 choices as though you were paying your public transport fare.

SITUATION 1

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1yuan per ride	Average 20seconds slower than smart cards	No	N/A	None	
Smart Cards	1yuan per ride	Straight getting on	N/A	0yuan	None. Only for public transport	

SITUATION 2

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1.2 yuan per ride	Average 80seconds slower than smart cards	No	N/A	None	
Smart Cards	1yuan per ride	Straight getting on	N/A	30yuan	None. Only for public transport	

SITUATION 3

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1yuan per ride	Average 60seconds slower than smart cards	Yes	N/A	None	
Smart Cards	0.4 yuan per ride	Straight getting on	N/A	20yuan	None. Only for public transport	

SITUATION 4

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1yuan per ride	Average 40seconds slower than smart cards	Yes	N/A	None	
Smart Cards	0.8 yuan per ride	Straight getting on	N/A	0yuan	Shopping, telephone, amusement ,parking, tolling and banking	

SITUATION 5

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1.4 yuan per ride	Average 20seconds slower than smart cards	No	N/A	None	
Smart Cards	0.8 yuan per ride	Straight getting on	N/A	50yuan	Shopping, telephone, entertainment ,parking, tolling and banking	

SITUATION 6

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	0.8 yuan per ride	Average 80seconds slower than smart cards	Yes	N/A	None	
Smart Cards	0.8 yuan per ride	Straight getting on	N/A	50yuan	Shopping, telephone, amusement	

SITUATION 7

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1 yuan per ride	Average 60seconds slower than smart cards	No	N/A	None	
Smart Cards	0.6 yuan per ride	Straight getting on	N/A	0yuan	Shopping, telephone, amusement	

SITUATION 8

Fare Payment Methods	Travel Cost	Boarding Time	Whether passengers can get change	Deposit	Multifunction	Choice
Cash	1yuan per ride	Average 40seconds slower than smart cards	No	N/A	None	
Smart Cards	1yuan per ride	Straight getting on	N/A	20yuan	Shopping, telephone, amusement ,pa rking and tolling	

Background Survey

1. How old are you?

☐ A. 16-25 ☐ B. 26-35 ☐ C. 36-45 ☐ D. 45-60 ☐ E. Over 60

2. Your gender? ☐ A. Male ☐ B. Female

3. Your educational level?

☐ A. High school or less
☐ B. Undergraduate student
☐ C. College graduate
☐ D. Postgraduate or equivalent

4. What is employment status?

☐ A. Employed full-time
☐ B. Employed part-time
☐ C. Unemployed
☐ D. Student, working full or part time
☐ E. Student, not working
☐ F. Homemaker
☐ G. Retired

5. Household income per month?

☐ A.<￥1500
☐ B. ￥1500-￥2999
☐ C. ￥3000-￥3999
☐ D. ￥4000-￥5999
☐ E. >￥6000

6. Do you have a personal vehicle (e.g., car, truck, motorcycle) available for transportation when you want it?

☐ A. Always
☐ B. Most of the time
☐ C. Sometimes
☐ D. Rarely
☐ E. Never or no personal vehicle

7. Would you like to pre pay your public transport fare?

☐ Yes, I would pre pay weekly.
☐ Yes, I would pre pay monthly.
☐ Yes, I would pre pay quarterly.
☐ No, I would not.

The End

Thank you for your co-operation!

ITS
Institute for Transport Studies

University of Leeds

Public Transport Fare Payment Survey

Dear Passengers,

The purpose of the questionnaire is to identify public transport users' choices on different fare payment methods: Cash, Travel cards and Smart cards. Your participation is critical to the success of the study. Would you please take just 10-15 minutes to complete the questionnaire and submit to the surveyors? Your co-operation would be greatly appreciated.

Any information provided will be dealt with in the strictest confidence and used for research purpose only, in accordance with the Data Protection Act 1998, UK.

Thank you!

Xu Hao
Research student
Institute for Transport Studies
University of Leeds
Leeds, U.K.
LS2 9JT
xhao@its.leeds.ac.uk

Before you start the game, please indicate your fare payment type by using tick "√".

☐ **If you are a cash fare payment user primarily, please answer the questionnaires related to cash: Cash vs. Travel Cards; Cash vs. Smart Cards (Pay as you go cards).**

☐ **If you are a travel card user primarily, please answer the questionnaires related to travel cards: Cash vs. Travel Cards; Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

☐ **If you are a smart card user primarily, please answer the questionnaires related to smart cards: Cash vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay as you go cards); Travel Cards vs. Smart Cards (Pay monthly cards).**

Please ensure that you have got the correct questionnaires before you start the game.

Travel Cards vs. Smart Cards

(For travel card or smart card users)

1. **Did you pay your fare by using travel cards before?**
   Yes. No. (if answer NO, please go to Q4)
2. **What type of travel cards did you use?**
   ☐ Monthly cards with limited bus route
   ☐ Quarterly cards with limited bus route
   ☐ Monthly cards with unlimited bus route
   ☐ Quarterly cards with unlimited bus route
3. **How many trips did you make by using travel cards in one month? About ( )**

4. **Did you pay your fare by using smart cards before?**
   Yes. No. (if answer NO, please skip Q5 and Q6)

5. **What type of smart cards did you use?**
   ☐ "Pay as you go" card
   ☐ Electronic travel card (a minimum payment required per month)
   ☐ Student smart card
   ☐ Elder smart card

6. **How many trips did you make by using smart cards in one month? About (   )**

Suppose that you could only choose fare payment methods from Travel Cards and Smart Cards for your ONE MONTH trips according to those factors as described below. Then which one would you prefer?

Travel Cards: One kind of cashless fare payment with limited or unlimited services depending on what kind of card type you buy. Normally card users have to renew their cards monthly.

Smart Cards (pay as you go cards): One kind of cashless fare payment combined with some new features, such as multifunction, overdraft, much more top-up means than travel cards and various ticket packages and so forth. Smart cards can provide discounted single fare. Deposit is required.

An example is shown below.

EXAMPLE

Fare Payment Methods	Travel Cost	Public transport services Covered	Top up methods	Choice
Travel Cards	29yuan per month	Limited route: Only one bus or light rail route service. If you want to take other routes, you have to pay by cash	Only at ticket offices	
Smart Cards	0.4yuan per ride	Any public transport modes and routes	At ticket offices and agencies	√

Now please make your 8 choices as though you were paying your public transport fare.

SITUATION 1

Fare Payment Methods	Travel Cost	Public transport services Covered	Top up methods	Choice
Travel Cards	29yuan per month	Limited route: Only one bus or light rail route service. If you want to take other routes, you have to pay by cash	Only at ticket offices	
Smart Cards	0.4yuan per ride	Any public transport modes and routes	At ticket offices and agencies	

SITUATION 2

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	37yuan per month	Unlimited routes with extra charge: 10% more than this value	Only at ticket offices	
Smart Cards	1yuan per ride	Any public transport modes and routes	At ticket offices, banks and agencies	

SITUATION 3

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	53yuan per month	Unlimited public transport modes and routes in urban area without extra charge	Only at ticket offices	
Smart Cards	0.8yuan per ride	Any public transport modes and routes	At ticket offices, agencies, banks, telephone and online	

SITUATION 4

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	37yuan per month	Limited route: Only one bus or light rail route service. If you want to take other routes, you have to pay by cash	At ticket offices and agencies	
Smart Cards	0.6yuan per ride	Any public transport modes and routes	At ticket offices, agencies, banks, telephone and online	

SITUATION 5

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	53yuan per month	Unlimited routes with extra charge: 15% more than this value	At ticket offices and agencies	
Smart Cards	0.4yuan per ride	Any public transport modes and routes	At ticket offices, agencies and banks	

SITUATION 6

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	29yuan per month	Unlimited routes with extra charge: 15% more than this value	At ticket offices and agencies	
Smart Cards	1yuan per ride	Any public transport modes and routes	At ticket offices, agencies, banks, telephone and online	

SITUATION 7

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	45yuan per month	Limited route: Only one bus or light rail route service. If you want to take other routes, you have to pay by cash	At ticket offices and agencies	
Smart Cards	0.8yuan per ride	Any public transport modes and routes	At ticket offices, agencies and banks	

SITUATION 8

Fare Payment Methods	Travel Cost	Public transport services covered	Top up methods	Choice
Travel Cards	29yuan per month	Unlimited modes and routes (any services in urban area) without any extra charge	At ticket offices, agencies and banks	
Smart Cards	0.6yuan per ride	Any public transport modes and routes	At ticket offices, agencies and banks	

Background Survey

1. How old are you?
☐ A. 16-25 ☐ B. 26-35 ☐ C. 36-45 ☐ D. 45-60 ☐ E. Over 60

2. Your gender? ☐ A. Male ☐ B. Female

3. Your educational level?
☐ A. High school or less
☐ B. Undergraduate student
☐ C. College graduate
☐ D. Postgraduate or equivalent

4. What is employment status?
☐ A. Employed full-time

☐ B. Employed part-time
☐ C. Unemployed
☐ D. Student, working full or part time
☐ E. Student, not working
☐ F. Homemaker
☐ G. Retired

5. Household income per month?

☐ A.<￥1500
☐ B. ￥1500-￥2999
☐ C. ￥3000-￥3999
☐ D. ￥4000-￥5999
☐ E. >￥6000

6. Do you have a personal vehicle (e.g., car, truck, motorcycle) available for transportation when you want it?

☐ A. Always
☐ B. Most of the time
☐ C. Sometimes
☐ D. Rarely
☐ E. Never or no personal vehicle

7. Would you like to pre pay your public transport fare?

☐ Yes, I would pre pay weekly.
☐ Yes, I would pre pay monthly.
☐ Yes, I would pre pay quarterly.
☐ No, I would not.

The End

Thank you for your co-operation!

Appendix C: Estimation Results in Segmentation Analysis

Table C-1: Estimation Results by Age Segments

Variables	Age 16-25	Age 26-35	Age 36-45	Over 45
1. Travel cost (Yuan)	-0.1162(-11.2)	-0.1133(-13.1)	-0.0983(-12.9)	-0.0967(-11.1)
2. Boarding time (second)	-0.00767(-3.2)	-0.00814(-3.3)	-0.00752(-3.8)	-0.00761(-4.1)
3. Overall assessment-RP:				
Dummy 1: Totally unsatisfied & Unsatisfied	-0.4461(-2.1)	-0.4581(-2.0)	-0.4621(-2.0)	-0.4812(-2.2)
Dummy 2: Satisfied & Totally satisfied	0.2636(2.0)	0.2891(1.9)	0.2891(1.8)	0.3387(2.0)
Base: Neutral				
5. Cash ticket type-RP:				
Dummy variable: Zonal fare; **Base:** Flat fare	0.4409(2.0)	0.4461(2.2)	0.4512(2.1)	0.4522(2.0)
18. Seat Availability by using travel cards or smart cards, comparing with cash-RP				
Dummy variable: Slightly better or Better or Much better	0.6493(3.9)	0.6502(4.0)	0.6920(3.9)	0.7022(3.8)
Base: No difference				
19. Top-up/purchase methods of travel cards-RP				
Dummy variable: ticket offices **or** banks	-0.4288(-2.2)	-0.4298(-2.1)	-0.4367(-2.0)	-0.4419(-2.0)
Base: Both ticket offices and banks				
20. Top-up/purchase methods of smart cards-RP				
Dummy 1: At ticket offices	-1.534(-5.6)	-1.609(-5.8)	-0.1745(-5.5)	-0.1892(-5.0)
Dummy 2: Banks	-1.195(-5.2)	-1.271(-5.1)	-0.1482(-4.8)	-0.1522(-4.9)
Dummy 3: Agencies	-0.6823(-2.8)	-0.769(-2.6)	-0.8997(-2.6)	-0.9225(-2.4)
Base: two or three top up methods used				
21. Difficulty of topping up/purchasing-RP				
Dummy 1: Very difficult & Difficult	-0.6827(-2.7)	-0.7895(-2.8)	-0.8949(-2.3)	-0.9032(-2.4)
Dummy 2: Easy & Very easy	0.5158(2.5)	0.5387(2.5)	0.6291(2.2)	0.6387(2.3)
Base: Neutral				
22. Whether passengers can get changes if they pay bid money value: 0: No; **1:** Yes --- **Cash (SP)**	0.1348(1.9)	0.1496(2.0)	0.1455(1.9)	0.1461(1.9)
23. Deposit –Smart cards(Yuan)**-SP**	-0.01179(-7.2)	-0.01202(-7.0)	-0.01198(-6.9)	-0.01208(-6.6)
24. Service-Travel cards SP:				
1: Limited routes: Only one bus, or light rail route service in urban area (servicetc1);	-0.7158(-6.6)	-0.7115(-6.8)	-0.6280(-6.0)	-0.6213(-6.1)
2: Unlimited routes with extra charge: 10% more than limited services (Servicetc2);	-0.7878(-6.6)	-0.7693(-6.4)	-0.6860(-5.8)	-0.6794(-5.9)
3: Unlimited routes with extra charge: 15% more than limited services (Servicetc3)	-0.8947(-7.4)	-0.8871(-7.4)	-0.7794(-6.9)	-0.7732(-6.9)
Base: Unlimited routes without any extra charges				

25. Service-Smart cards-SP: **1:** Unlimited routes with extra charge: 10% more than limited services (servicesc1); 2: Unlimited routes with extra charge: 15% more than limited services (servicesc2); 3: Unlimited routes with extra charge: 20% more than limited services (servicesc3) **Base:** Unlimited routes (any services in urban area) without any extra charges	-0.1545(-1.0) -0.2742(-1.9) -0.5427(-3.6)	-0.1643(-1.1) -0.2821(-2.0) -0.5393(-3.3)	-0.1504(-0.9) -0.2624(-1.9) -0.4786(-3.0)	-0.1529(-1.0) -0.2642(-1.9) -0.4781(-3.1)
26. Overdraft-smart cards-SP: Overdraft function in smart cards: 0: No; 1 Yes	0.452 (4.6)	0.4793(4.5)	0.4216 (4.1)	0.4268 (4.2)
27. Multifunction-Smart cards-SP: **1:** Shopping, telephone, entertainment (mf1); 2: Shopping, telephone, entertainment, parking and tolling (mf2); 3: Shopping, telephone, entertainment, parking, tolling and banking (mf3) **Base:** No. only for public transport	0.2579(2.3) 0.7111 (6.0) 1.026 (7.5)	0.2674 (2.2) 0.7285 (6.0) 1.012 (7.1)	0.2368 (2.3) 0.6437 (5.7) 0.8865 (7.3)	0.2216 (2.1) 0.6078 (5.5) 0.8449 (6.9)
28. Geographic Area-Smart cards-SP: **1:** Urban and rural areas (Geo1); 2: Dalian and other nearby cities (Geo2); 3: Within one province (Geo3) **Base:** Only urban area	0.3858 (3.4) 0.6809 (5.0) 0.9215 (6.3)	0.4022 (3.4) 0.6968 (4.9) 0.9313 (6.0)	0.341 (3.3) 0.6860 (4.6) 0.8029 (5.9)	0.3397 (3.1) 0.5758 (4.7) 0.7843 (5.5)
29. Top-up-travel cards-SP: **1:** Ticket offices and agencies (topuptc1); 2: Ticket offices and agencies and banks (topuptc2) **Base:** Ticket offices	0.09817 (0.4) 0.7983 (3.2)	0.09512 (0.5) 0.8169 (3.5)	0.08922 (0.4) 0.7145 (3.1)	0.08694 (0.4) 0.7094 (3.1)
30. Top-up-smart cards-SP: **1:** Ticket offices, banks, agencies (topup1); 2: Ticket offices, banks, agencies, self-adding value machine (topup2); 3:Ticket offices, banks, agencies, self-adding value machine, telephone and Interne (topup3) **Base:** Only at ticket offices	0.3695 (2.6) 0.7472 (4.4) 1.4548 (5.6)	0.3807 (2.7) 0.7512 (4.6) 1.5058 (5.7)	0.3331 (2.4) 0.6575 (4.4) 1.312 (5.2)	0.24 (2.2) 0.6513 (4.3) 1.2272 (5.1)
ASC-travel cards (RP):	0.2269 (0.9)	0.2417 (0.8)	0.2689 (1.0)	0.2298 (0.9)
ASC-smart cards (RP):	0.7182 (2.6)	0.6793 (2.3)	0.7011(2.4)	0.6766 (2.2)
ASC-travel cards (SP1):	0.3651 (2.2)	0.3393 (2.3)	0.3932 (2.4)	0.3765 (2.4)
ASC-smart cards (SP2):	-0.5017 (-4.6)	-0.2998 (-4.4)	-0.3895 (-4.9)	-04665 (-4.8)
ASC-smart cards (SP3):	-0.6256 (-5.5)	-0.5901 (-5.4)	-0.5731 (-5.5)	-0.3978 (-5.1)
ASC-smart cards (SP4):	-0.1185 (-1.2)	-0.0978 (-1.0)	-0.1025(-1.0)	-0.9971 (-1.1)
θ1 (SP1, 2 and 4):	1.413 (10.1)	1.389 (11.2)	1.401(12.0)	1.426 (12.2)
θ2 (SP3):	1.803 (9.5)	1.722 (9.9)	1.749 (9.7)	1.832 (9.5)
Likelihood:	-690.238	-729.021	-787.845	-986.876
No. of Observations:	1050	1467	1691	2167
Rho-squared value w.r.t. cons:	0.4453	0.4372	0.4301	0.4097

Table C-2: Estimation Results by Sex Segments

Variables	Male	Female
1. Travel cost (Yuan)	-0.0981 (-13.3)	-0.1125 (-12.8)
2. Boarding time (second)	-0.007619 (-4.4)	-0.006452 (-4.5)
3. Overall assessment-RP:		
Dummy 1: Totally unsatisfied & Unsatisfied	-0.4221 (-2.2)	-0.4943 (-2.3)
Dummy 2: Satisfied & Totally satisfied	0.3089 (2.0)	0.3563 (2.1)
Base: Neutral		
4. Cash ticket type-RP:		
Dummy variable: Zonal fare; **Base:** Flat fare	0.4390 (1.9)	0.4822 (2.0)
5 . Seat Availability by using travel cards or smart cards, comparing with cash-RP		
Dummy variable: Slightly better or Better or Much better	0.6713 (3.8)	0.7109 (3.8)
Base: No difference		
6. Top-up/purchase methods of travel cards-RP		
Dummy variable: ticket offices **or** banks	-0.4209 (-2.0)	-0.4671 (-2.0)
Base: Both ticket offices and banks		
7. Top-up/purchase methods of smart cards-RP		
Dummy 1: At ticket offices	-1.514 (-5.5)	-1.798 (-5.4)
Dummy 2: Banks	-1.169 (-4.9)	-1.522 (-4.7)
Dummy 3: Agencies	-0.8701 (-2.6)	-0.8969 (-2.5)
Base: two or three top up methods used		
8. Difficulty of topping up/purchasing-RP		
Dummy 1: Very difficult & Difficult	-0.8392 (-2.6)	-0.8967 (-2.5)
Dummy 2: Easy & Very easy	0.5819 (2.2)	0.6238 (2.5)
Base: Neutral		
9. Whether passengers can get changes if they pay bid money value: 0: No; **1:** Yes --- **Cash (SP)**	0.1520 (2.0)	0.1262 (1.9)
10. Deposit –Smart cards(Yuan)**-SP**	-0.01379 (-6.9)	-0.01153 (-6.6)
11. Service-Travel cards SP:		
1: Limited routes: Only one bus, or light rail route service in urban area (servicetc1);	-0.6396 (-6.5)	-0.6874 (-6.4)
2: Unlimited routes with extra charge: 10% more than limited services (Servicetc2);	-0.6995 (-6.1)	-0.7627 (-6.2)
3: Unlimited routes with extra charge: 15% more than limited services (Servicetc3)	-0.7838 (-7.1)	-0.8663 (-7.2)
Base: Unlimited routes without any extra charges		

12. Service-Smart cards-SP:		
1: Unlimited routes with extra charge: 10% more than limited services (servicesc1);	-0.1599 (-1.0)	-0.1474 (-0.9)
2: Unlimited routes with extra charge: 15% more than limited services (servicesc2);	-0.2815 (-2.0)	-0.2644 (-1.9)
3: Unlimited routes with extra charge: 20% more than limited services (servicesc3)	-0.4885 (-3.3)	-0.5141 (-3.4)
Base: Unlimited routes (any services in urban area) without any extra charges		
13. Overdraft-smart cards-SP: Overdraft function in smart cards: 0: No; 1 Yes	0.4836 (4.4)	0.4253 (4.5)
14. Multifunction-Smart cards-SP:		
1: Shopping, telephone, entertainment (mf1);	0.2462 (2.2)	0.2475 (2.3)
2: Shopping, telephone, entertainment, parking and tolling (mf2);	0.6524 (5.6)	0.6874 (5.7)
3: Shopping, telephone, entertainment, parking, tolling and banking (mf3)	0.9045 (7.1)	0.9821 (7.2)
Base: No. only for public transport		
15. Geographic Area-Smart cards-SP:		
1: Urban and rural areas (Geo1);	0.3512 (3.4)	0.3611 (3.4)
2: Dalian and other nearby cities (Geo2);	0.6082 (4.4)	0.6536 (4.8)
3: Within one province (Geo3)	0.8054 (6.1)	0.8809 (6.2)
Base: Only urban area		
16. Top-up-travel cards-SP:		
1: Ticket offices and agencies (topuptc1);	0.09157 (0.5)	0.09786 (0.5)
2: Ticket offices and agencies and banks (topuptc2)	0.7191 (3.3)	0.7673 (3.5)
Base: Ticket offices		
17. Top-up-smart cards-SP:		
1: Ticket offices, banks, agencies (topup1);	0.3414 (2.3)	0.3465 (2.4)
2: Ticket offices, banks, agencies, self-adding value machine (topup2);	0.6661 (4.1)	0.7211 (4.2)
3:Ticket offices, banks, agencies, self-adding value machine, telephone and Interne (topup3)	1.3194 (5.5)	1.4197 (5.5)
Base: Only at ticket offices		
ASC-travel cards (RP):	0.2090 (1.0)	0.2278 (1.0)
ASC-smart cards (RP):	0.6122 (2.2)	0.7538 (2.1)
ASC-travel cards (SP1):	0.3278 (2.0)	0.4091 (2.2)
ASC-smart cards (SP2):	-0.5611 (-4.3)	-0.7022 (-5.0)
ASC-smart cards (SP3):	-0.7369 (-4.9)	-0.6739 (-5.4)
ASC-smart cards (SP4):	-0.2065 (-1.0)	-0.2135 (-1.2)
θ1 (SP1, 2 and 4):	1.4122 (11.4)	1.443 (11.6)
θ2 (SP3):	1.807 (9.5)	1.821 (10.0)
Likelihood:	-2801.3166	-2791.02
No. of Observations:	3068	3307
Rho-squared value w.r.t. cons:	0.4322	0.4216

