A Comprehensive Analysis of Discrete Choice Modelling Specifications for Modelling Route and Stop Choice Behaviour of Transit Users

Author:
Hassan, Mohammad

Publication Date:
2019

DOI:
https://doi.org/10.26190/unsworks/21620

A Comprehensive Analysis of Discrete Choice Modelling Specifications for Modelling Route and Stop Choice Behaviour of Transit Users

By

Mohammad Nurul Hassan

MSc (Civil Engineering), BURP

A thesis presented in the fulfilment of the requirements for the degree of Doctor of Philosophy

School of Civil and Environmental Engineering

Faculty of Engineering

The University of New South Wales

December 2019

Read in the name of your Lord who created.
He created man from a clot.
Read and your Lord is the most honourable.
Who taught by the pen.
Taught man what he knew not.

(Qur'an Chapter 96 verse 1-5)

UNSW SYDNEY | Australia's Global University

Thesis/Dissertation Sheet

Surname/Family Name	:	HASSAN
Given Name/s	:	MOHAMMAD NURUL
Abbreviation for degree as give in the University calendar	:	PhD
Faculty	:	Faculty of Engineering
School	:	School of Civil and Environmental Engineering
Thesis Title	:	A Comprehensive Analysis of Discrete Choice Modelling Specifications for Modelling Route and Stop Choice Behaviour of Transit Users

Abstract 350 words maximum: (PLEASE TYPE)

Public transit demand modelling typically considers two main aspects: mode choice and transit assignment. In such frameworks of modelling transit users' behaviour, three important decisions are considered: choice of mode, choice of access to a stop and choice of a route. The literature argues that transit users may evaluate different strategies while choosing transit options such as minimizing travel time, travel cost or a number of transfers. Furthermore, there are other important aspects regarding the behaviour of users, like how people make decisions such as random utility maximisation (RUM) or random regret minimisation (RRM), and the influence of users' socio-economic and demographic attributes and the trip attributes on these decisions. This study addresses all of these aspects in a number of discrete choice modelling formulations to evaluate the appropriateness of these specifications for access stop choice and route choice modelling.

It was found that models based on the RUM theory generally fit the data of this thesis better. However, a hybrid RRM-RUM specification showed better prediction capability. More specifically, in the route choice models of this study, models with smaller choice set sizes work better than models based on larger choice sets. However, the prediction capability is better in the models with larger choice sets. Another finding of this thesis is that accounting for different strategies in modelling the behaviour of transit users is quite important. This thesis also examined the effectiveness of the popular simple random sampling technique in different discrete choice formulations in the route choice modelling context and proposed approaches to improve the quality of the choice modelling specifications. The use of the proposed sampling mechanisms can be expanded to other contexts like destination choice, housing search, and route choice, where sampling of alternatives is inevitable. Finally, the findings of this thesis research can provide insightful modelling methods to be used by transport/transit planning professionals to improve the quality of their demand models.

31.08.18
Date

FOR OFFICE USE ONLY Date of completion of requirements for Award:

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed:

Date: 14/12/2019

COPYRIGHT STATEMENT

Signed:

Date: 14/12/2019

AUTHENTICITY STAEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.'

Signed:

Date: 14/12/2019

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The student contributed greater than 50% of the content in the publication and is the "primary author", ie. the student was responsible primarily for the planning, execution and preparation of the work for publication
- The student has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not.

☐ *This thesis contains no publications, either published or submitted for publication(if this box is checked, you may delete all the material on page 2)*

☒ *Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement (if this box is checked, you may delete all the material on page 2)*

☐ *This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below*

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Name	**Signature**	**Date (dd/mm/yy)**
Mohammad Nurul Hassan		**14/12/2019**

ABSTRACT

Public transit demand modelling typically considers two main aspects, namely: mode choice and transit assignment. In such a framework of modelling transit users' behaviour, three important decisions are considered: choice of mode, choice of access to a stop and choice of a route. The literature argues that transit users may evaluate different strategies such as minimising travel time, travel cost or number of transfers while choosing transit stop or route. Furthermore, there are other important aspects regarding the behaviour of users like how to model individual behaviour in making decisions based on the behavioural theories like random utility maximisation (RUM) or random regret minimisation (RRM), and the influence of users' socioeconomic, demographics, and trip attributes on these decisions. This study addresses all these aspects using different discrete choice modelling formulations to evaluate the appropriateness of these specifications for access stop choice and route choice modelling.

In this study, it is found that models based on the RUM theory generally show a better fit to the data which is used in this thesis. However, a hybrid RUM-RRM specification shows better prediction capability. Particularly for the transit route choice models in this study, models with smaller choice set sizes work better than the model based on larger choice sets. However, the prediction capability is better in models with larger choice sets. Another finding of this thesis is that accounting for different strategies in modelling the behaviour of transit users is quite important. This thesis also examines the effectiveness of the widely used simple random sampling technique in different discrete choice formulations in the route choice modelling context and proposes approaches to improve the quality of the choice modelling specifications. The use of the proposed sampling mechanisms can be expanded to other contexts like destination choice, housing search, and route choice, where sampling of alternatives is inevitable. Finally, the findings of this thesis research can provide insightful modelling methods to be used by transport/transit planning professionals to improve the quality of their demand models.

ACKNOWLEDGEMENT

At first, I would like to express my utmost gratitude to the Almighty for all the blessings, including the opportunity to take this thesis to this final stage.

I want to thank the Commonwealth Government's contribution to this research by providing me the "Australian Government Research Training Program Scholarship."

I want to mention the people whose support, encouragement and blessings made it possible for me to finalise the thesis. Firstly, I would like to thank my supervisors, Professor Travis Waller and Associate Professor Taha Hossein Rashidi, for their incredible support and invaluable suggestions throughout the PhD candidature. I am grateful to Taha for his excellent supervision and guidance and Travis for his encouragement and thoughtful comments. I want to thank Professor Mark Hickman from UQ and Dr Neema Nassir from Melbourne University for the transit data and TBSP codes which were used in this research. Thanks to all the members of the rCITI family, including Sai, Ted, Raed, Ali, Abdul Majid, Ahmed, Asif, Xiang, Shantanu, Alireza, Divya, Kiran and Maria. A very special thanks to Neeraj for his extraordinary support in every aspect during my candidature. Special gratitude to Professor Yaser Hawas from UAEU for his encouragement to pursue the PhD. Special thanks to my niece Shayla for her help in coding. Thanks to Mandy Lisson of "The Expert Editor" for the editorial service.

I want to thank two anonymous reviewers for their thoughtful and constructive comments which help me to improve the thesis.

I am grateful to my parents and parents-in-law for their never-ending love and blessings. Thanks to all my extended family members and friends for their encouragement and support. Special thanks to my friend Zobair for his assistance towards "the straight path" which made me calm and focussed during difficult circumstances.

Finally, heartiest gratitude to my wife Rowshan and my sons Rameen and Rishan, who were the sufferers of this long and challenging journey. Without their love, understanding and patience I might not have been able to write this acknowledgement.

LIST OF RELEVANT PUBLICATIONS

The list of journal publications and conference proceedings and under review journal publications that have contributed towards the development of this thesis are as follows:

Publication in Journals

- **Hassan M. N**., Rashidi T. H., Waller S. T., Nassir N. and Hickman M. Modelling Transit Users Stop Choice Behaviour: Do Traveler Strategize? *Journal of Public Transportation* 19 (3) 98-116, 2016
- **Hassan M. N.,** Najmi A. and Rashidi T. H., A Two-Stage Recreational Destination Choice Study Incorporating Fuzzy Logic in Discrete Choice Modelling. *Transportation Research Part F* 67 (123-141), 2019.
- **Hassan M. N.**, Rashidi T. H. and Nassir N., Consideration of Different Travel Strategies and Choice Set Sizes in Transit Path Choice Modelling. *Accepted* in "Transportation".

Publications in Conference Proceedings

- **Hassan M. N**., Rashidi T. H. and Nassir N., (2018). Developing A Fuzzy Logic Based Sampling Protocol For Transit Route Choice Modelling. *Transportation Research Board 97th Annual Meeting,* January 2018, Washington DC, USA.
- **Hassan M. N.**, Rashidi T. H., Waller S. T., Nassir N. and Hickman M., 2016. Do Travellers Strategize When Selecting Their Transit Stop?, *Transportation Research Board 95th Annual Meeting*, January 2016, Washington DC, USA
- **Hassan M. N.,** Rashidi T. H., Waller S. T., Nassir N. and Hickman M., (2015). Study of Users Travel Approaches in Transit Stop Choice Modelling, presented in*14th International Conference on Travel Behaviour Research*, 19-23 July 2015, Windsor, UK.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1 INTRODUCTION

1.1 Introduction and Rationale of the Research

Public transportation demand estimation is essential for urban planners, transport authorities, and transit agencies. These models help to improve efficiency and cost-effectiveness in the planning of transit supply. A review of the literature identifies three key issues (presented in Figure 1.1.) of transit demand estimation. These are: what type of model, how do people make choices, and who chooses what?