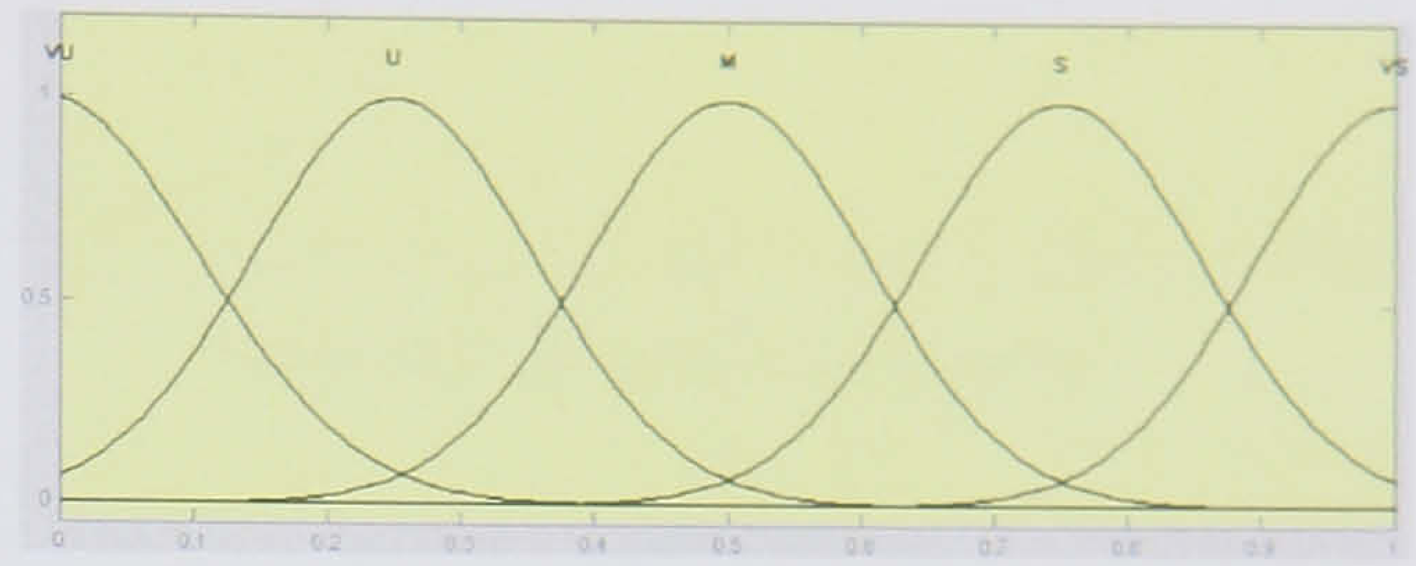
Table C-3: Estimation Results by Household Income Segments

Variables	Less than ¥1500	¥1500-¥2999	¥3000-¥3999	>¥4000
1. Travel cost (Yuan)	-0.1032(-10.1)	-0.1076(-11.1)	-0.0987(-11.9)	-0.0969(-11.8)
2. Boarding time (second)	-0.00650(-3.9)	-0.0111(-4.0)	-0.0106(-4.1)	-0.00958(-4.0)
3. Overall assessment-RP:				
Dummy 1: Totally unsatisfied & Unsatisfied	-0.4682 (-2.2)	-0.4390 (-2.1)	-0.4522 (-2.0)	-0.4589 (-2.1)
Dummy 2: Satisfied & Totally satisfied	0.3269 (1.9)	0.3311(2.0)	0.3487 (1.9)	0.3673 (1.9)
Base: Neutral				
4. Cash ticket type-RP: **Dummy variable:** Zonal fare; **Base:** Flat fare	0.4489 (2.0)	0.4315 (1.9)	0.4692 (1.9)	0.4714 (1.9)
5. Seat Availability by using travel cards or smart cards, comparing with cash-RP **Dummy variable:** Slightly better or Better or Much better **Base:** No difference	0.6893 (3.3)	0.6739 (3.2)	0.7012 (3.3)	0.7131 (3.3)
6. Top-up/purchase methods of travel cards-RP **Dummy variable:** ticket offices **or** banks **Base:** Both ticket offices and banks	-0.4439 (-1.9)	-0.4396 (-1.9)	-0.4565 (-2.0)	-0.4637 (-2.0)
7. Top-up/purchase methods of smart cards-RP				
Dummy 1: At ticket offices	-1.712 (-5.1)	-1.6955 (-5.1)	-1.8134 (-5.3)	-1.8345 (-5.2)
Dummy 2: Banks	-1.339 (-4.5)	-1.3491 (-4.7)	-1.4518 (-4.8)	-1.4601 (-4.4)
Dummy 3: Agencies	-0.8801 (-2.2)	-0.8923 (-2.3)	-0.9036 (2.3)	-0.9233 (-2.1)
Base: two or three top up methods used				
8. Difficulty of topping up/purchasing-RP				
Dummy 1: Very difficult & Difficult	-0.8792 (-2.2)	-0.8801 (-2.4)	-0.8997 (-2.5)	-0.8987 (-2.2)
Dummy 2: Easy & Very easy	0.6034 (2.3)	0.5988 (2.3)	0.6126 (2.1)	0.6233 (2.2)
Base: Neutral				
9. Whether passengers can get changes if they pay bid money value: 0: No; **1:** Yes --- **Cash (SP)**	0.1187 (1.9)	0.1463 (1.9)	0.1442 (1.8)	0.1474 (1.8)
10. Deposit –Smart cards(Yuan)**-SP**	-0.0126 (-6.5)	-0.0127(-6.4)	-0.01086(-6.1)	-0.09897(-6.0)
11. Service-Travel cards SP:				
1: Limited routes: Only one bus, or light rail route service in urban area (servicetc1);	-0.6347 (-6.2)	-0.6811 (-6.0)	-0.6322 (-6.1)	-0.6303 (-6.2)
2: Unlimited routes with extra charge: 10% more than limited services (Servicetc2);	-0.71 (-6.0)	-0.7188 (-5.9)	-0.6954 (-5.6)	-0.7 (-5.6)
3: Unlimited routes with extra charge: 15% more than limited services (Servicetc3)	-0.7864 (-6.6)	-0.8425 (-6.2)	-0.7863 (-6.2)	-0.7951 (-6.1)
Base: Unlimited routes without any extra charges				

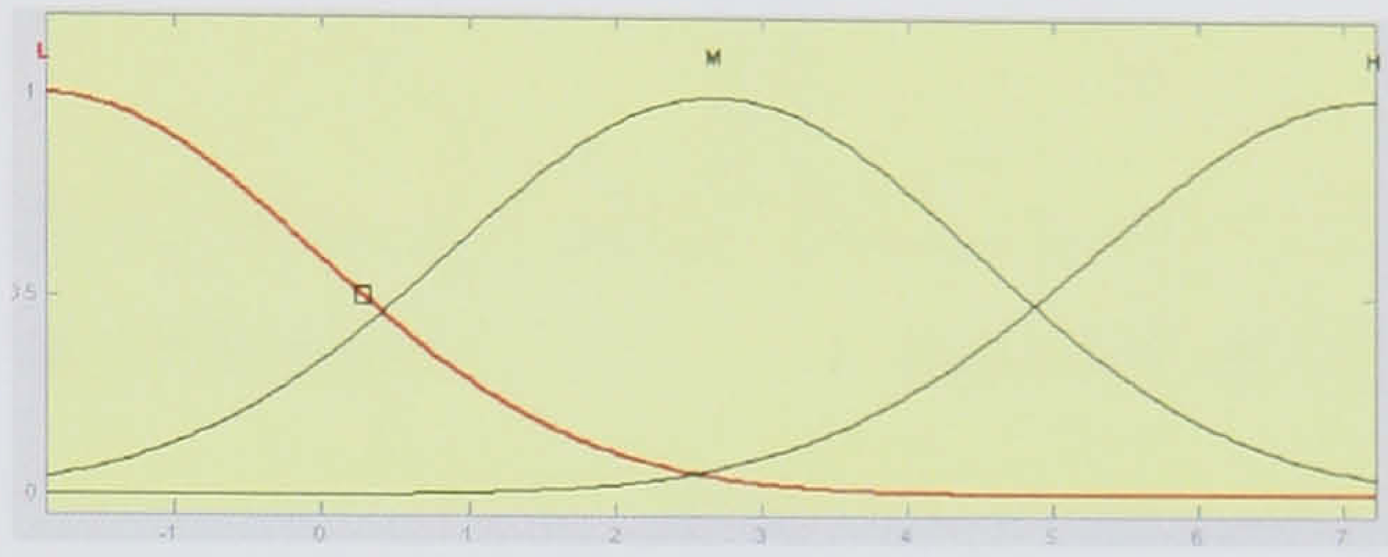
12. Service-Smart cards-SP:				
1: Unlimited routes with extra charge: 10% more than limited services (servicesc1);	-0.1331 (-0.9)	-0.1528 (-1.0)	-0.1571 (-0.9)	-0.1648 (-0.9)
2: Unlimited routes with extra charge: 15% more than limited services (servicesc2);	-0.2436 (-2.0)	-0.2668 (-2.0)	-0.2736 (-1.9)	-0.2802 (-2.0)
3: Unlimited routes with extra charge: 20% more than limited services (servicesc3)	-0.4716 (-3.2)	-0.5143 (-3.1)	-0.4890 (-3.0)	-0.4955 (-3.0)
Base: Unlimited routes (any services in urban area) without any extra charges				
13. Overdraft-smart cards-SP: Overdraft function in smart cards: 0: No; 1 Yes	0.3922 (4.3)	0.4584 (4.2)	0.4336 (3.9)	0.4383 (4.0)
14. Multifunction-Smart cards-SP:				
1: Shopping, telephone, entertainment (mf1);	0.2249 (2.3)	0.2572 (2.2)	0.242 (2.1)	0.2482 (2.2)
2: Shopping, telephone, entertainment, parking and tolling (mf2);	0.6305 (4.9)	0.6951 (5.1)	0.648 (5.0)	0.6468 (4.9)
3: Shopping, telephone, entertainment, parking, tolling and banking (mf3)	0.8906 (7.0)	0.6973 (6.8)	0.9038 (6.6)	0.8979 (6.6)
Base: No. only for public transport				
15. Geographic Area-Smart cards-SP:				
1: Urban and rural areas (Geo1);	0.3313 (3.3)	0.3744 (3.2)	0.3536 (3.4)	0.3588 (3.1)
2: Dalian and other nearby cities (Geo2);	0.3841 (4.4)	0.6499 (4.6)	0.6105 (4.4)	0.6041 (4.6)
3: Within one province (Geo3)	0.7977 (6.1)	0.8737 (6.0)	0.8169 (5.6)	0.8233 (5.7)
Base: Only urban area				
16. Top-up-travel cards-SP:				
1: Ticket offices and agencies (topuptc1);	0.09538 (0.4)	0.09321 (0.5)	0.08956 (0.3)	0.08989 (0.3)
2: Ticket offices and agencies and banks (topuptc2)	0.6987 (3.3)	0.7801 (3.5)	0.7230 (3.1)	0.7233 (3.0)
Base: Ticket offices				
17. Top-up-smart cards-SP:				
1: Ticket offices, banks, agencies (topup1);	0.3282 (2.5)	0.3637 (2.1)	0.3418 (2.2)	0.3462 (2.1)
2: Ticket offices, banks, agencies, self-adding value machine (topup2);	0.6512 (4.1)	0.7155 (4.0)	0.6638 (3.8)	0.6642 (3.3)
3:Ticket offices, banks, agencies, self-adding value machine, telephone and Interne (topup3)	1.3024 (5.5)	1.4268 (5.4)	1.3167 (4.9)	1.2267 (5.1)
Base: Only at ticket offices				
ASC-travel cards (RP):	0.1639 (1.0)	0.1529 (1.0)	0.0989 (0.9)	0.0837 (0.9)
ASC-smart cards (RP):	0.3981 (2.2)	0.3841 (2.1)	0.1209 (1.9)	0.09619 (2.1)
ASC-travel cards (SP1):	0.2197 (2.5)	0.2248 (2.0)	0.3092 (1.9)	0.2044 (2.0)
ASC-smart cards (SP2):	-0.4022 (-3.1)	-0.4198 (-2.1)	-0.229 (-2.2)	0.1297 (1.7)
ASC-smart cards (SP3):	-0.4906 (-4.0)	-0.5022 (-2.3)	-0.2901 (-2.0)	-0.1135 (-1.2)
ASC-smart cards (SP4):	-0.1451 (-1.2)	-0.1493 (-1.0)	0.0985 (1.0)	0.0586 (0.9)
θ1 (SP1, 2 and 4):	1.371 (11.9)	1.434 (11.4)	1.238 (11.0)	1.241 (11.7)
θ2 (SP3):	1.719 (10.0)	1.892 (9.8)	1.622 (9.7)	1.602 (9.4)
Likelihood:	-1402.518	-1109.2901	-799.852	-521.673
No. of Observations:	2962	1785	957	671
Rho-squared value w.r.t. cons:	0.4153	0.4219	0.4234	0.4246

Appendix D: Calibrated Membership Functions

Input MF of Travel Cost

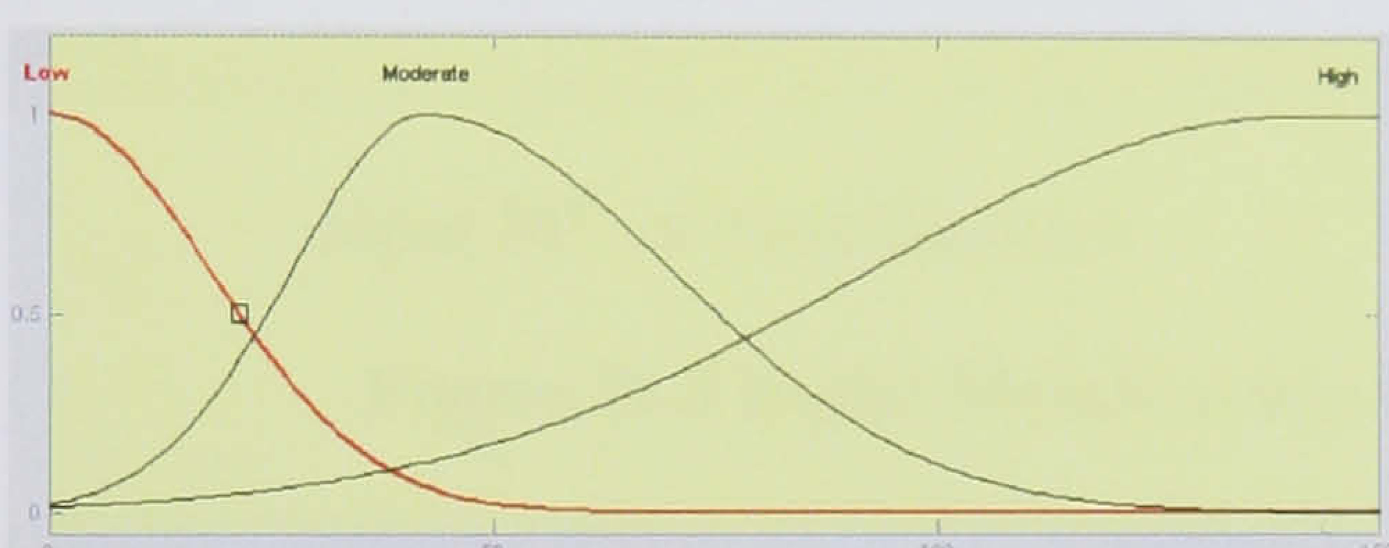


Input MF of Overall Assessment

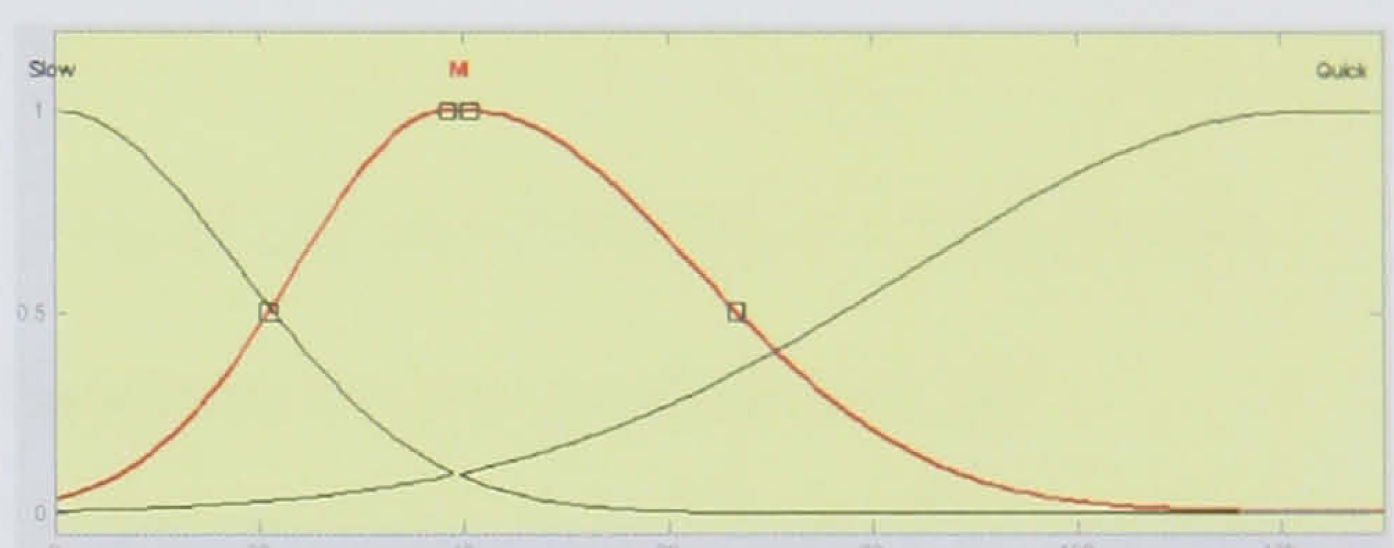


Input MF of Gumbel Error Distribution

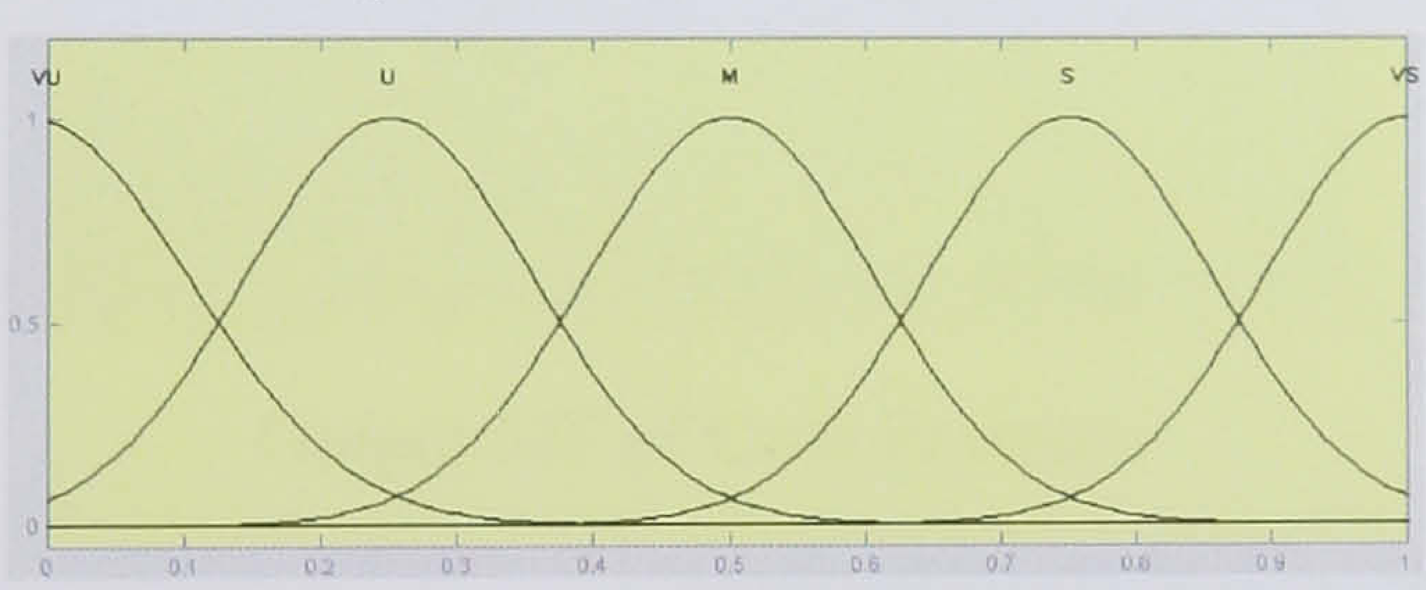
Figure D-1 Input Membership Functions for Cash Fare Payment in RP