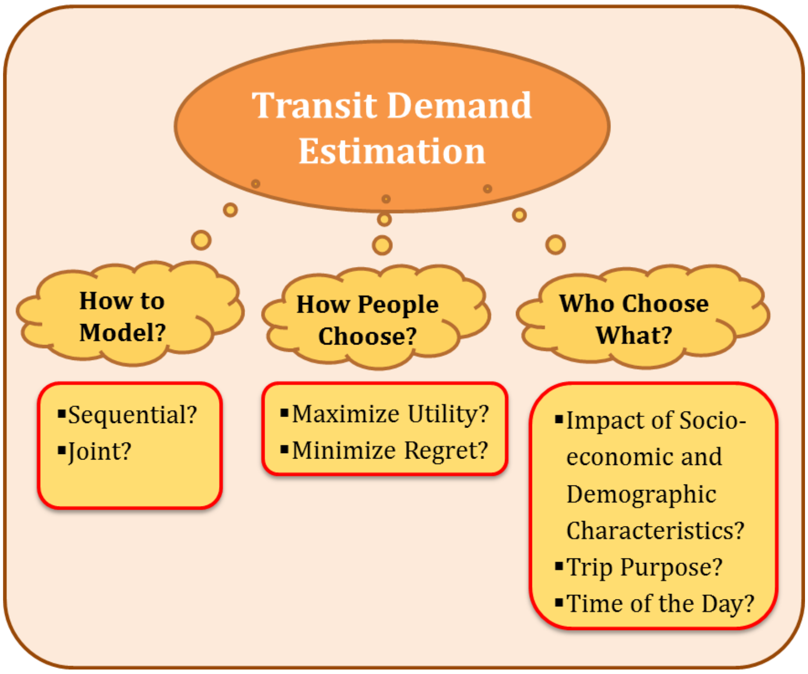

Figure 1.1 Transit Demand Estimation issues

1.1.1 The "Choices" of Transit Research

Usually, transit demands are estimated in the mode choice and transit assignment steps of travel demand modelling. Researchers have typically concentrated on developing mode choice models and transit assignment models separately. In the conventional travel demand modelling structure, transit assignment comes after the transit mode choice step, which needs re-evaluation, as transit users might choose travel routes while evaluating transit modes. Again, a transit route choice model is needed for the transit assignment step. Recently, researchers have begun studying transit stop choice behaviour when focusing on modelling transit mode and route choice decisions. As a result, the question of "how do you model these choices?" has become an important issue to investigate in transit demand modelling.

Transit mode choice models are mainly studied in generic mode choice modelling (Bhatta and Larsen, 2011; Habib, 2012; Habib and Sasic, 2014; Hess et al., 2007). Habib (2011) investigated mode choice behaviour considering work start time and work duration. Habib and Sasic (2014) developed a model to explore mode choice behaviour of non-work travel at peak travel periods. Hess et al. (2007) examined the choice of departure time and mode from a stated preference survey conducted in two different countries (the UK and the Netherlands). Eluru et al. (2012) developed separate mode and route choice models to understand the behaviour of McGill University commuters. Besides these, some researchers have addressed the transit mode choice issue separately as the behaviour of transit users were found to be different from auto users. Beimborn et al. (2003) studied transit mode choice from a captivity perspective, finding that the traditional models could not estimate transit share behaviour effectively, as they underestimated the share of captive users while they overestimated the share of choice users.

Route choice models in the literature generally consider different route attributes, especially the components of travel time including access time, waiting time, in-vehicle travel time, walking time, transfer time, and egress time (Van der Waard, 1988; Bovy and Hoogendoorn-Lancer, 2005; Raveau et al., 2011; Eluru et al., 2012; Tan et al., 2016; Anderson et al., 2017). Some studies investigated the transfer component in detail (Anderson et al., 2017; Garcia-Martinez et al., 2018; Schakenbos et al., 2016; Vrtic and Axhausen, 2003). Some relate route choice modelling with the trip purpose (Vrtic and Axhausen, 2003; Anderson et al., 2017). Studies have also incorporated taste heterogeneity (Eluru et al., 2012; Tan et al., 2016; Anderson et al., 2017; Nielsen et al., 2000; Fosgerau et al., 2007a, b) and route overlap (Tan et al., 2016; Anderson et al., 2017) in the modelling

specification. Although mode characteristics are integrated into many of these studies, none of the studies considers the stop attributes in the modelling structure.
Transit stop choice models are a relatively new idea which assumes that transit users choose a transit mode by selecting an access stop (Chakour and Eluru, 2013; Debrezion et al., 2009; Mahmoud et al., 2014; Nassir et al., 2015b). Here, the idea of choosing an access stop consequently identifies the selected mode and shows that the selection of a mode can be an auxiliary choice if the primary choice is the access stop. Nassir et al. (2015b) incorporated route attributes at the stop level by considering aggregate characteristics of the bundle of routes initiated from a stop.