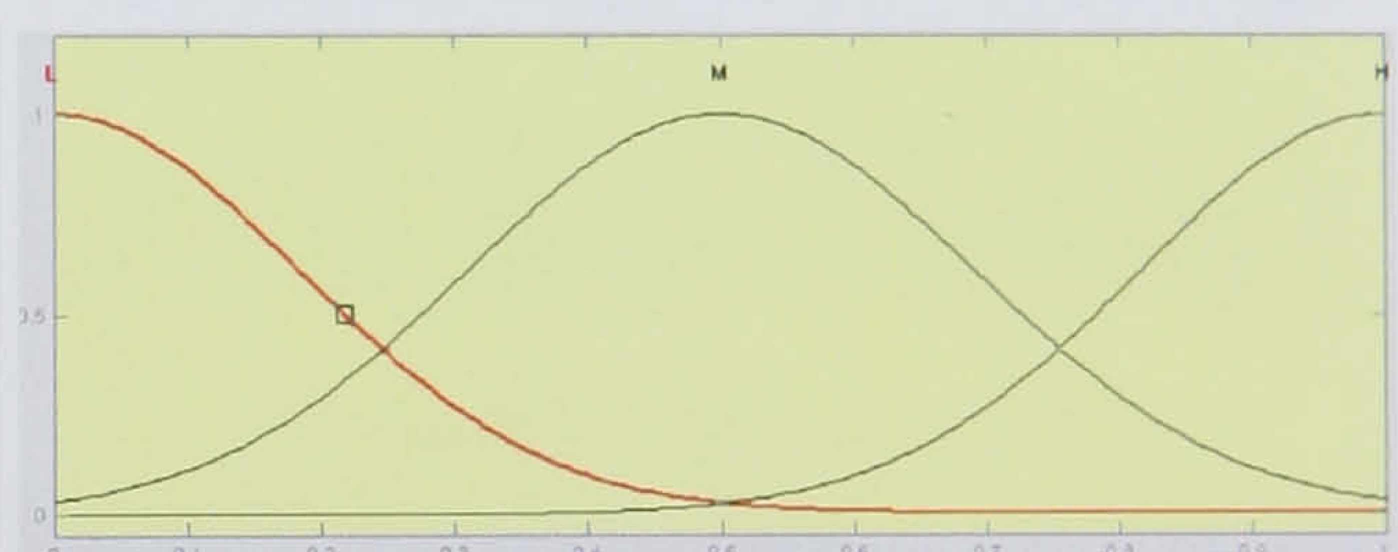
Input MF of Travel Cost



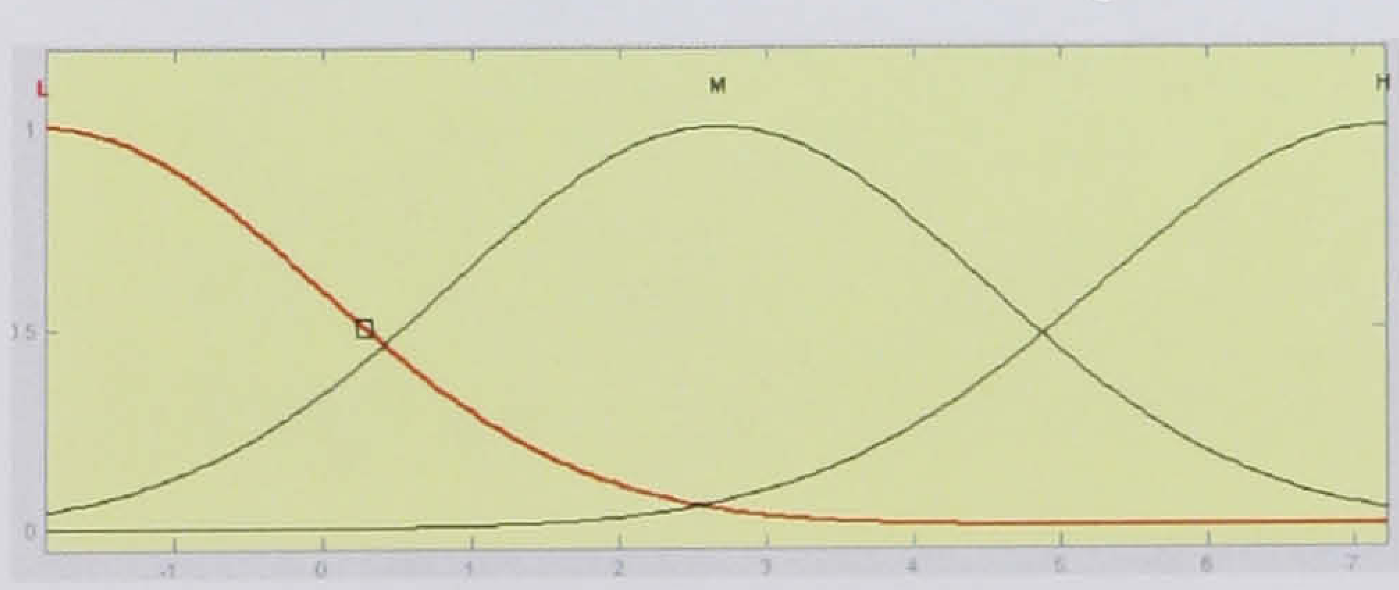
Input MF of Boarding Time Difference



Input MF of Overall Assessment

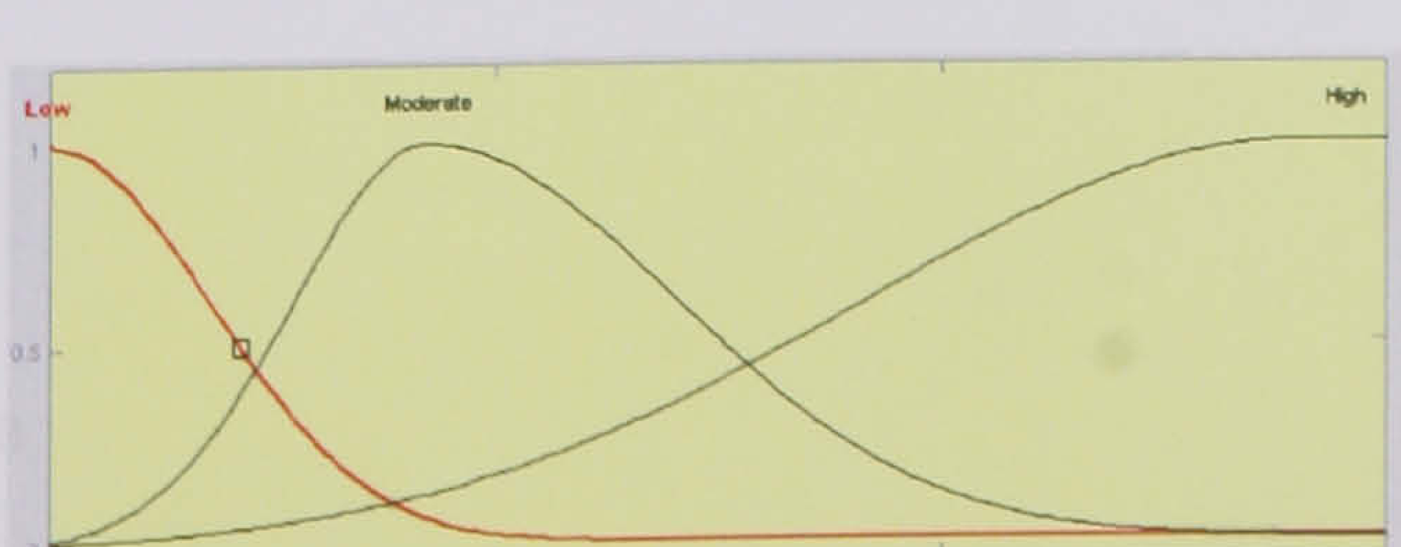


Input MF of Seat Availability

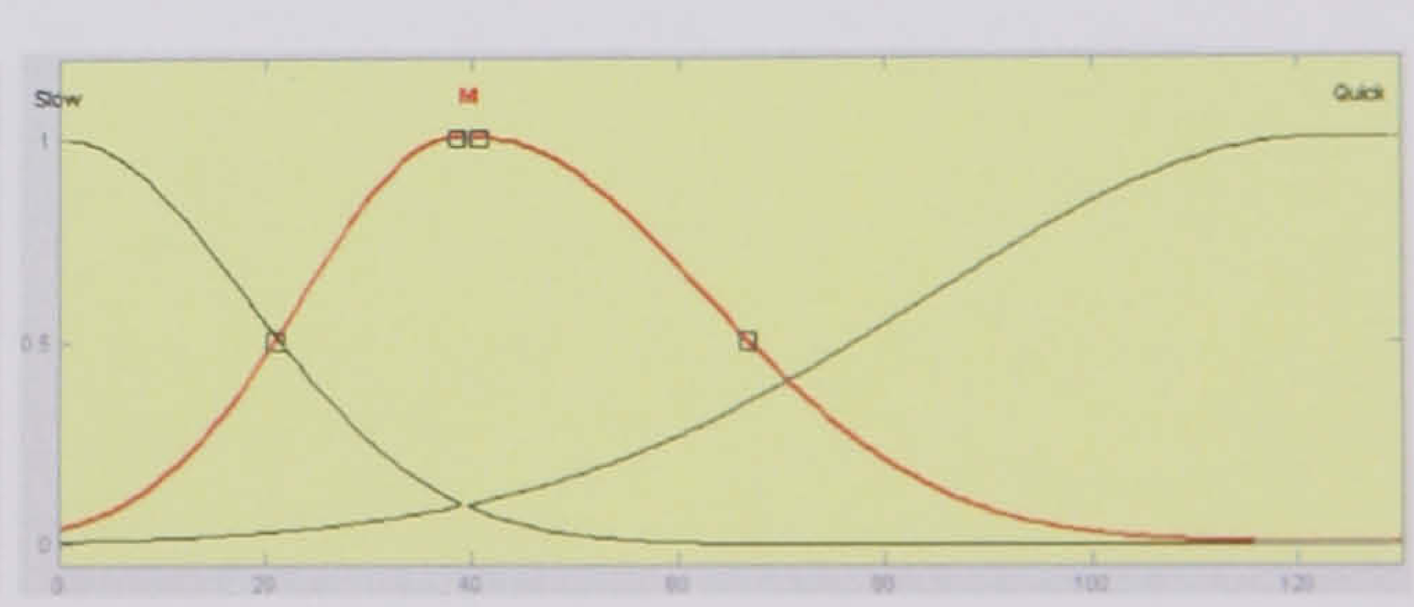


Input MF of Gumbel Error Distribution

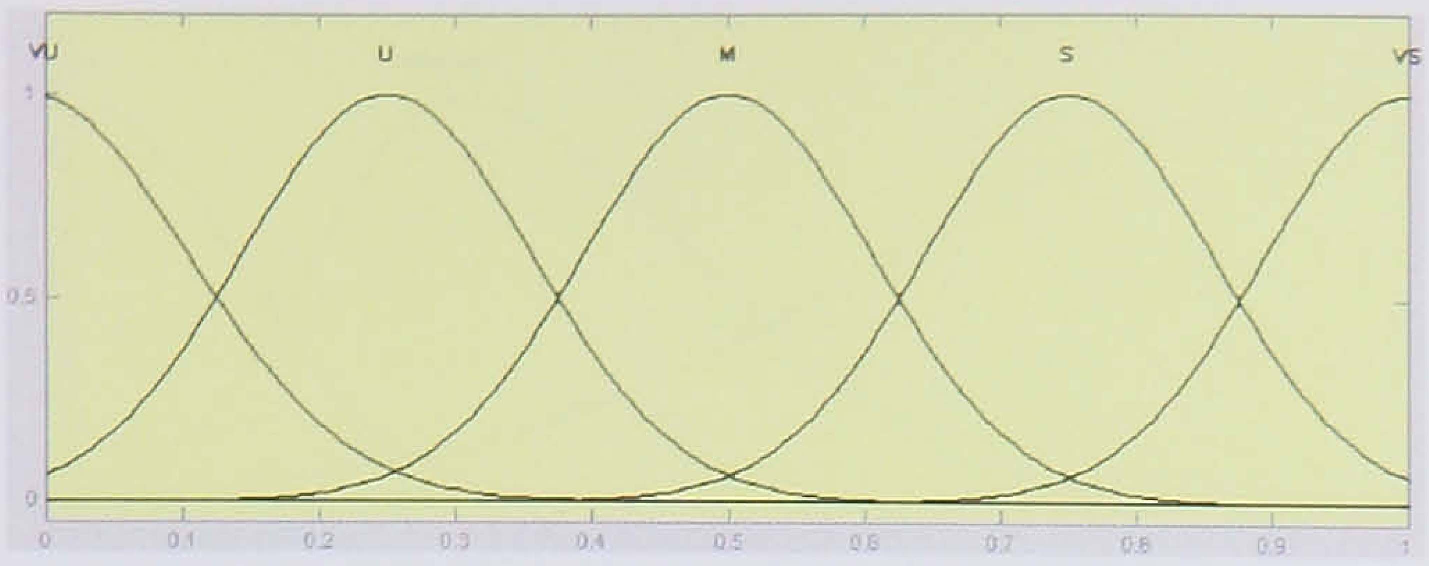
Figure D-2 Input Membership Functions for Travel Cards in RP



Input MF of Travel Cost

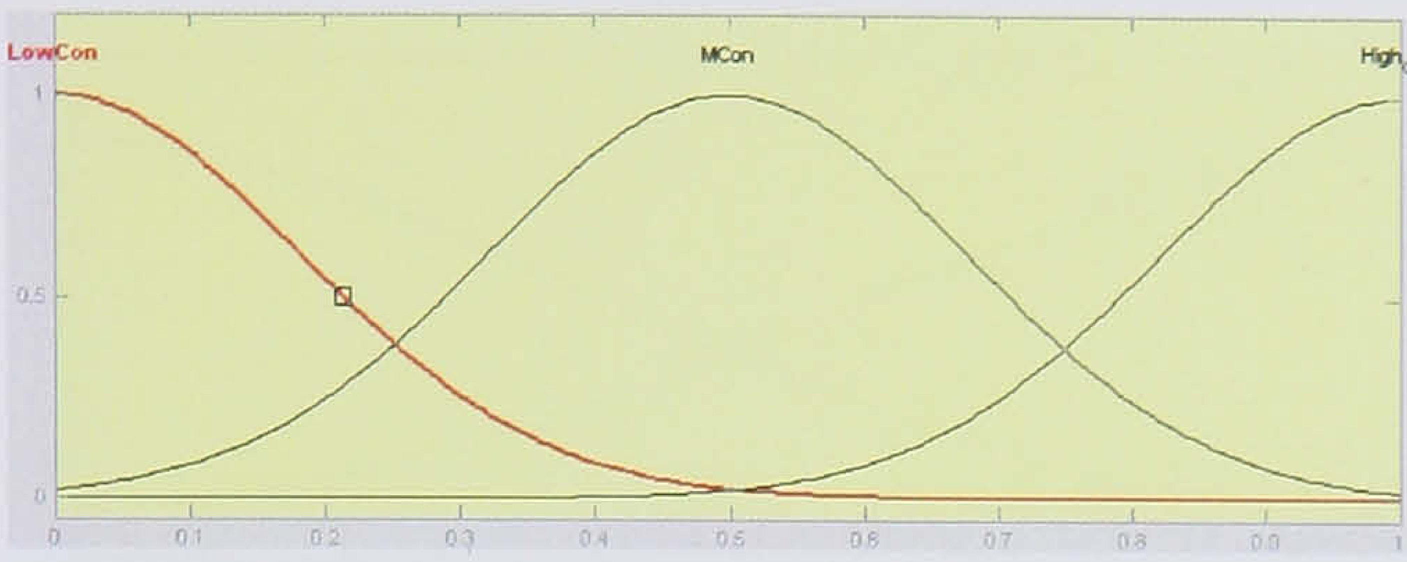


Input MF of Boarding Time Difference



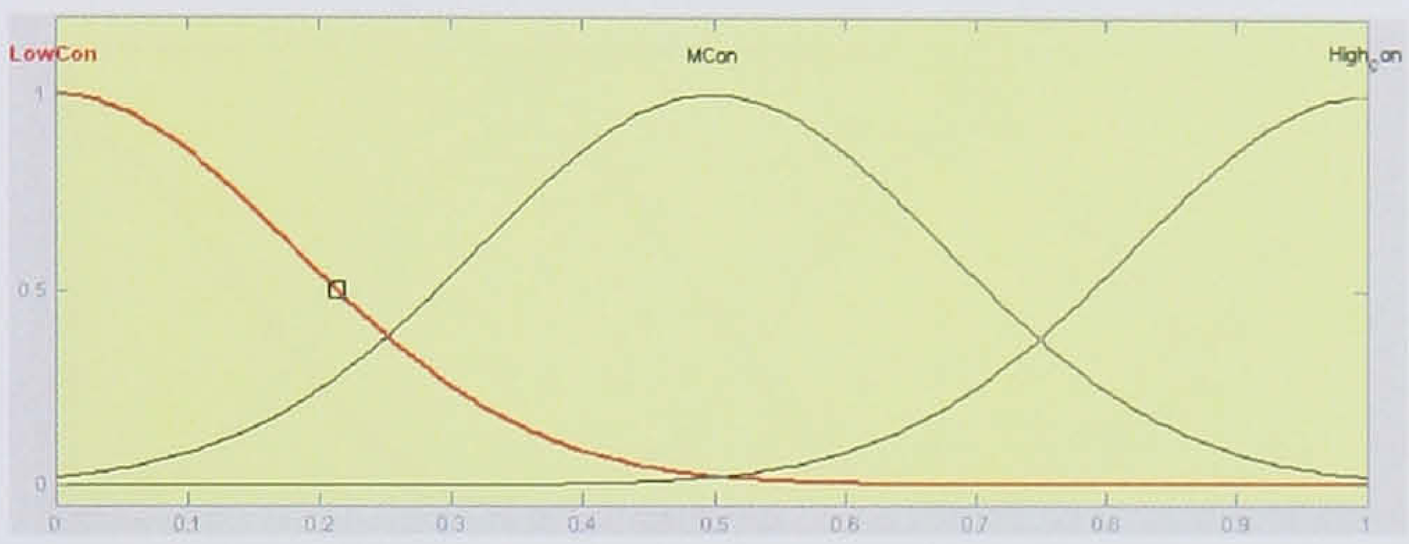
Input MF of Overall Assessment

Input MF of Seat Availability



Input MF of Top-up Methods

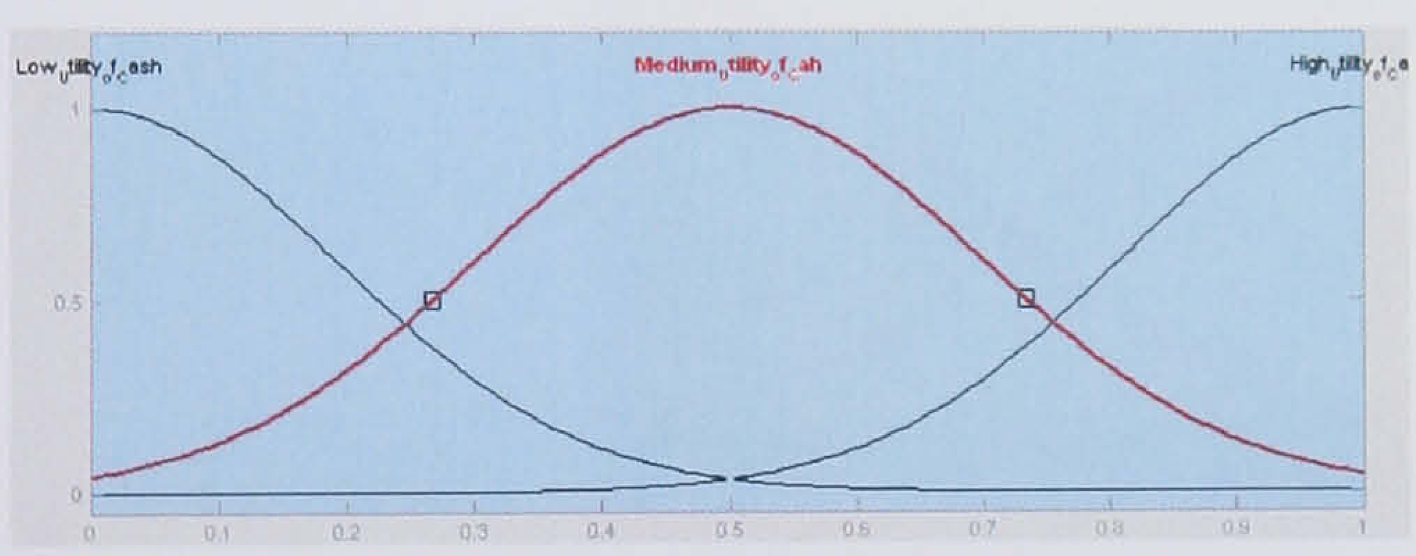
Input MF of Difficulty of Top-up



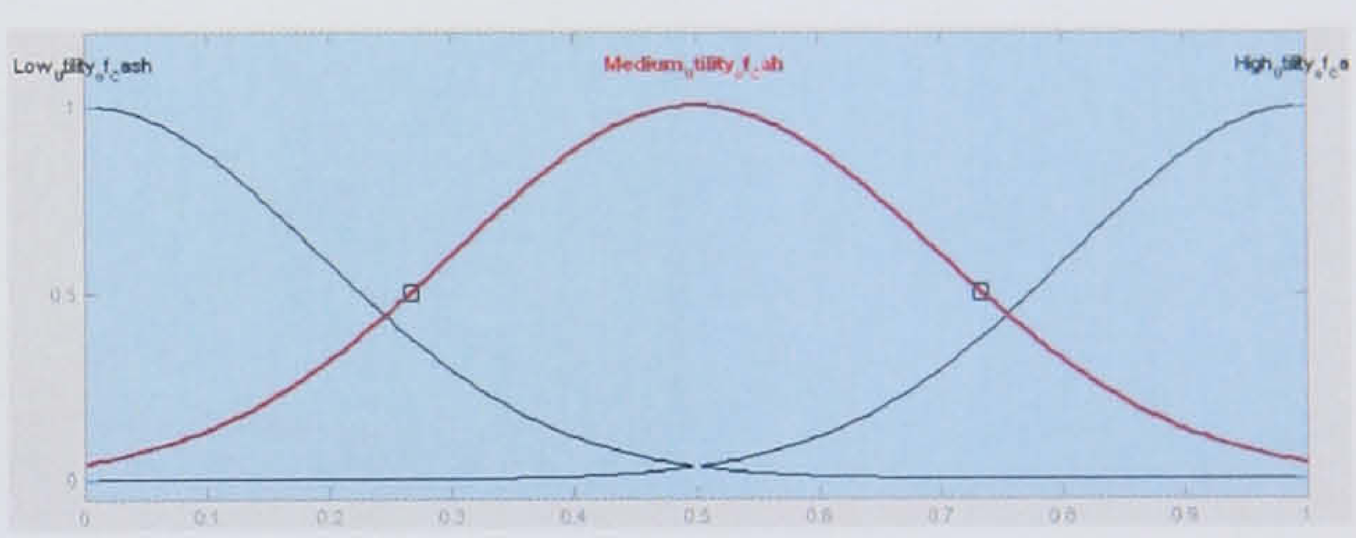
Input MF of Multifunction

Input MF of Gumbel Error Distribution

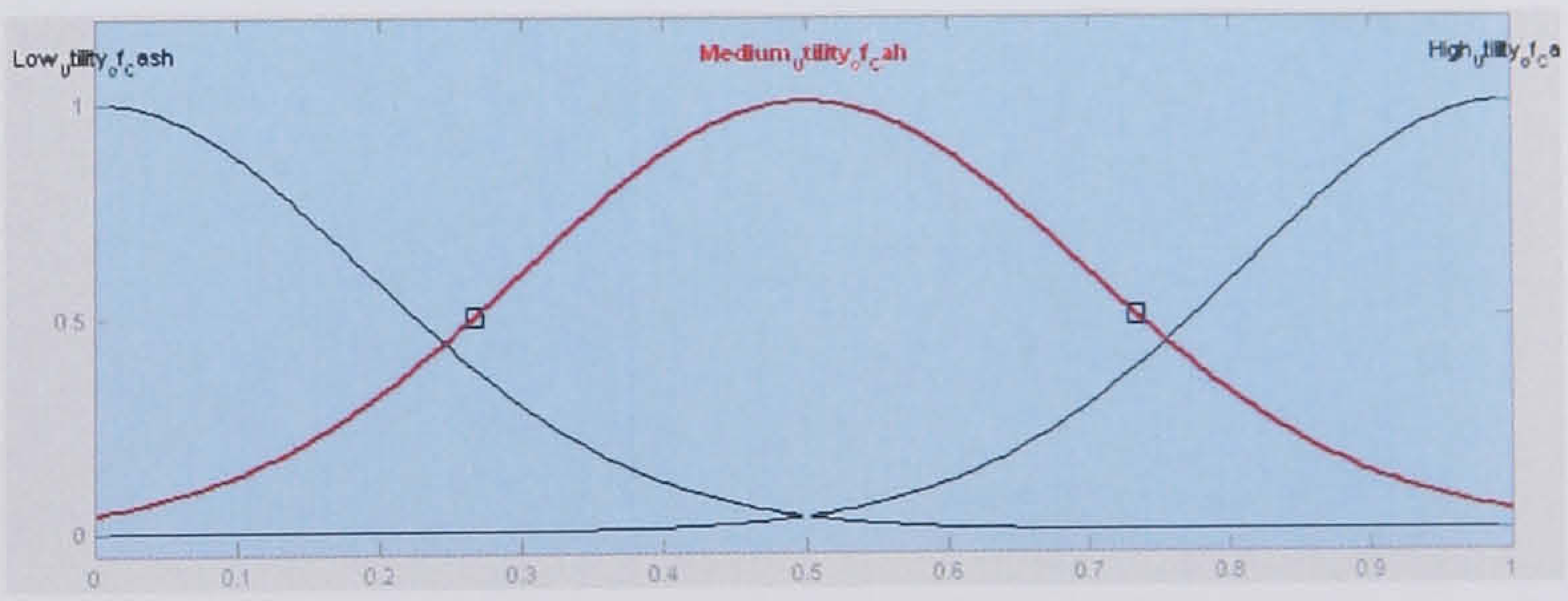
Figure D-3 Input Membership Functions for Smart Cards in RP



Output MF of Cash Payment

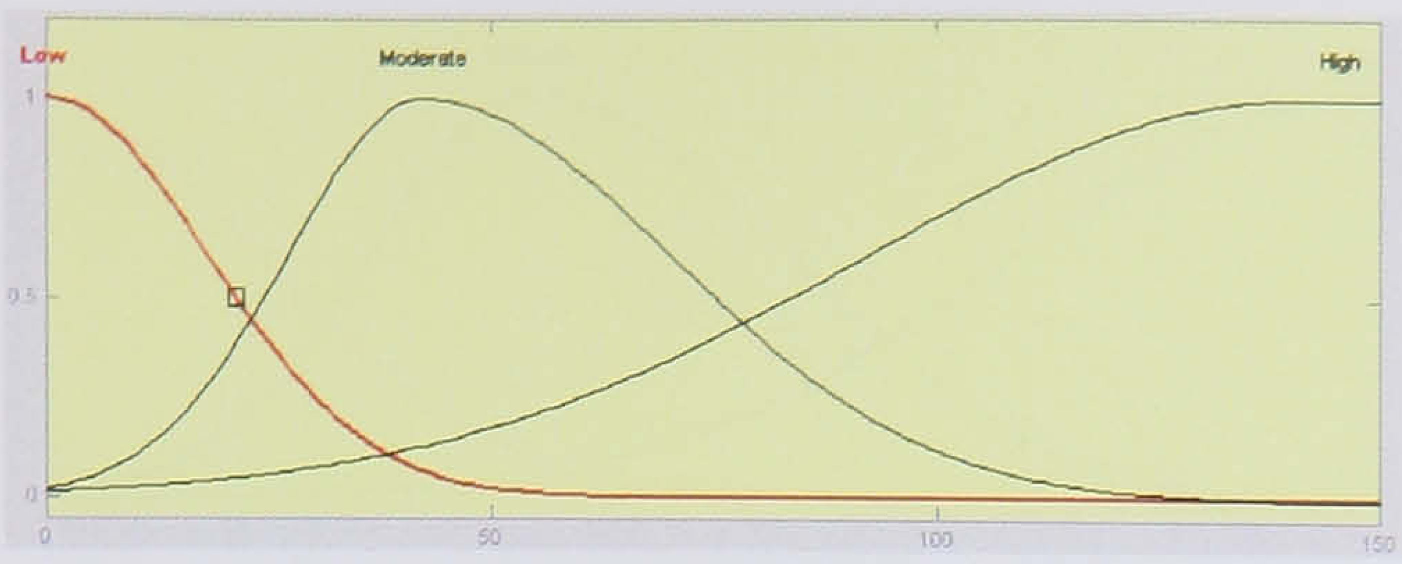


Output MF of Travel Cards



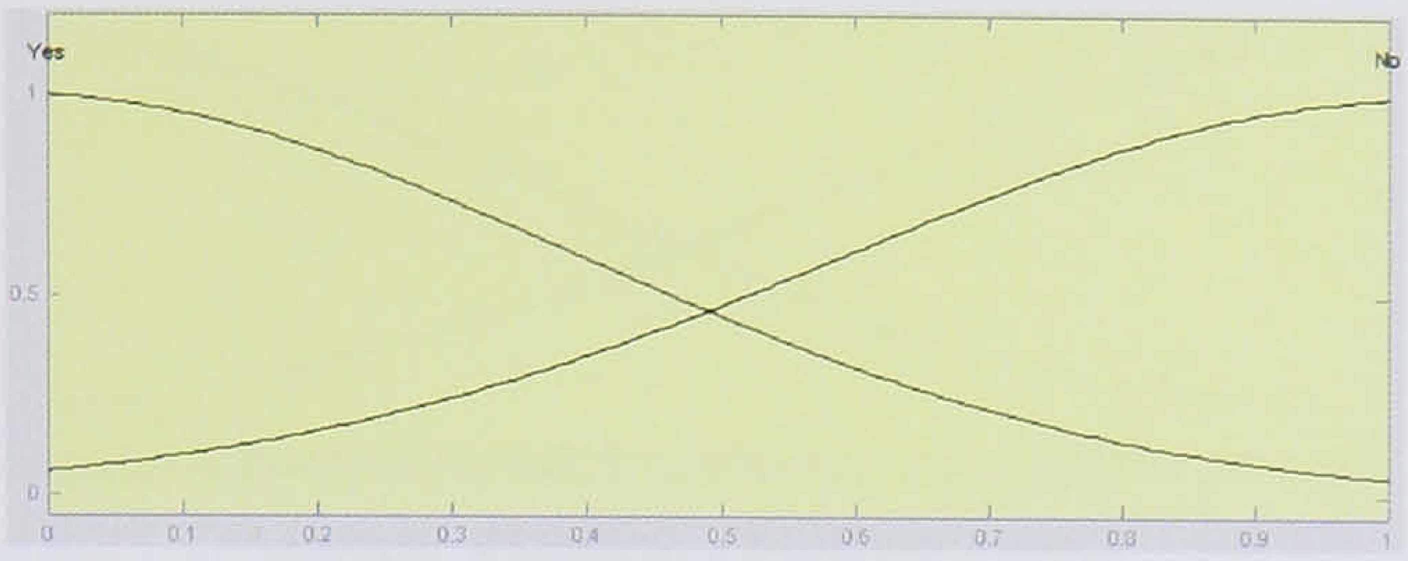
Output MF of Smart Cards

Figure D-4 Output Membership Functions in RP



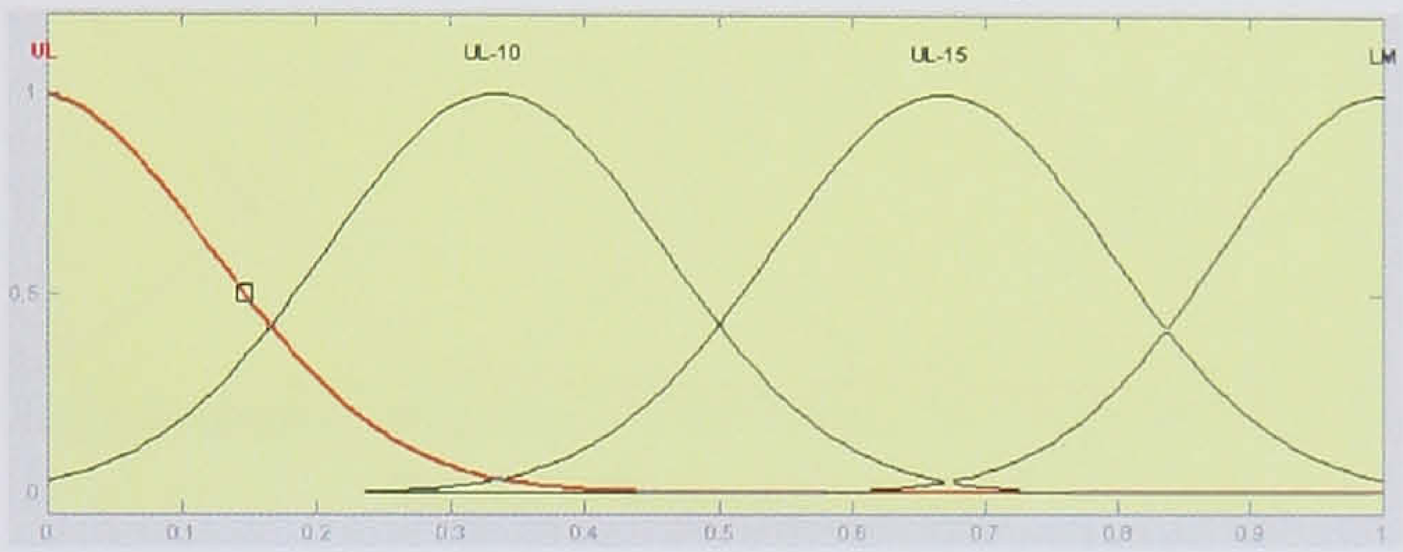
Input MF: Cost of Cash

Input MF: Boarding time of cash



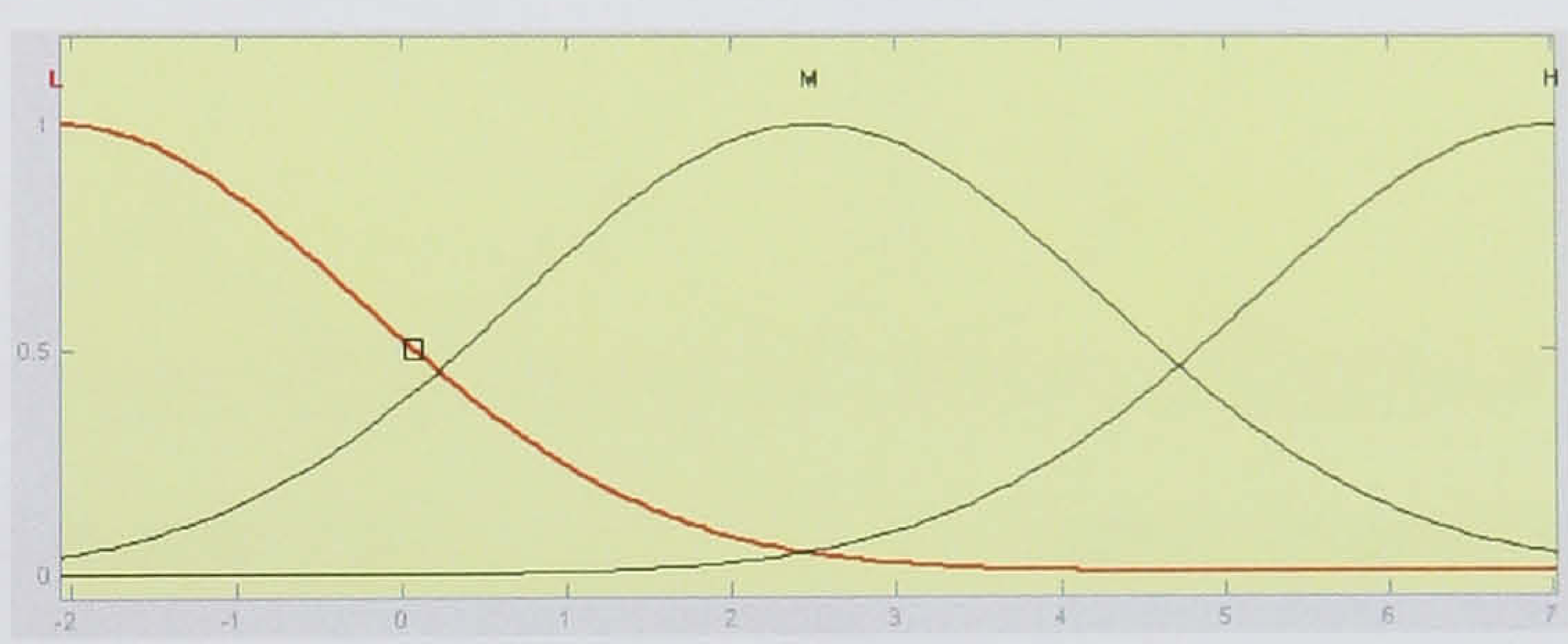
Input MF: Getting changes back

Input MF: Cost of TC



Input MF: Service Routes of Travel Cards

Input MF: Gumbel Error Distribution of Cash



Input MF: Gumbel Error Distribution of Travel Cards

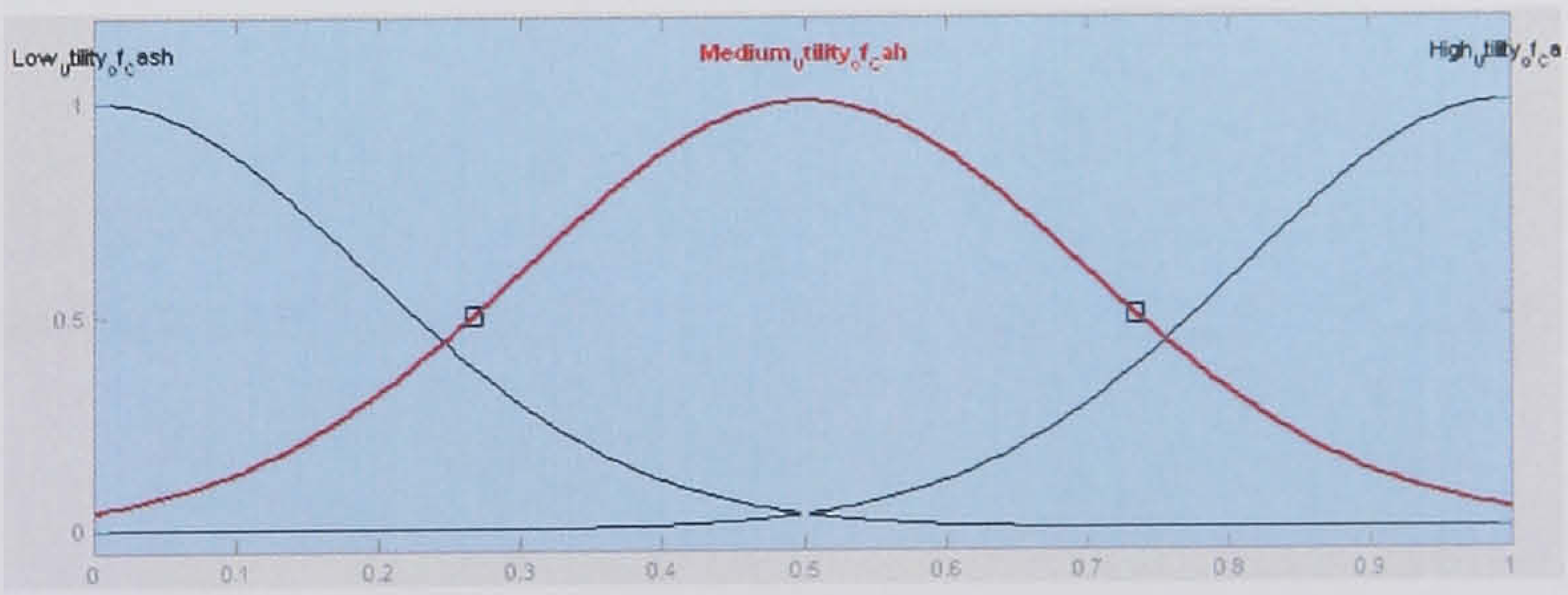
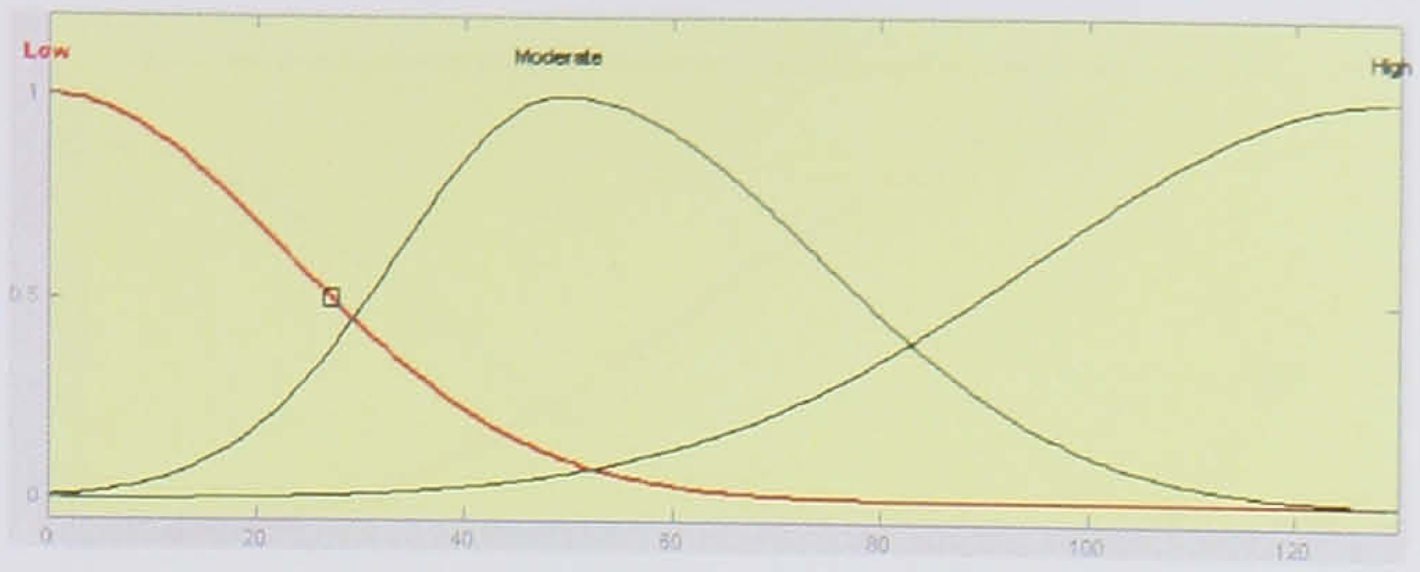
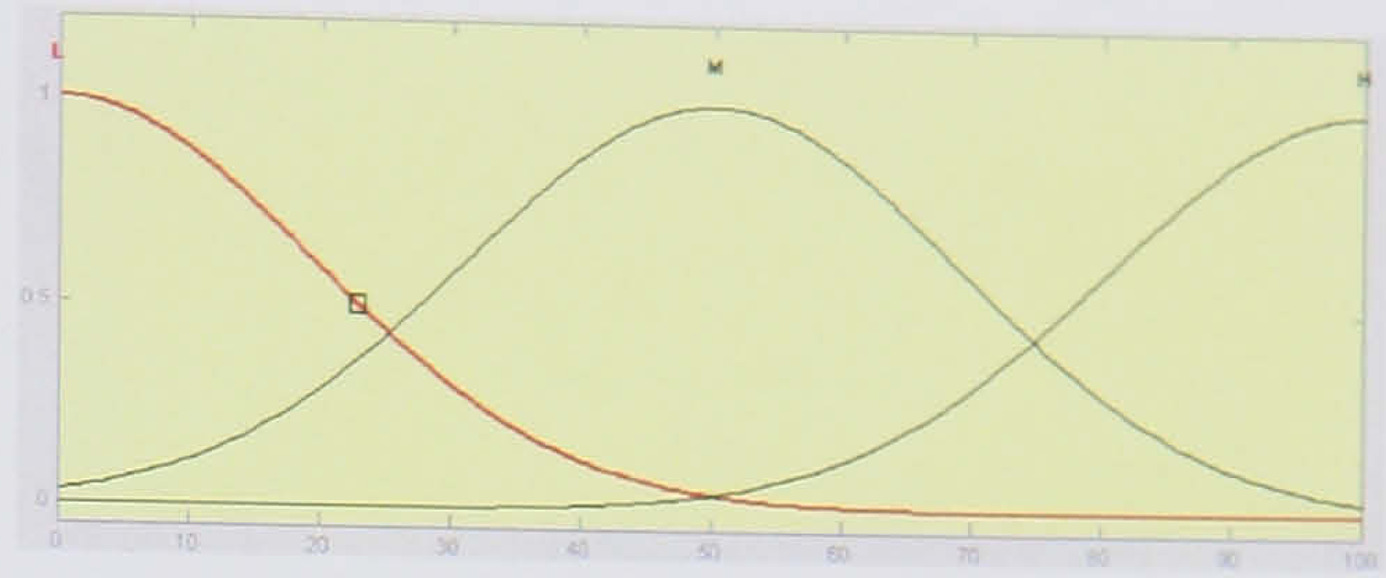