The critical question is, how should these choices be modelled: individually, sequentially or jointly? Though most of the researchers used an individual approach, the joint and sequential approach to include all these three choice components of transit could also be investigated.

1.1.2 Selection of Choice

The second issue of demand estimation is to model the users' choice selection processes. Most researchers perceive choice selection as selecting the alternative that yields the maximum utility (Random Utility Maximization or RUM). These researchers used this concept in most of their transit choice models (mode, stop and route) and consequently established this as a classical discrete choice modelling approach. Along with other researchers, Wen et al. (2012), Hess et al. (2013), and Zhou and Lu (2011) used the RUM concept in mode choice. Lam et al. (1999), Yang and Lam(2006), and Nielsen and Frederiksen (2006) used RUM in route choice problems while Nassir et al. (2015b) and Debrezion et al. (2009) used it in stop choice models.

Apart from RUM, Random Regret Minimization (RRM) theory was adopted by Chorus et al. (2010; 2008), who argued that users might avoid negative emotions rather than maximising certain benefits. This might occur when the alternatives are not fully compensatory, especially in the case of risky travel choice. RRM might then be a good choice where users are not satisfied with the existing transit system. Consequently, both of these theories could be explored in transit research.

1.1.3 Users' Choice Behaviour

Choosing from the alternatives depends not only on the attributes of the alternatives but also on the characteristics of the transit users who make the decision. The majority of researchers have proposed models which incorporate different network attributes of the alternatives. However, the impact of users' socioeconomic and demographic characteristics has not been adequately studied. Therefore, "who chooses what?" is an important question to address in transit demand research.

Users' choice behaviour is very challenging but essential to model as it is used to forecast travellers' behaviour, predict future traffic conditions and understand travellers' reactions and adaptation to sources of information (Prato, 2009). Some researchers have modelled users' choice behaviour through a learning process (Nuzzolo et al., 2001; Wahba and Shalaby, 2009). One of these researchers suggested a utility-based model which accounts for both real-time information and the experience of network attributes through a day to day learning process using an exponential filter (Nuzzolo et al., 2001). Another study proposed a Micro-Simulation Learning-Based Approach for Transit Assignment (MILATRAS), where an adaptive trip choice behaviour was assumed (Wahba and Shalaby, 2009). It also analysed the interaction between transit network performance and passengers' decisions. Another study used game theory to model the behaviour of transit users aiming to minimise their expected travel times (Schmöcker et al., 2013). Vij et al. (2013) classified users' behaviour according to their modality style and presented a latent class model to predict mode choice.

1.1.4 Strategy of Choice

The concept of travel strategy was first introduced by Nguyen and Pallottino (1988a) and Spiess and Florian (1989) in transit assignment research. Here, travel strategy referred to a set of coherent decision rules that allowed the passenger to travel from their origin to their destination, and an optimal strategy is to minimise the passenger's travel time. However, in this concept, the optimal strategy did not consider other aspects like minimising the number of transfers or minimising walking time. In reality, transit users may want to walk less or use more direct routes to avoid transfers. Again, different studies (Beimborn et al., 2003; Fonzone et al., 2010; Fonzone et al., 2013; Kurauchi et al., 2014; Nassir et al., 2015b) support this assumption.

1.1.5 Problems of Transit Choice Modelling

In the literature, two major challenges were identified for transit route choice modelling: generation of a choice set and estimation of discrete choice modelling (Prato 2009). The choice set generation problem has been studied for a long time, starting from the labelling approach (Ben-Akiva et al., 1984), to the random walk approach (Frejinger et al., 2009). However, researchers are still working on increasing the efficiency of methods to incorporate heterogeneous and realistic alternatives in the choice set as the composition of the choice set, affects the estimation (Nerella and Bhat, 2004; Prato and Bekhor, 2007; Bliemer and Bovy, 2008). The problem is critical for large and dense multi-modal transit networks which offer a large number of alternatives to passengers. This research aims to improve the choice set composition method to incorporate more likely alternatives in the choice set.

The model estimation is critical as the model specifications need to incorporate different aspects: the likely variables that are considered by the users, taste and preference heterogeneity, and the correlation for the commonality of the route. The modelling process needs to incorporate transit choice components (mode, stop, route) and the selection of the strategy. Furthermore, the appropriate modelling approach should be identified to reflect users' behaviour. Therefore, this thesis will also develop appropriate modelling specifications to address these issues.

1.2 Aim and Objective of the Research

This research aims to develop an appropriate modelling structure that incorporates the three choices of mode, stop and route to estimate transit demand. The objectives include:

1. Investigate the potential modelling structure to model the choices
2. Understand travel behaviour and selection preferences
3. Identify possible issues related to transit demand estimation

1.3 Scope of the Research

The scope of the research is limited to the following:

1. Use a state-of-art choice set generation algorithm to generate initial choice sets;
2. Investigate existing sampling mechanisms and suggest an appropriate sampling mechanism to form the final choice sets;
3. Use different discrete choice model structures to model the transit stop and transit route choice behaviour;
4. Investigate the influence of different strategies and passengers' socio-demographics on transit stop and route choice;
5. Investigate the effect of choice set size in route choice modelling;
6. Study the effects of sampling methods on the choice set formation and overall model performance.

1.4 Organisation of the Thesis

This thesis presents a comprehensive analysis of discrete choice modelling specifications for modelling the route and stop choice behaviour of transit users. This thesis is organised into seven chapters. A synopsis of these chapters, particularly how the chapters are connected with the transit demand estimation problem and the objectives of the study, is provided in Figure 1.2. A brief description of the chapters is given below.

1.4.1 Chapter 1: Introduction

The Introduction describes the background, rationale, aim and objectives of the research. This chapter also discusses the relationship of the thesis with the problem statement and the objective of the research.

1.4.2 Chapter 2: Methodology

The second chapter discusses the overall methodology of the study. A four-stage procedure adopted for the study is discussed in this chapter as is RUM modelling structures tested in the study. Furthermore, this chapter proposes a new sampling technique adapted for the route choice modelling in Chapter 4 and the testing of the effects of the sampling in Chapter 7.

1.4.3 Chapter 3: Modelling Transit Users Access Stop Choice Behaviour

The third chapter formulates the access stop choice model. In this chapter, four types of RUM modelling structures are considered. All the modelling structures consider different strategies, modes and route attributes along with stop attributes. A comparison of different structures is provided along with the behavioural interpretation.