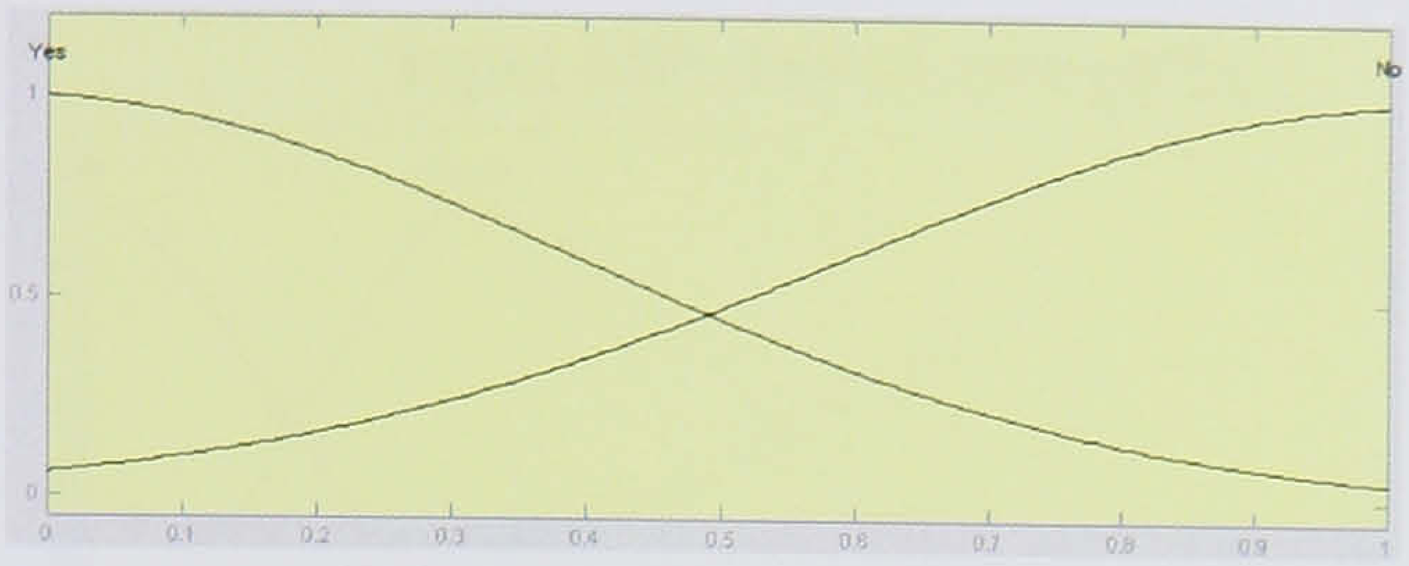
Figure D-5 Membership Functions of Cash and Smart Cards in SP-1