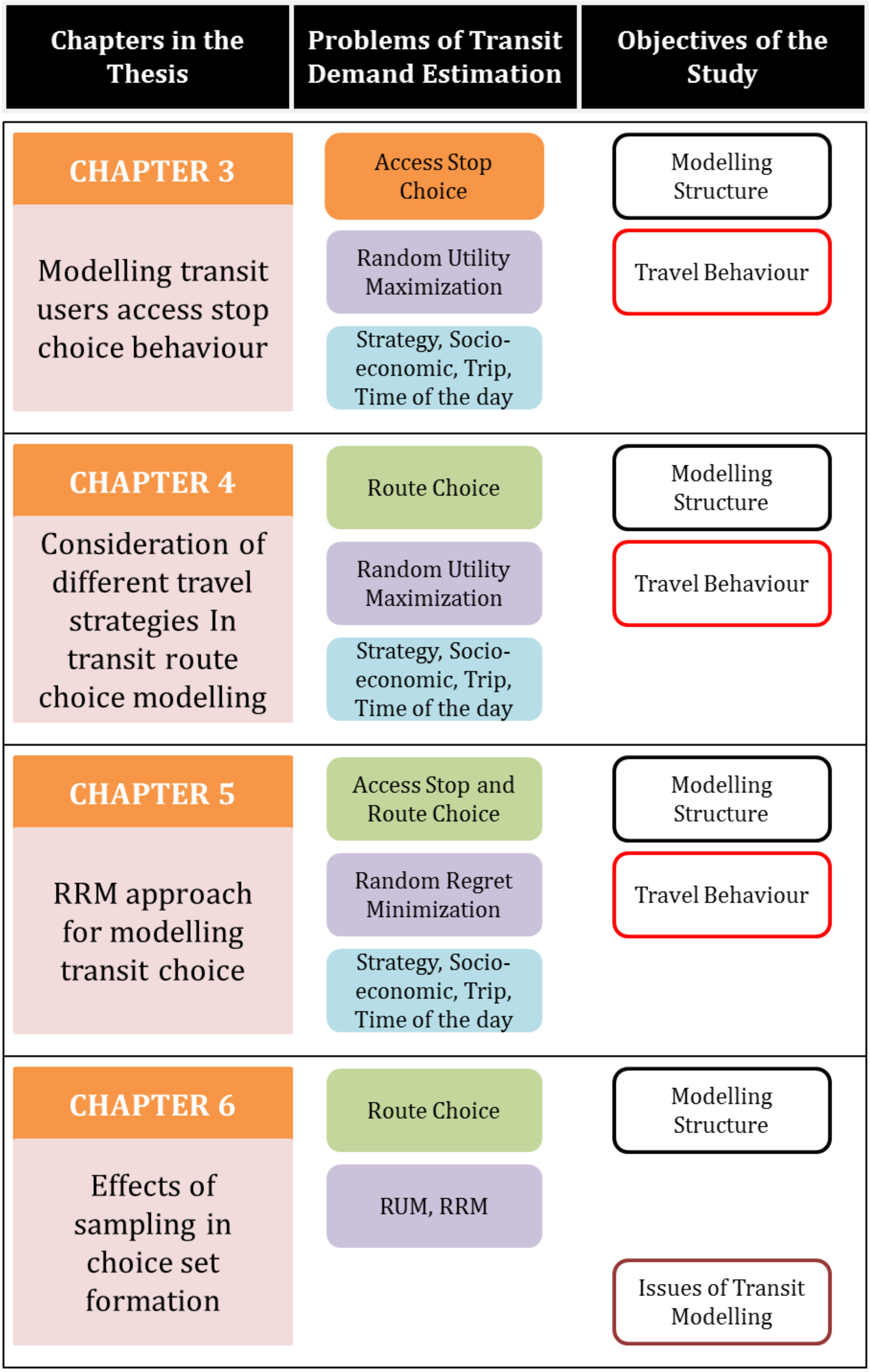

Figure 1.2 Thesis Relationships with the Problem and the Objective

1.4.4 Chapter 4: Consideration of Different Travel Strategies in Transit Route Choice Modelling

Chapter 4 discusses the formation of route choice modelling. In this chapter, two types of RUM modelling structures are considered. These modelling structures consider the sampling of alternatives as the number of route choice options is generally very high. The route attributes are considered along with the strategy, mode and stop attributes. Finally, a comparison of different structures is provided along with the behavioural interpretation.

1.4.5 Chapter 5: RRM Approach for Modelling Transit Stop and Route Choice

In Chapter 5, the RRM approach for stop and route choice modelling is discussed with the evaluation and formation of the RRM methods. This chapter compares RUM and RRM approaches and discusses the findings.

1.4.6 Chapter 6: Effects of Sampling in Choice Set Formation

The sixth chapter discusses the issue of sampling alternatives in the choice set formation. This issue was identified in this study while modelling route choice. Different sampling mechanisms which are commonly used are discussed. Experiments performed to assess the functionality and prediction ability of different sampling methods and protocols are also evaluated.

1.4.7 Chapter 7: Conclusions and Future Research Potentials

The seventh and final chapter makes concluding remarks and discusses potential future research directions.

2 METHODOLOGY

This chapter includes the detailed methodology of this study, divided into four segments, as shown in Figure 2.1. The first segment includes acquisition and cleaning of the data. The second segment generates the route and stop choice set from the data prepared in the first segment. The third segment involves a series of discrete choice modelling exercises for stop and route choice. The fourth segment compiles the findings of the modelling exercises and recommends an appropriate modelling structure for transit choice modelling. The flow and structure of these segments are illustrated below.

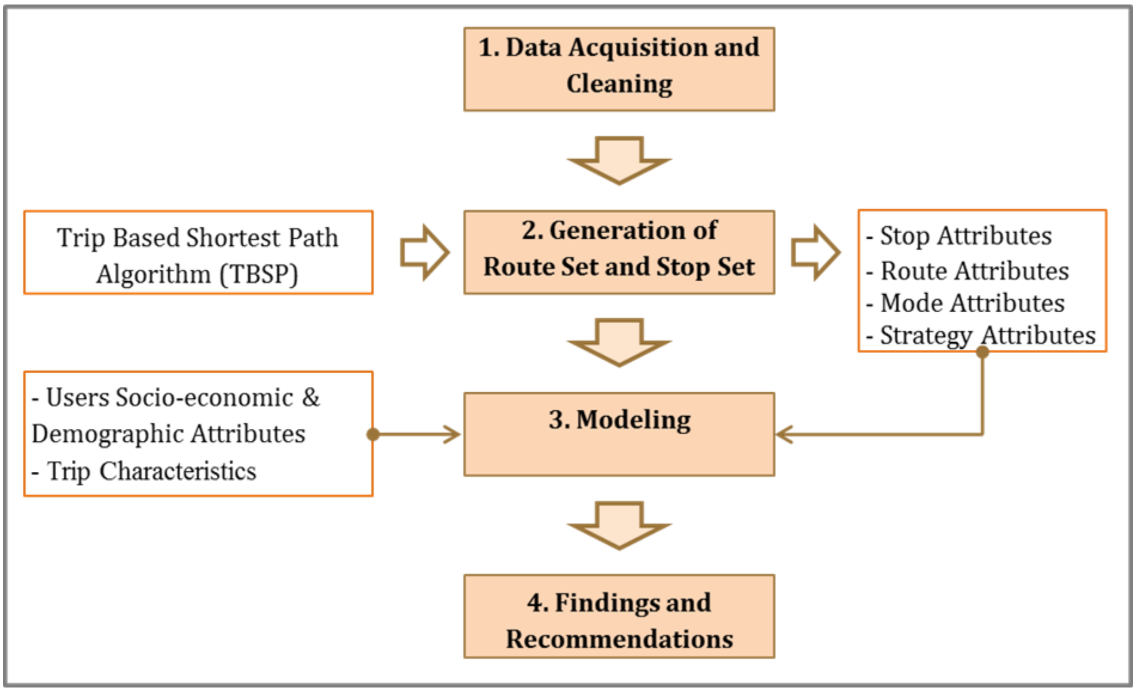

Figure 2.1 Structure of the Study

2.1 Data Acquisition and Cleaning

The proposed methods and techniques of this research are tested in the context of Southeast Queensland, Australia. The study area includes the Brisbane Statistical Division, the Gold Coast City Council, and the Sunshine Coast Regional Council areas. All data were first acquired for a research project at the University of

Queensland (Malekzadeh, 2015; Nassir et al., 2015b), and were later used in this research.

2.1.1 Household Travel Survey (HTS)

The trip dataset used in this research was taken from the household travel survey (HTS) of May 2009 conducted in Southeast Queensland (SEQ), Australia. This survey includes all occupied private residential households within the study area defined by the Brisbane Statistical Division, the Gold Coast LGA (Local Government Area) and the Sunshine Coast within the LGAs of Noosa, Maroochy and Caloundra (see Figure 2.2). The survey included the travel undertaken by persons aged five and above on all days of the week during the survey period. The survey also included all individuals (including visitors) staying at these households on the night before the household's travel day.

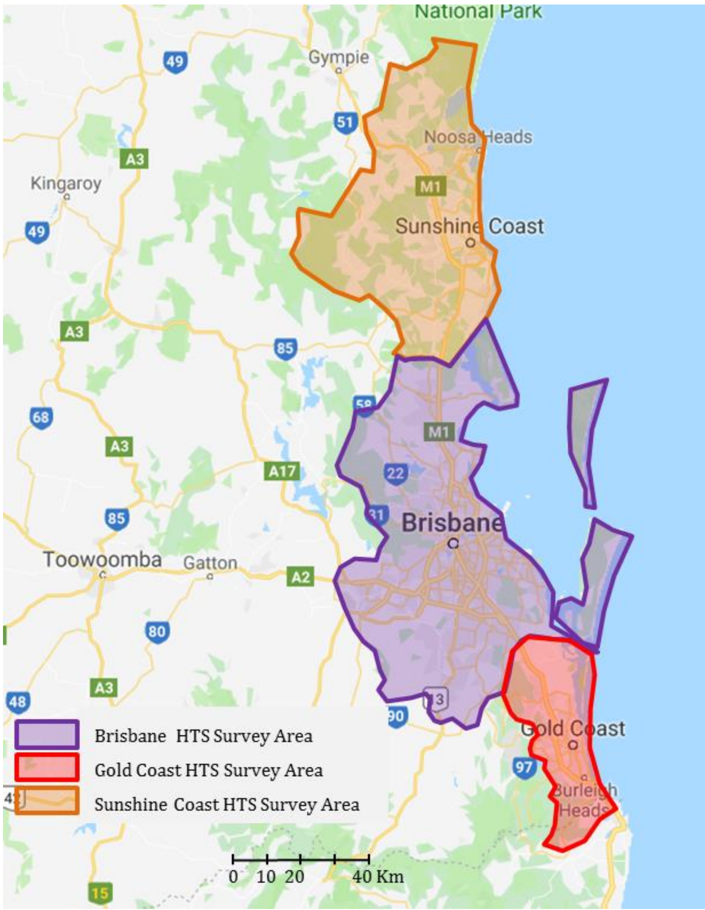

Figure 2.2 HTS 2009 Study Area (DTMR, 2010a, b, c)