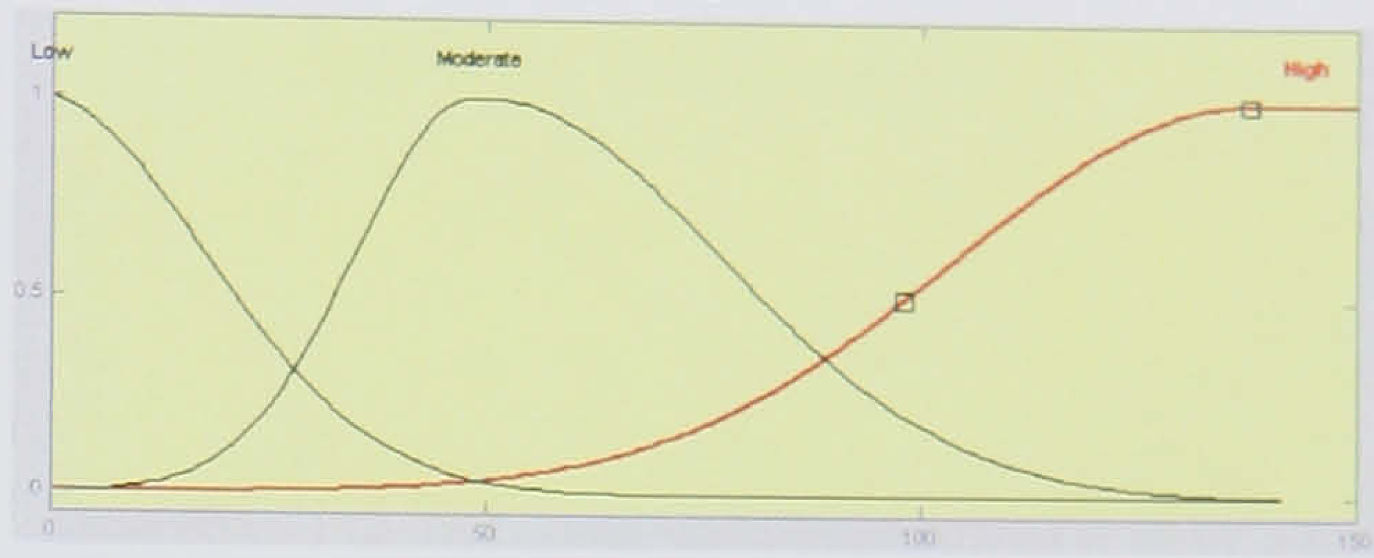
Input: Cost of Cash



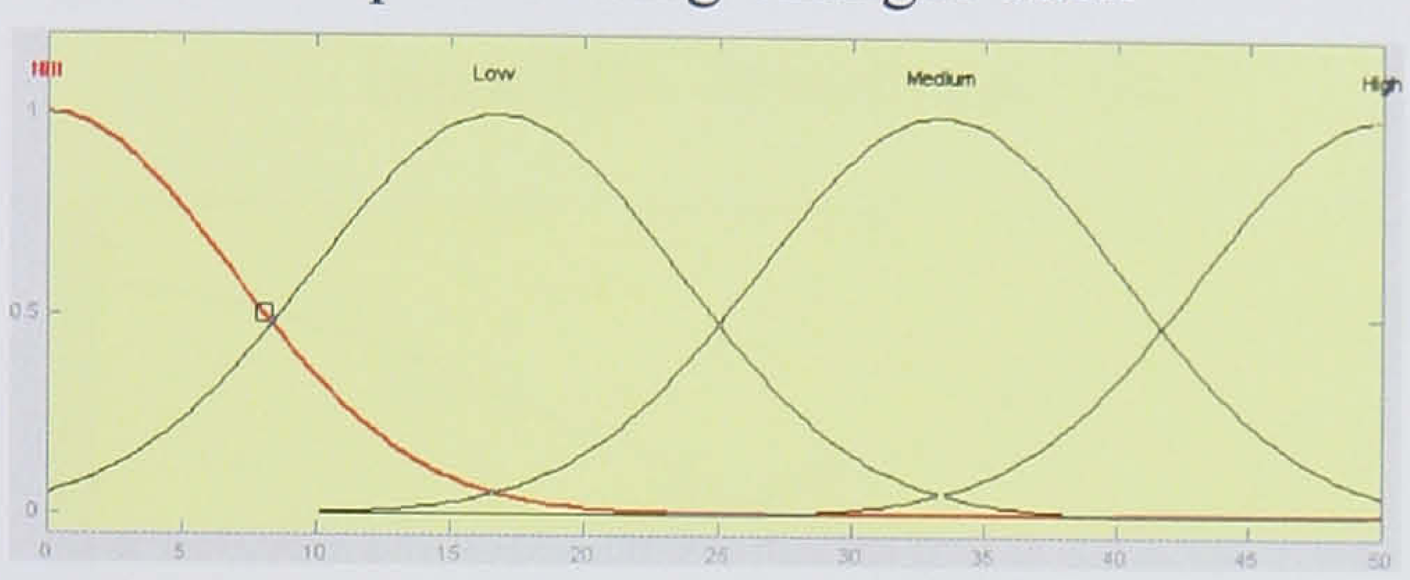
Input: Boarding time of cash



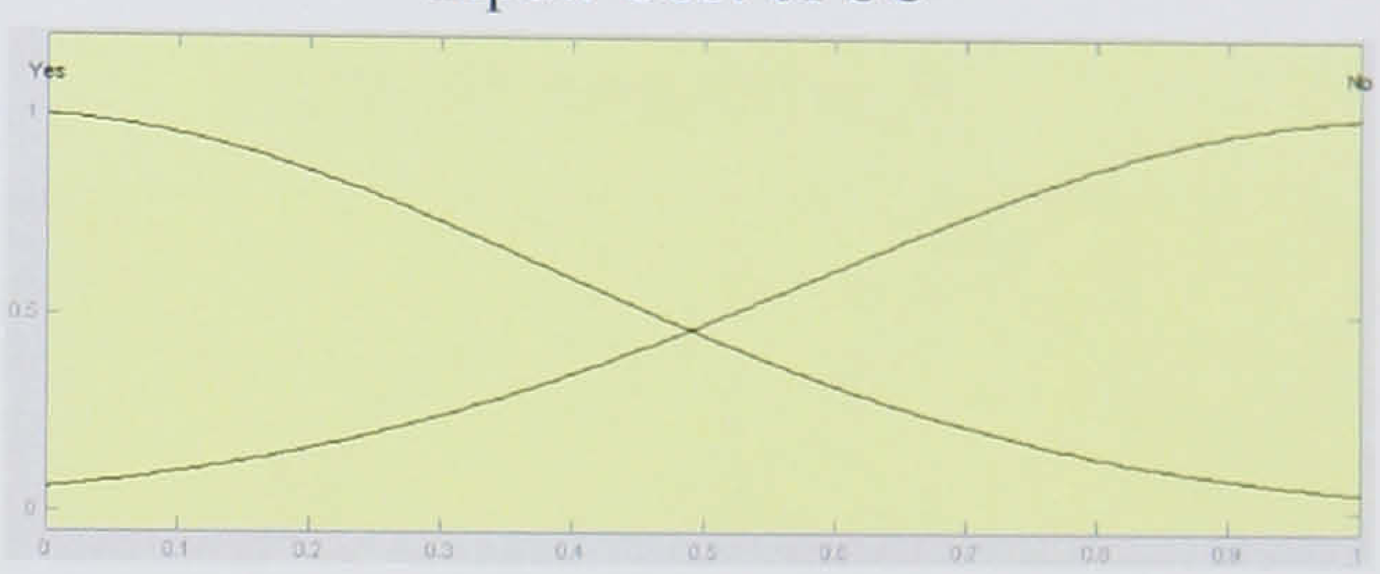
Input: Getting changes back



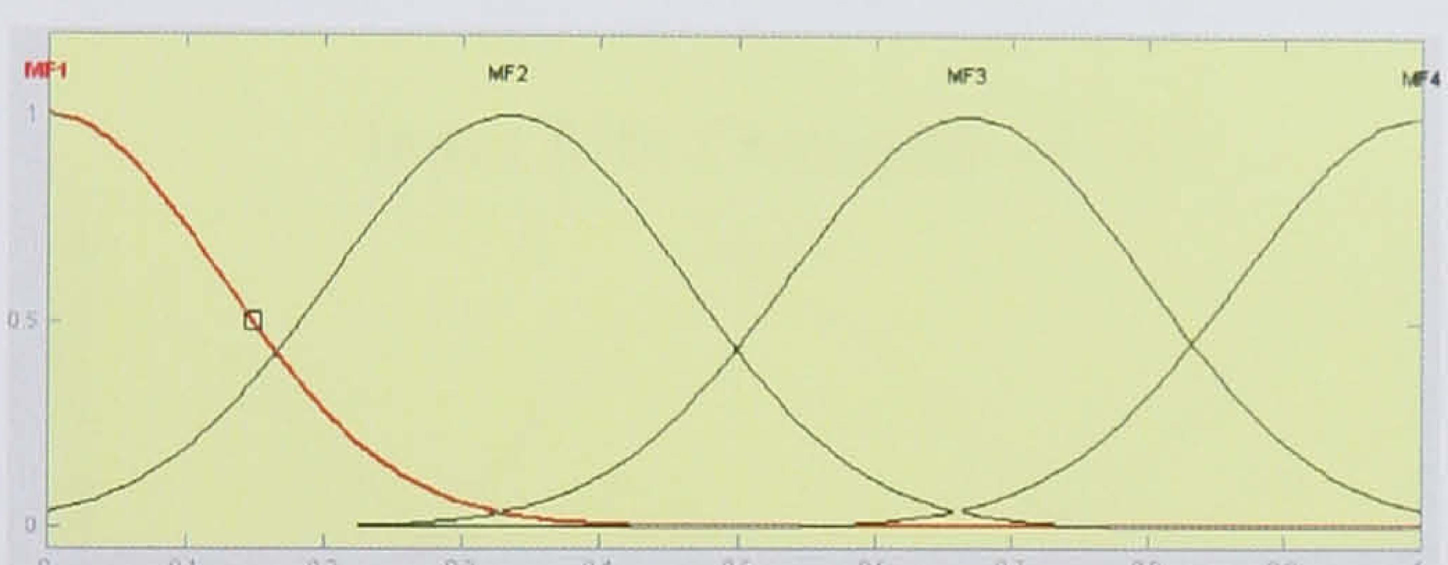
Input: Cost of SC



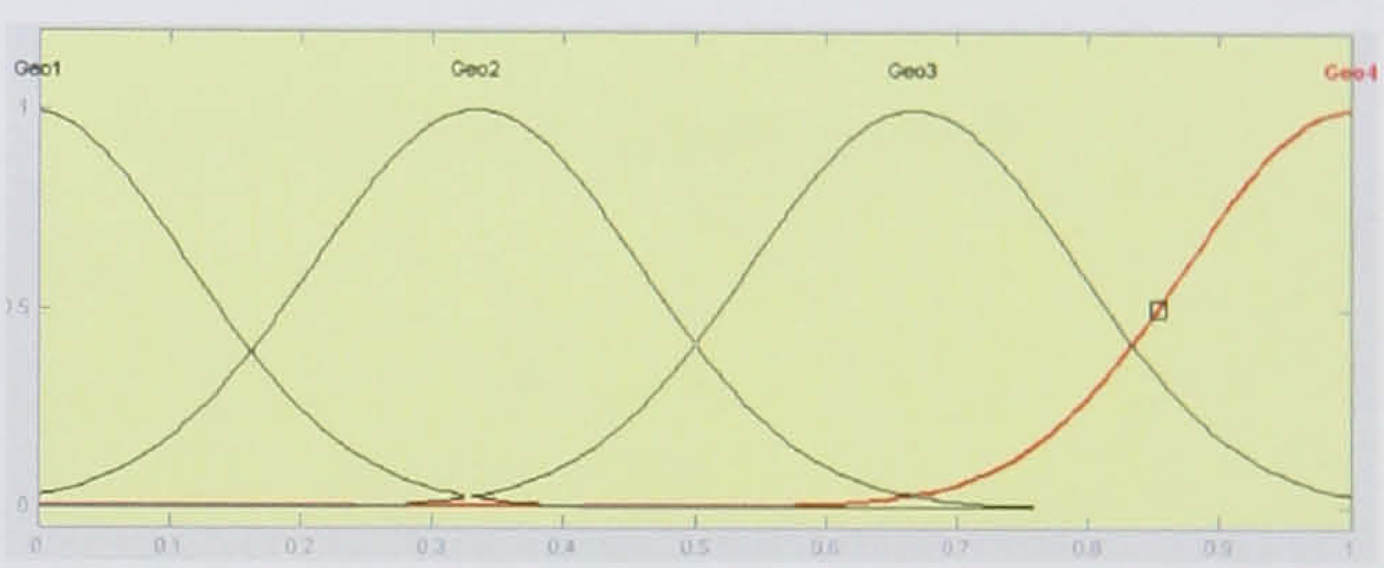
Input: Deposit of SC



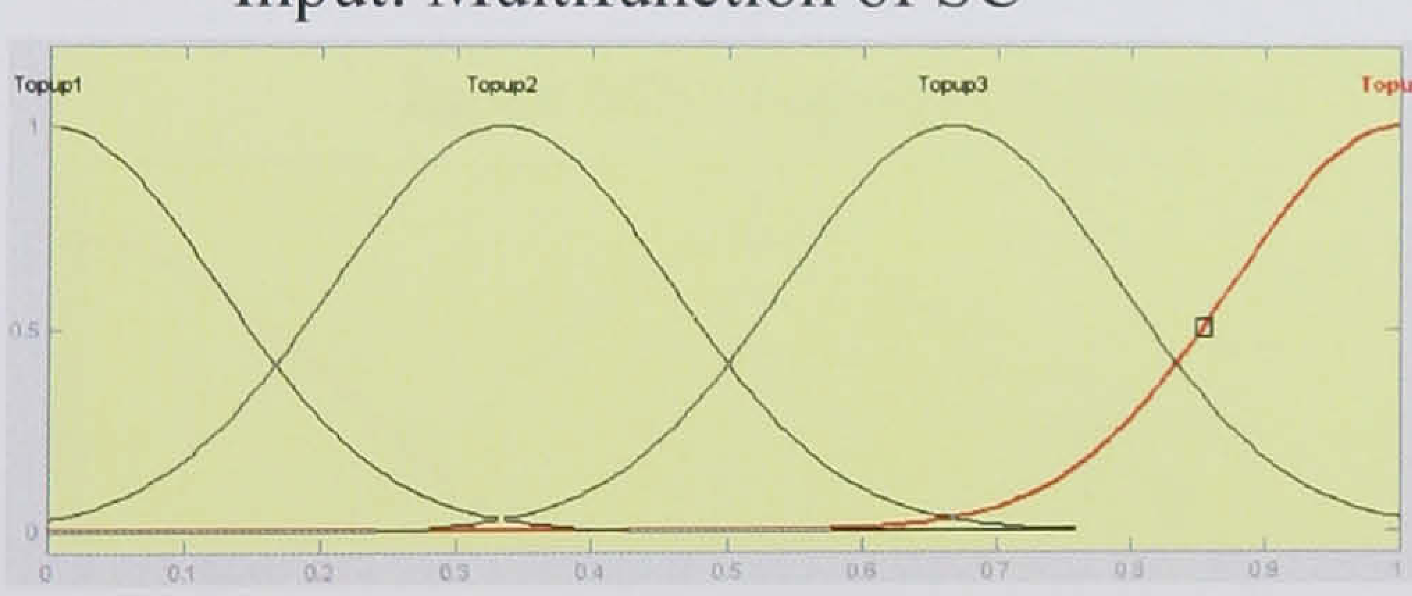
Input: Overdraft of SC



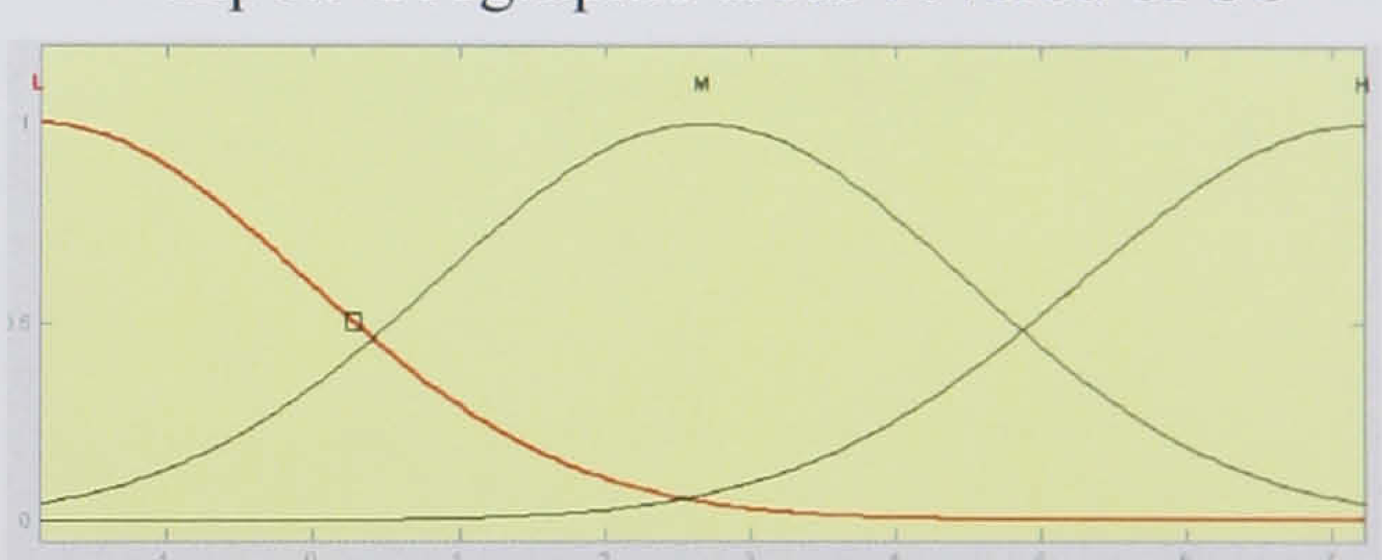
Input: Multifunction of SC



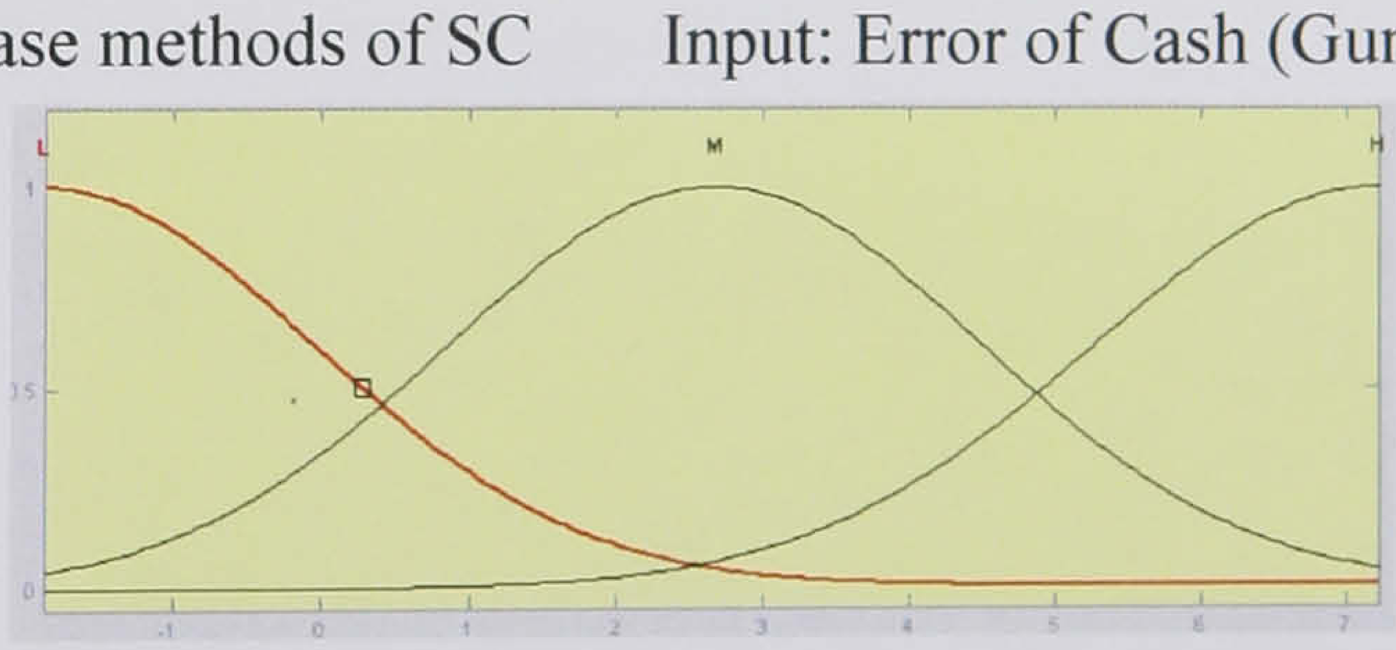
Input: Geographic areas covered of SC



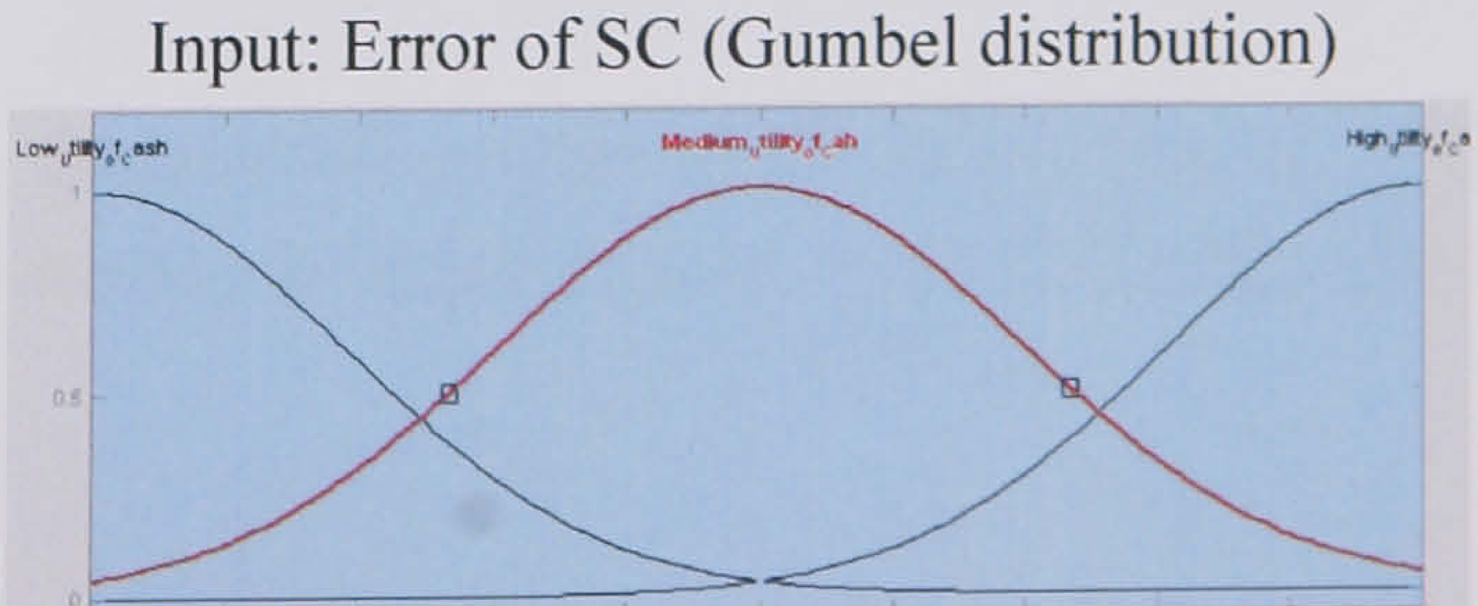
Input: Top-up/purchase methods of SC



Input: Error of Cash (Gumbel distribution)

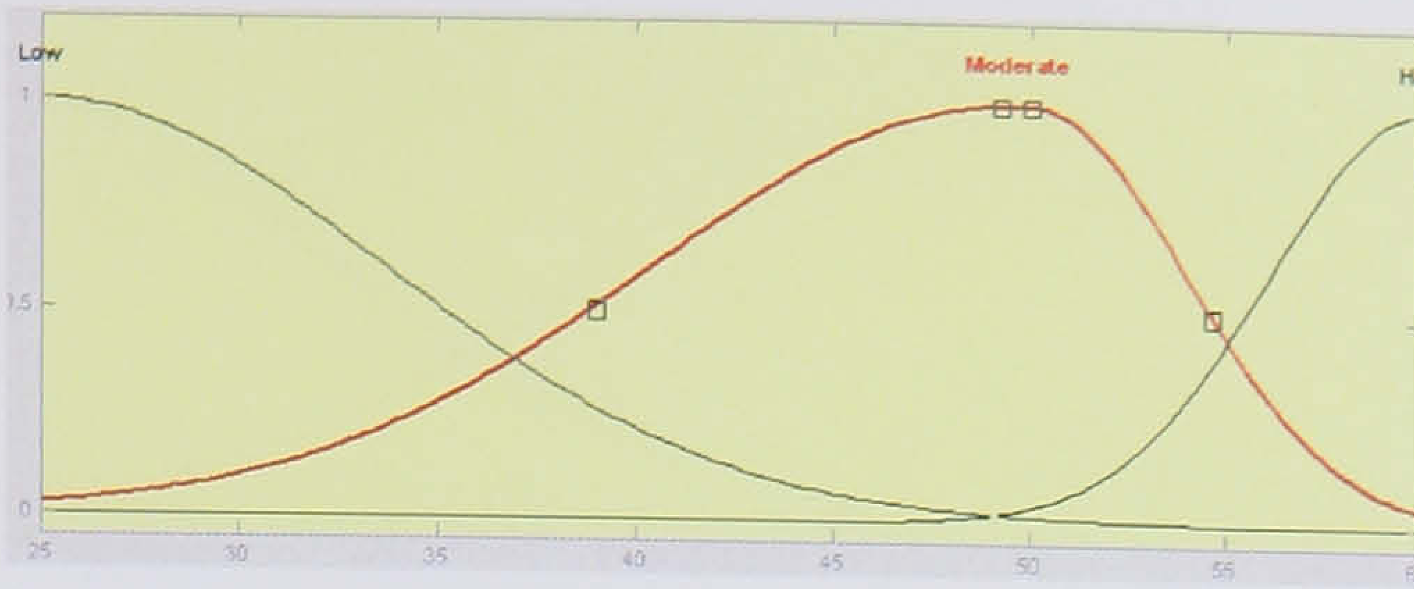


Input: Error of SC (Gumbel distribution)

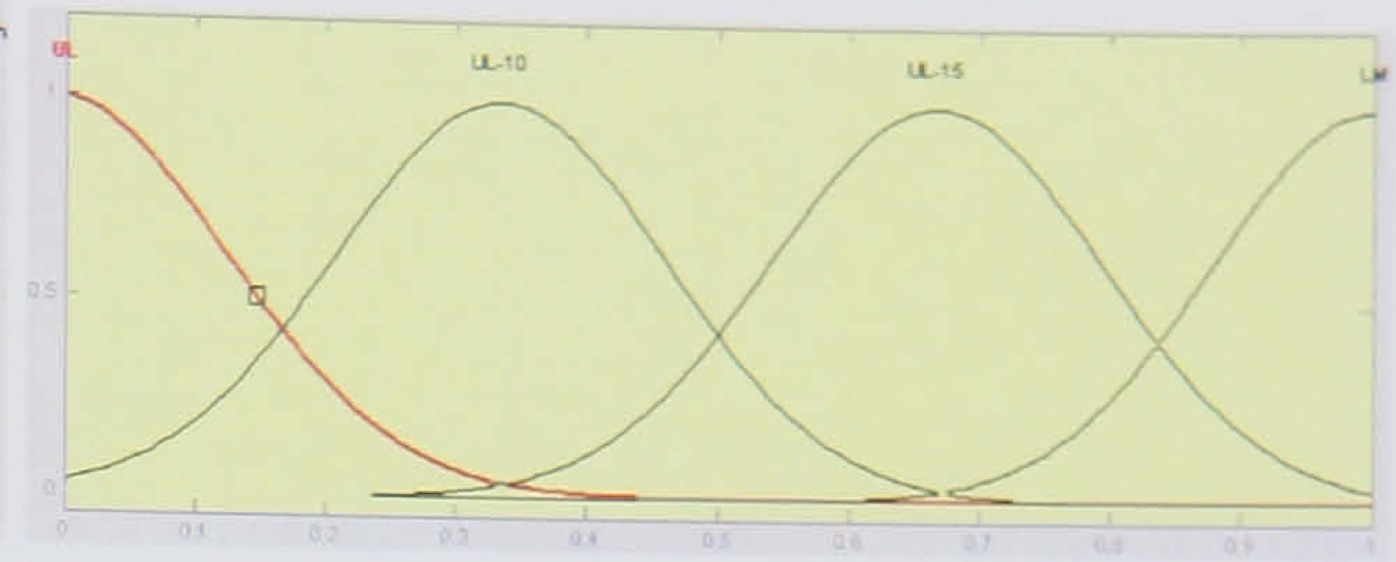


Output MF (choice probability of Cash)

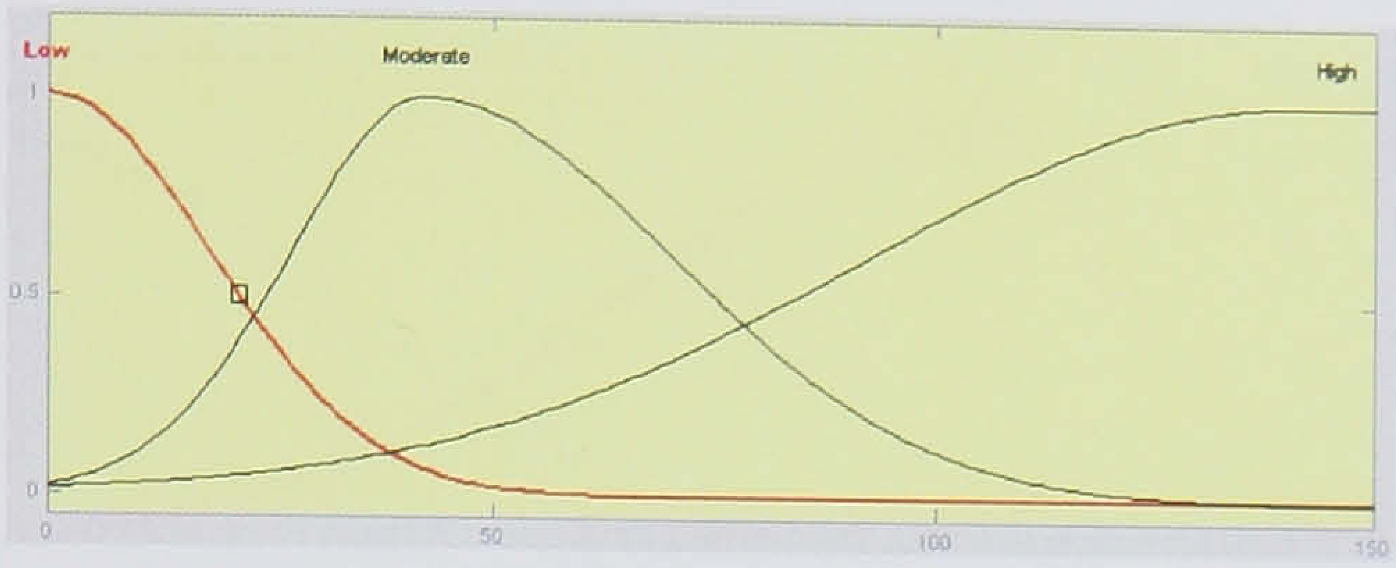
Figure D-6 Membership Functions of Cash and Smart Cards in SP-2



Input MF: Travel cost of TC

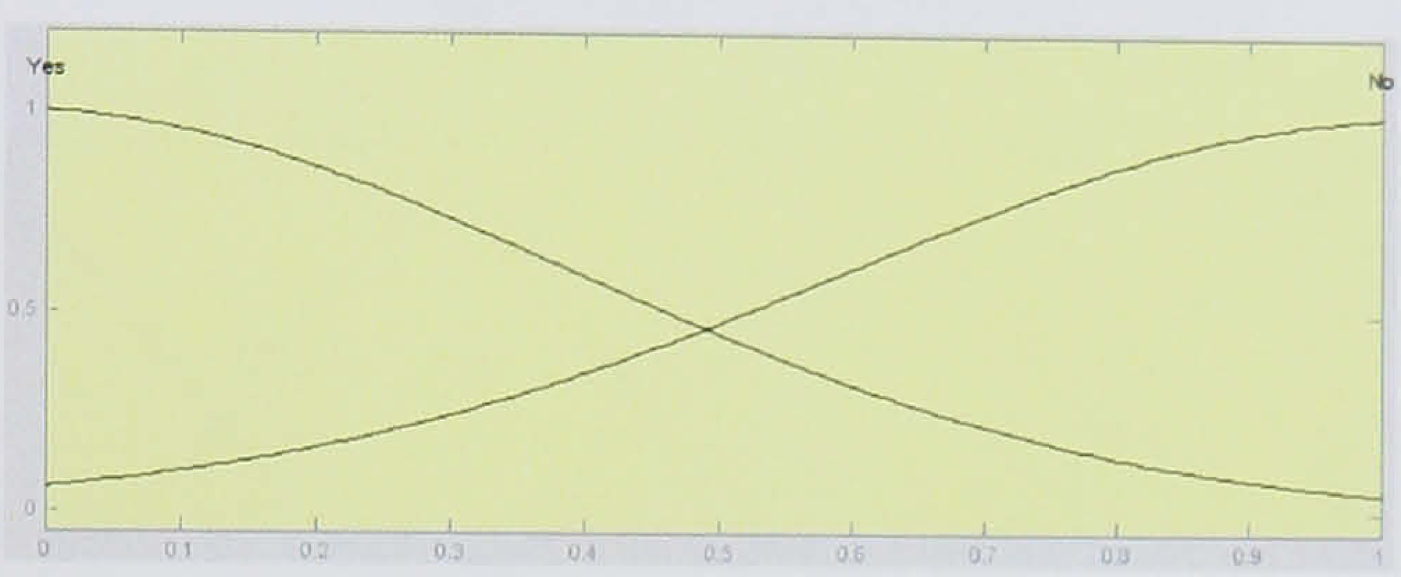


Input MF: Service of TC

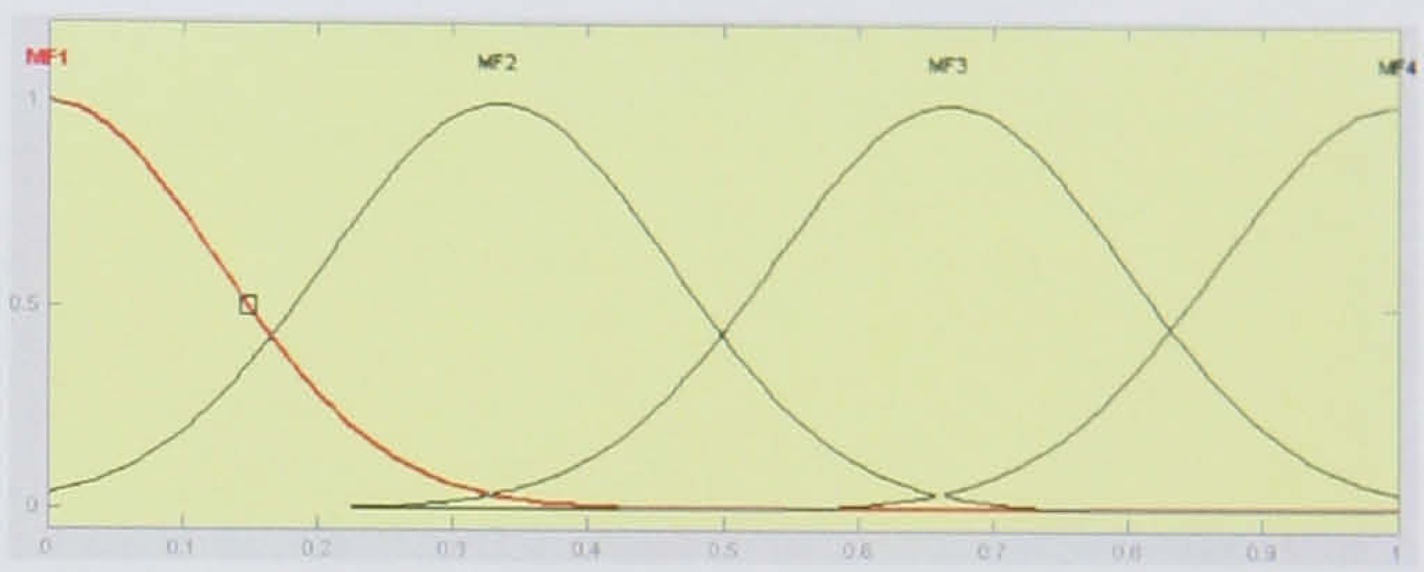


Input MF: Travel cost of SC

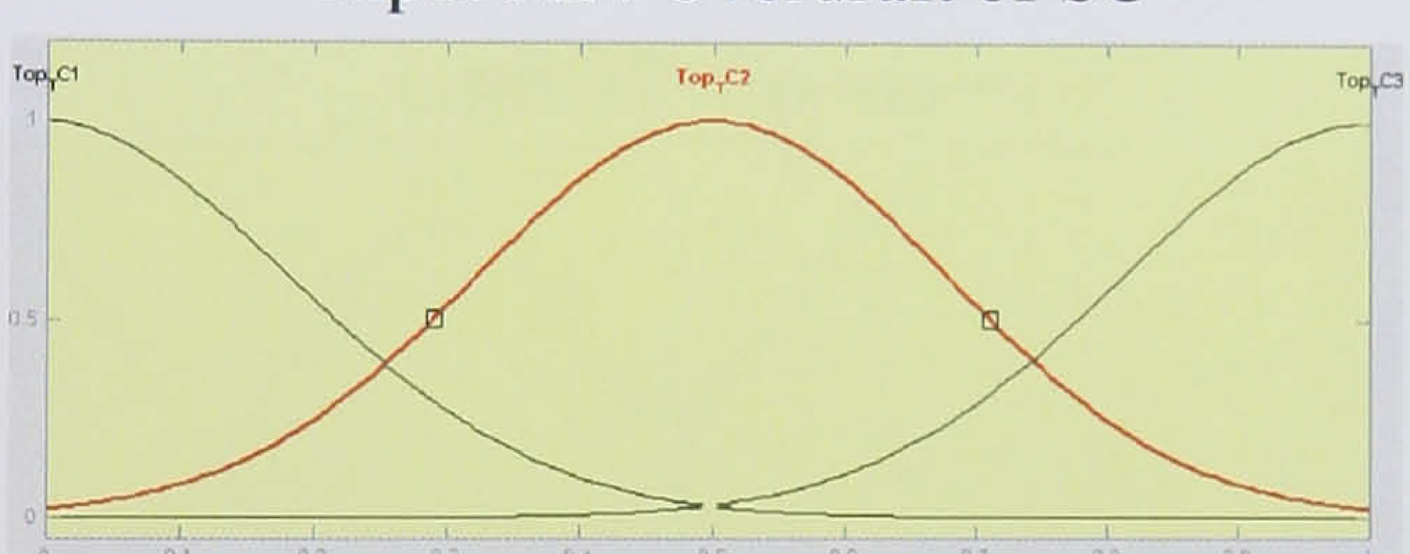
Input MF: Deposit of SC



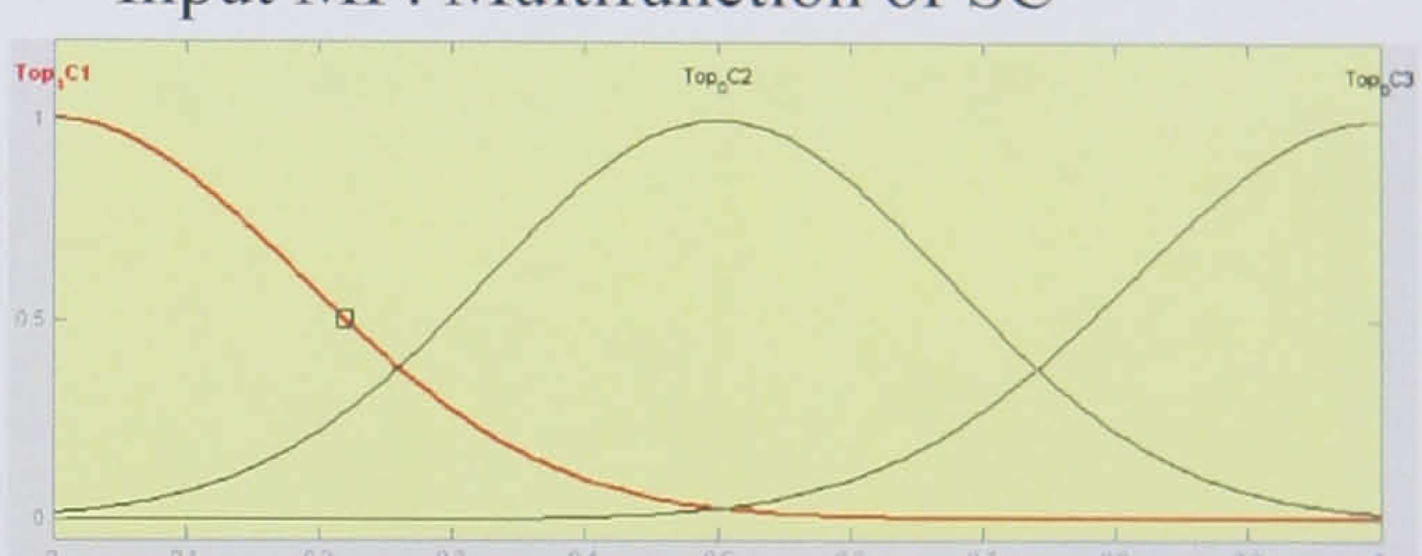
Input MF: Overdraft of SC



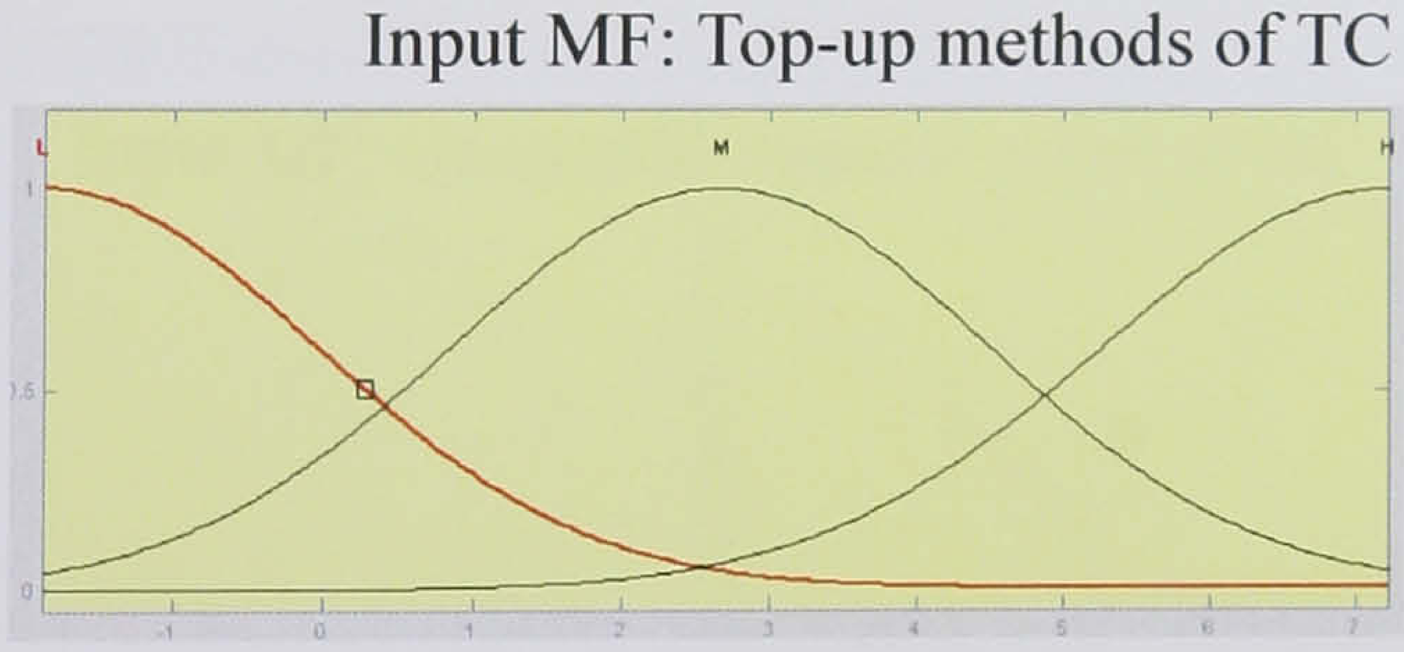
Input MF: Multifunction of SC



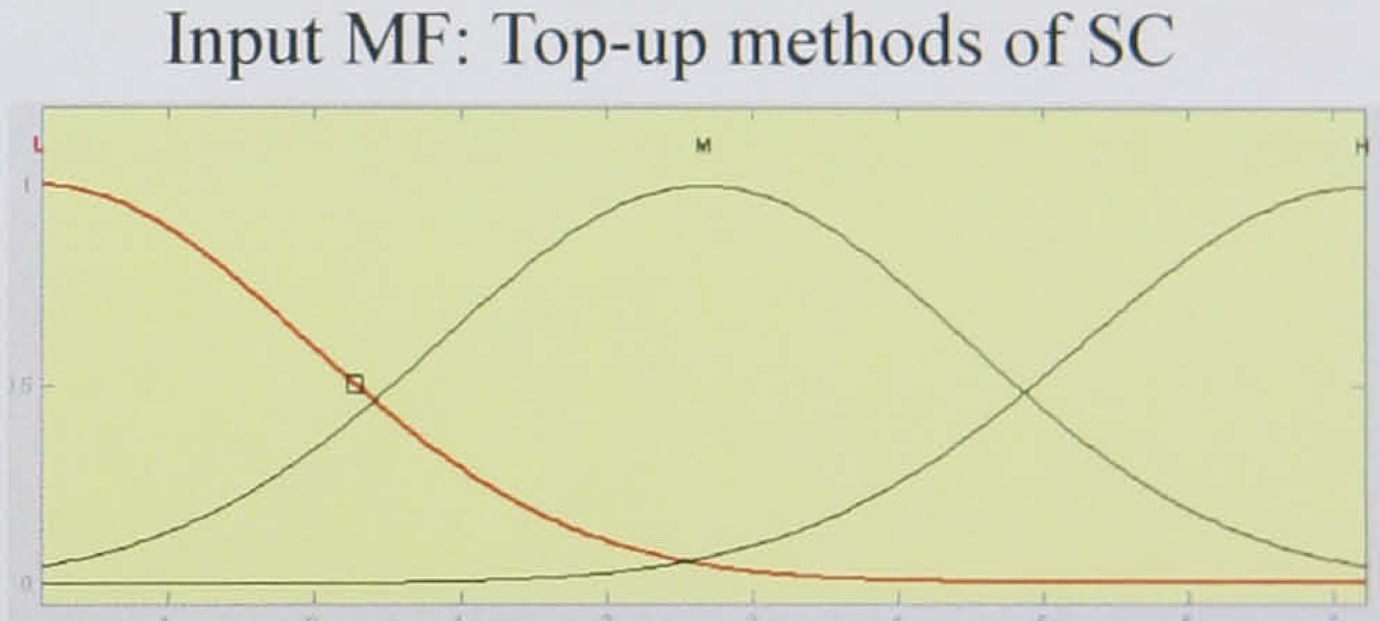
Input MF: Top-up methods of TC



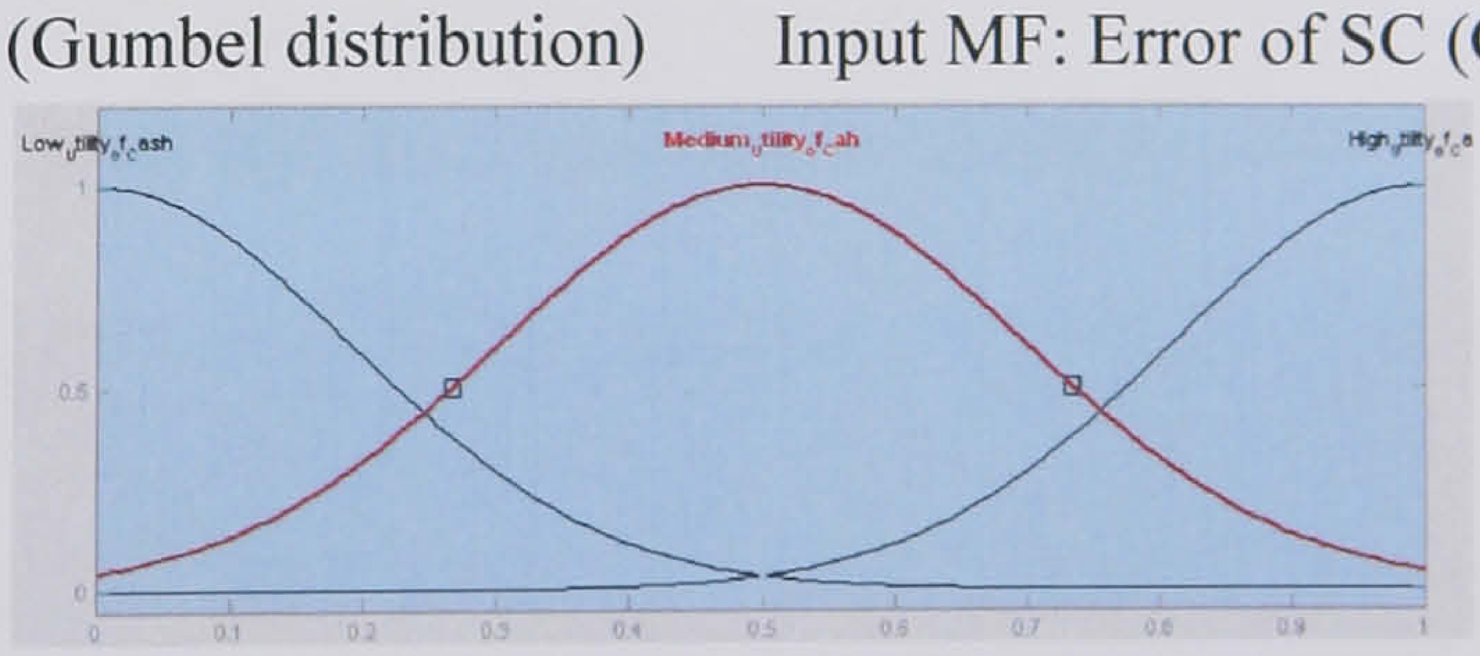
Input MF: Top-up methods of SC



Input MF: Error of TC (Gumbel distribution)

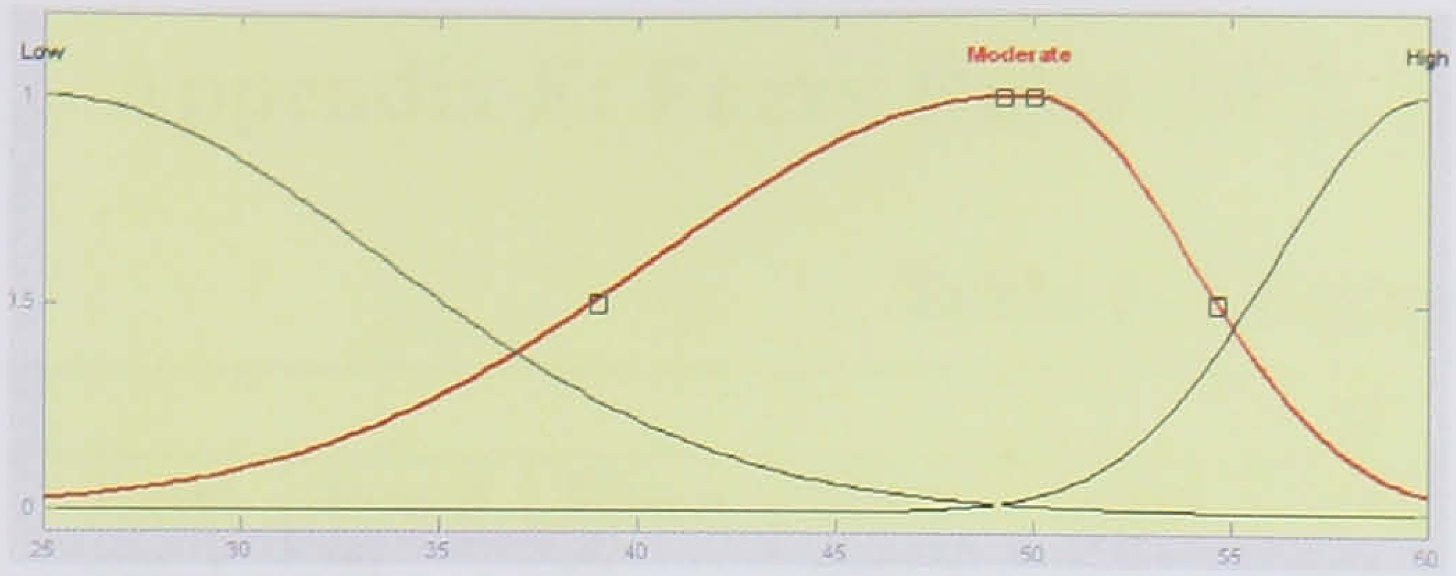


Input MF: Error of SC (Gumbel distribution)

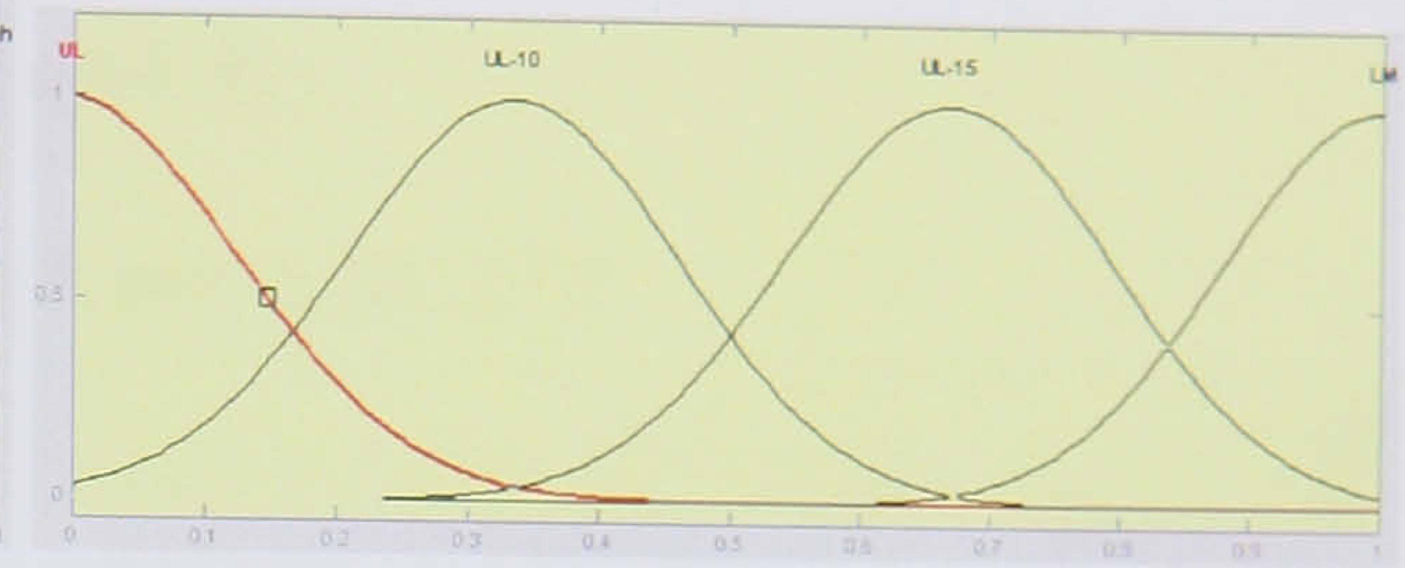


Output MF (choice probability of Travel Cards)

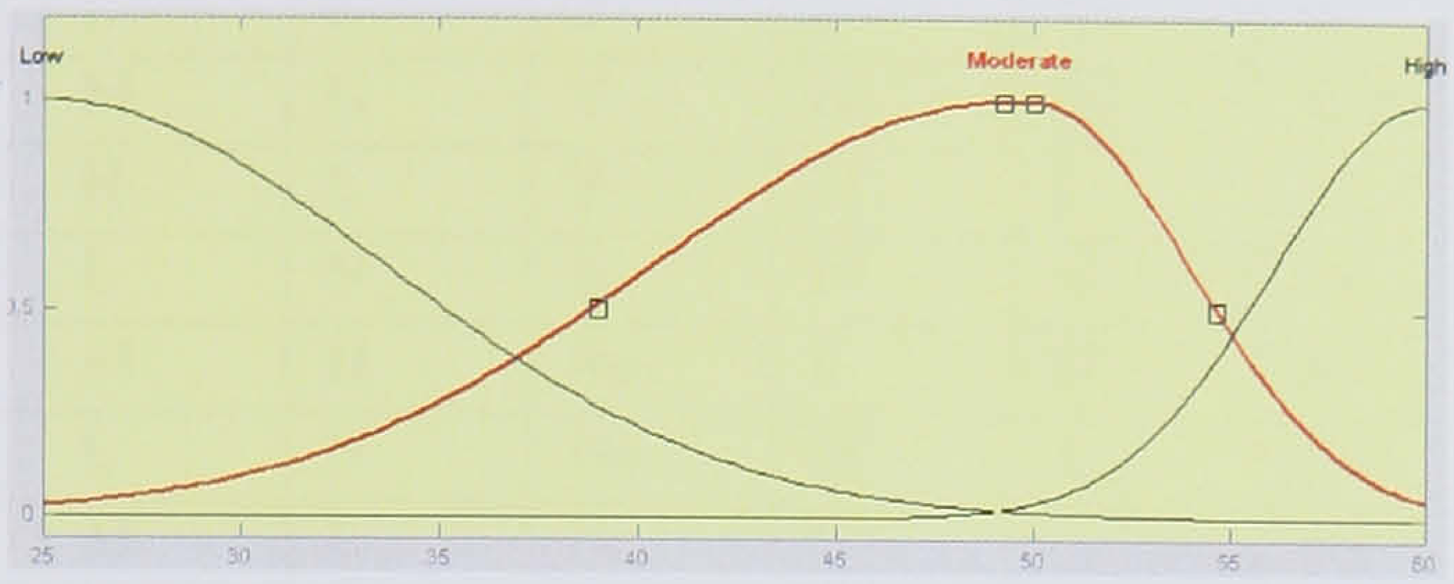
Figure D-7 Membership Functions of Travel Cards and Smart Cards in SP-3



Input MF: Travel cost of TC

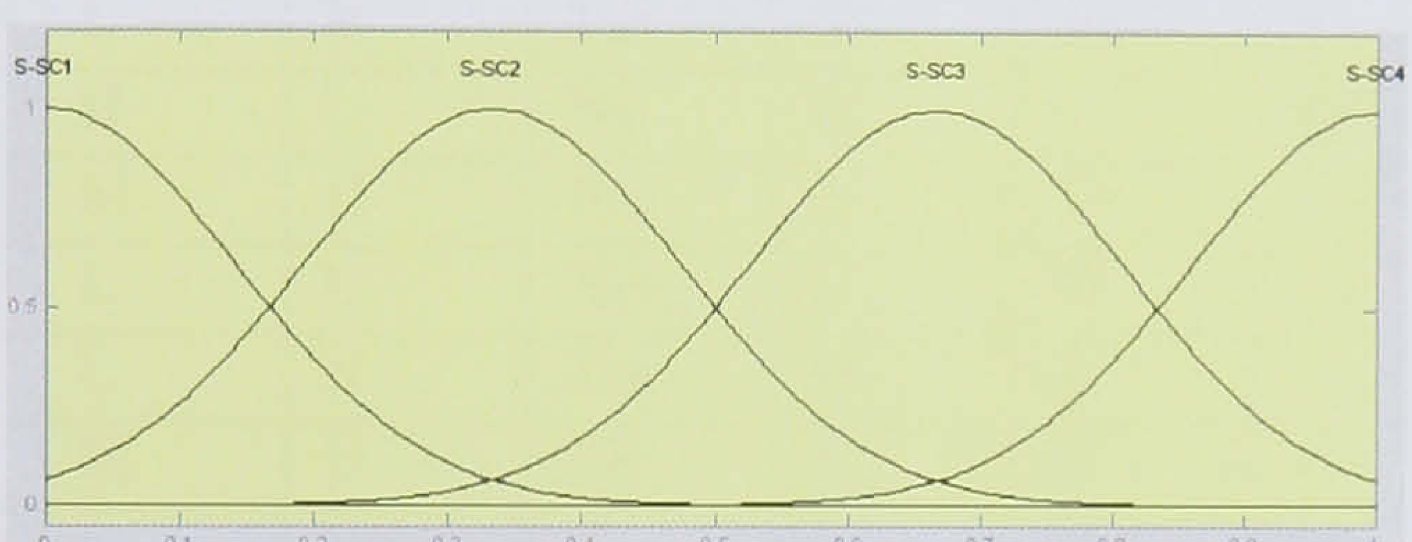


Input MF: Service of TC

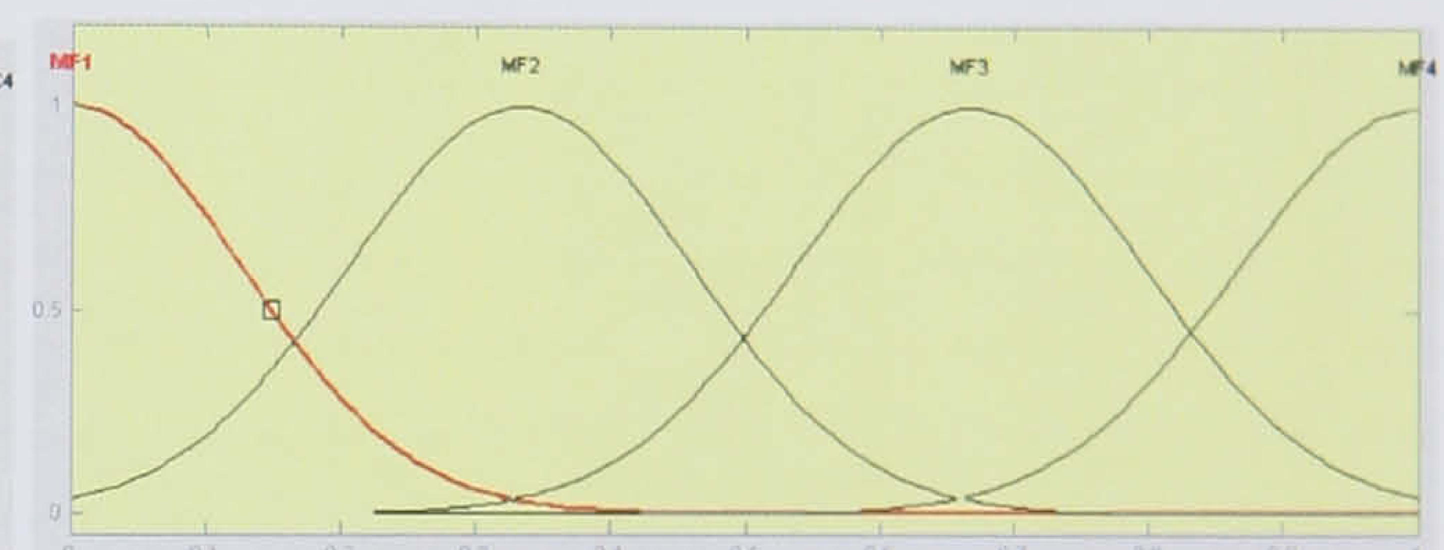


Input MF: Travel cost of SC

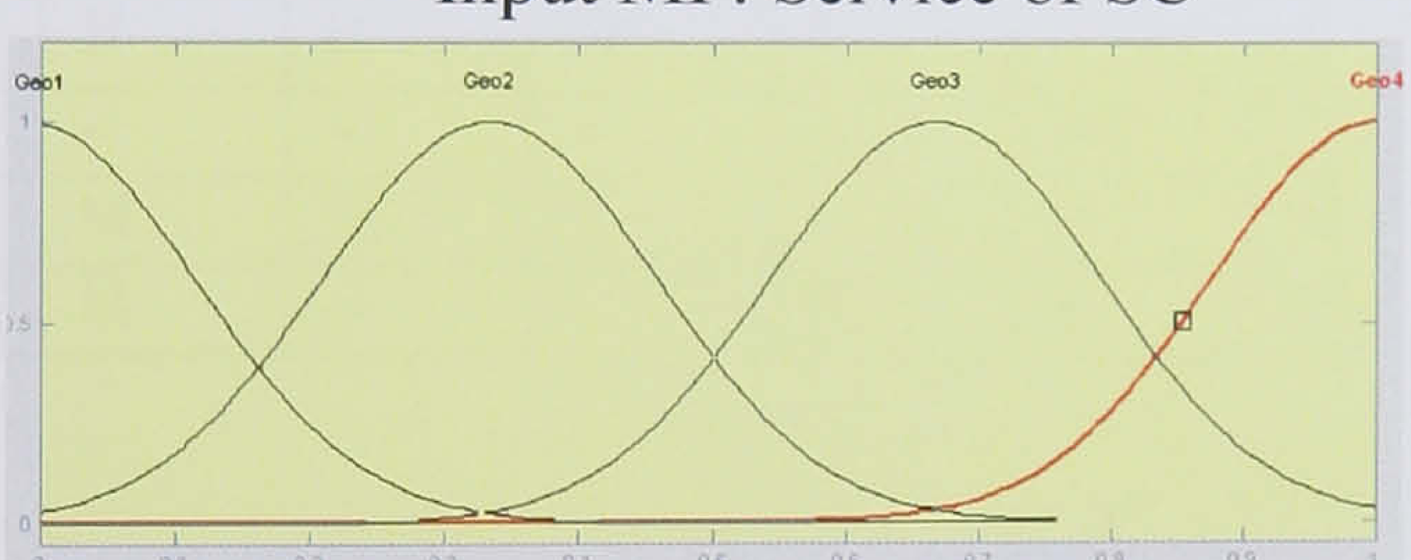
Input MF: Deposit of SC



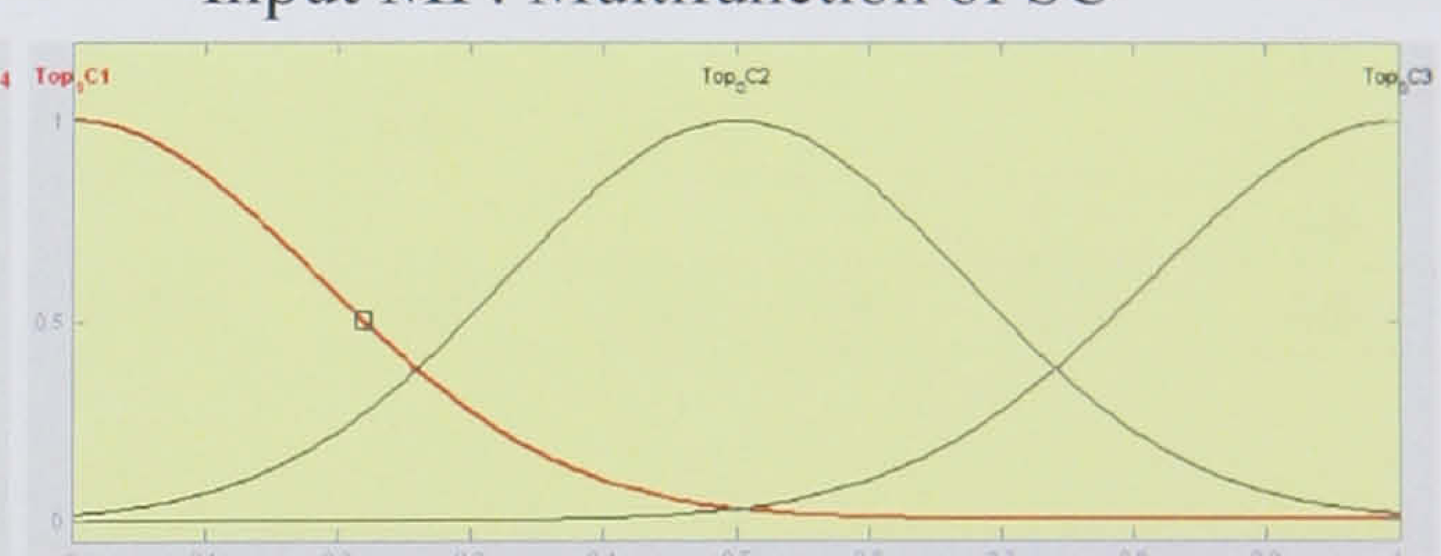
Input MF: Service of SC



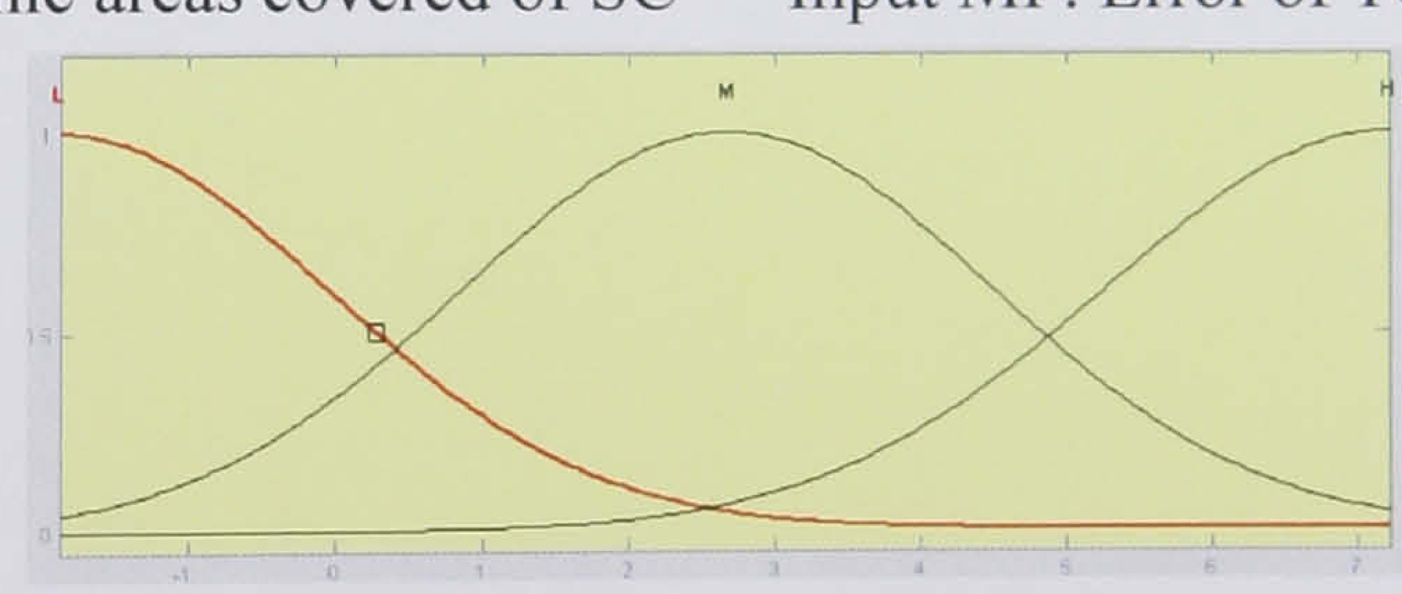
Input MF: Multifunction of SC



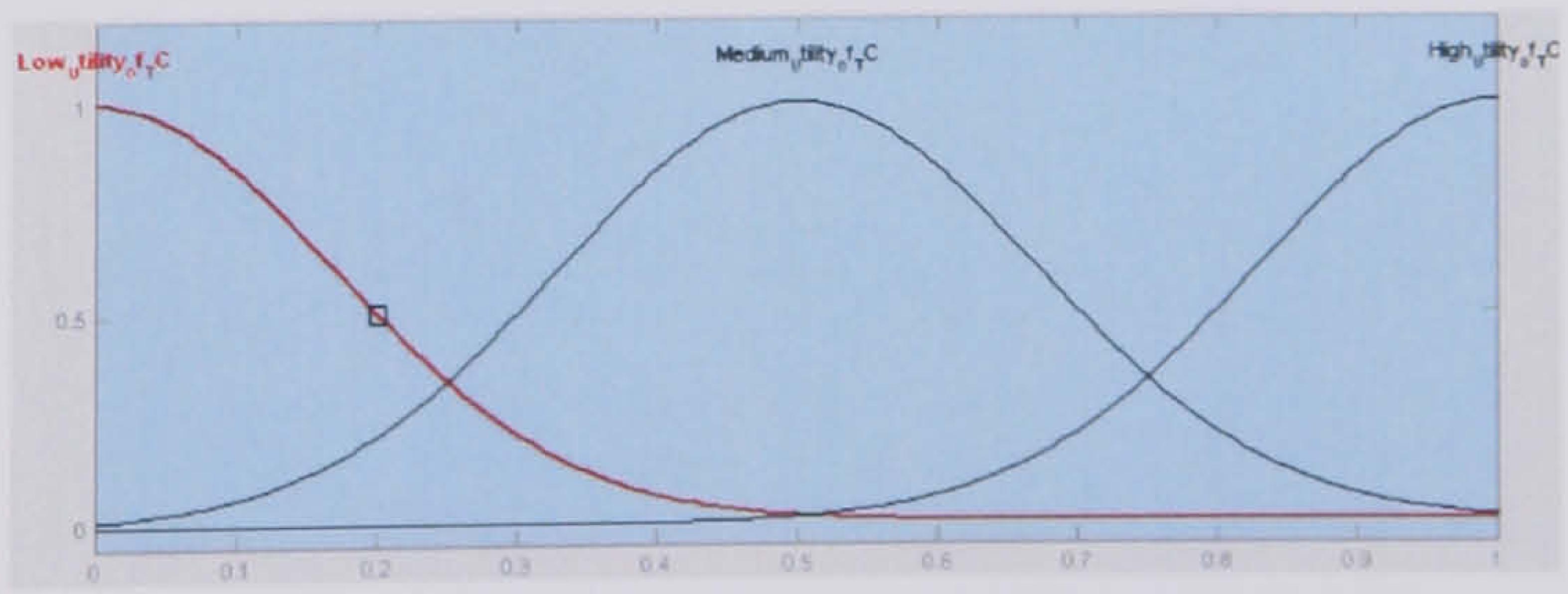
Input MF: Geographic areas covered of SC



Input MF: Error of TC (Gumbel distribution)



Input MF: Error of SC (Gumbel distribution)



Output MF (choice probability of Travel Cards)

Figure D-8 Membership Functions of Travel Cards and Smart Cards in SP-4

Appendix E: Fuzzy Rules (SP2, 3 and 4)

Table E-1 Fuzzy Rules in the SP-2

	IF											THEN
	Travel cost_C	Boarding time_C	Changes C	Travel cost_SC	Deposit SC	Overdraft SC	Multifunction SC	Geo_SC	Top-up SC	Error C	Error SC	Probability choosing cash
1	L	L	Y	L	Nil	No	--	--	--	--	L	L
2	M	N	Y	H	H	Yes	Good	--	--	M	L	M
3	H	L	Y	M	L	No	Better	--	--	M	M	L
4	L	N	Y	H	H	No	Bad	--	--	L	H	H
5	M	H	No	L	H	Yes	N	--	--	H	L	L
6	L	N	No	M	M	No	--	N	--	M	M	H
7	M	L	No	L	L	Yes	--	Better	--	M	L	L
8	H	L	Yes	H	H	No	--	Good	--	L	L	M
9	M	N	No	H	Nil	Yes	--	Good	--	L	M	H
10	L	N	Yes	H	Nil	Yes	--	Bad	--	L	H	H
11	H	L	No	M	L	Yes	--	--	N	M	M	L
12	M	H	No	H	M	Yes	--	--	Bad	M	L	M
13	L	L	Yes	L	M	No	--	--	Better	L	L	M
14	L	N	Yes	M	H	No	--	--	Good	M	H	H
15	M	H	No	L	H	Yes	--	--	Better	H	M	L
16	L	L	Y	H	H	No	Bad	Bad	Bad	L	--	H
17	L	L	Y	H	M	Yes	N	N	N	L	L	M
18	H	H	No	--	--	--	--	--	--	--	--	L
19	--	--	--	--	Nil	Yes	Better	Better	Better	--	--	L
20	M	M	Y	M	H	No	---	--	--	--	--	M
21	M	--	--	H	H	No	Bad	Bad	Bad	M	L	M

Table E-2 Fuzzy Rules in the SP-3

	IF										THEN
	Travelcost_TC	Service_TC	Travel cost SC	Deposit_SC	Overdraft_SC	MF SC	Top-up TC	Top-up SC	Error TC	Error SC	Prob. choosing TC
1	L	N	L	L	N	Bad	N	N	L	L	M
2	M	Better	H	H	N	N	Good	N	L	H	H
3	H	Good	H	Nil	Y	Good	N	Good	M	L	L
4	L	Bad	M	H	Y	Better	Bad	Good	M	M	M
5	H	N	L	M	N	Bad	Good	N	L	M	M
6	M	Better	H	M	Y	Bad	N	N	L	H	H
7	M	Good	M	H	Y	Good	Bad	Bad	M	L	H
8	L	Good	M	L	N	N	Good	N	H	H	H
9	H	Better	H	L	Y	N	Bad	Good	H	L	L
10	M	Bad	L	M	N	Good	N	Good	M	M	M
11	M	Good	L	Nil	Y	Better	N	Good	M	L	L
12	H	Bad	L	--	--	--	--	--	--	--	L
13	L	Better	M	--	--	--	--	--	--	--	H
14	L	--	H	--	--	--	--	--	--	--	H
15	M	--	L	--	--	Better	Bad	Better	--	--	L
16	M	Better	M	--	--	--	--	--	--	--	M
17	L	Good	L	--	--	--	--	--	--	--	M
18	H	Better	H	H	--	--	Good	--	--	--	M

Table E-3 Fuzzy Rules in the SP-4

	IF									THEN
	Travel cost TC	Service TC	Travel cost SC	Deposit_SC	Service_SC	MF SC	Geo_SC	Error TC	Error SC	Prob. choosing TC
1	L	Bad	M	N	Better	--	--	L	M	L
2	L	Better	M	H	Better	Good	Good	M	M	H
3	M	Better	L	H	Bad	Better	Bad	L	M	L
4	L	N	M	N	N	N	N	H	L	M
5	M	Better	M	L	Bad	Bad	Bad	L	H	H
6	H	Better	H	H	Good	Good	Good	M	M	M
7	H	Good	M	L	Better	--	--	M	L	L
8	H	--	--	N	Better	Better	Better	--	--	L
9	M	Bad	L	L	--	--	--	--	--	L
10	L	Better	H	--	--	--	--	--	--	H
11	--	--	M	H	Bad	Bad	Bad	--	--	H
12	M	N	M	L	--	--	--	--	--	M
13	L	Good	L	H	--	--	--	--	--	M
14	M	Better	H	N	Bad	--	--	--	--	M