All travel records (1,693 journeys) using public transport (which includes three modes: bus, train, and ferry), with walking legs of access, egress, and transfer(s), were extracted from the HTS data. These 1,693 journeys included 1,435 transit trips with no transfers, 229 trips with a single transfer, 26 trips with two transfers and three trips with three transfers. Transit mode choice of these records shows that bus is the most common mean of transit for the access stop (1,176 trips) followed by the train station (292 trips) and ferry terminal (24).

Information regarding the socio-demographic characteristics of the respondents (e.g. age, gender, income, employment status) and the household (e.g. number of members in the household, dwelling type, number of household vehicles) was gathered from the HTS-2009 dataset (DTMR, 2010a, b, c). Again, the information regarding the trip (e.g. origin, trip destination, the purpose of travel, mode of travel, time of travel) was utilised in the study. A summary of the socio-demographic and trip characteristics of the respondents (used in this study) is provided in Table 2.1.

2.1.2 Public Transit Network Data

The transit network data for this study was acquired from General Transit Feed Specification or GTFS. GTFS is a service developed by Google which enables transit authorities to publish their transit schedule and real-time service data. This service allows users to plan their trips from an origin to a destination using public transport in Google Maps (www.google.com.au/maps) and also provides an opportunity for agencies and researchers to evaluate public transport performance. GTFS has six required fields of information: agency, stops, routes, trips, stop times and calendar (Malekzadeh, 2015). Besides these, some optional information such as calendar dates, fare attributes, fare rules, shapes, frequencies, transfers and other information can also be used in GTFS (Malekzadeh, 2015). The GTFS data revealed that the SEQ transit network included 14,441 stops, 767 routes, and 33,897 timetabled vehicle trips, with different services on weekdays and weekends.

Table 2.1 Summary of HTS-2009 Socio-demographic and Trip Data

Traveller's Information		
1	Age	Mean age 35.55, standard deviation 19.2
2	Gender	Male 44%, Female 46%, Other/No Response 10%
3	Work	Full-time 37%,
4	Country of Birth	Australia 72%, Other Country 28%
5	Car License	Full Licence 48%, Probationary Licence 3%, Learners Permit 10%, No Car Licence 39%
6	Vehicle Ownership	Owned 57%, Not owned 43%
Traveller's Household Information		
7	HH Structure	Sole Person 13%, Couple No Kids 20%, Couple With Kids 36%, One-Parent 8%, Other HH Structures 23%
8	HH Income	Less than $399 per week 51%, $400-1300 per week 38%, More than $1300 per week 11%
9	Property Ownership	Owned 57%, Rent/other 43%
10	Dwelling Type	Separate House 80%, Flat/Apartment 15%, Townhouse 5%
11	Number of Bedrooms	Mean: 3.13, standard deviation: 0.98
12	Total Vehicles in the HH	Mean: 1.37, standard deviation: 0.99
Trip Information		
13	Purpose of travel	Work 67%, other 33%
14	Travel Day	Weekday 90%, Weekend 10%
15	Departure Time	AM Peak 28%, PM Peak 17% Off-Peak 55%
16	Access Stop	Bus 79%, Train 19%, Ferry 2%
17	Travel Mode Combination	Walk-Transit-Walk 84%, Walk-Transit-Transit-Walk 10%, Walk-Transit-Walk-Transit-Walk 4%, Other 2%

Walking network data is also used in this study to calculate walking distances/times. The walkway network data, consisting of local streets, sidewalks, crosswalk connections, walking ramps, footways, and stairways for SEQ, was obtained from "OpenStreetMap" (http://www.openstreetmap.org/). This walk network included about 250,000 nodes and 340,000 links. This network is used to calculate the walking distance from an origin location to the access stop, intermediate walking for transfers and from the egress stop to the destination. These road and walkway networks were developed by Malekzadeh (2015).

Besides these, the Queensland Department of Transport and Main Roads (DTMR) provided another dataset containing information about stop facilities such as shelter, stop lighting, street lighting, access walkways, boarding slabs, and the number of maps. A summary of this dataset is provided in Table 2.2.

Table 2.2 Summary of DTMR Stop Information

Sl.	Attribute	Description
1	Sheltered Stop	Total 3821 (26.5%)
2	Stops with Street Lighting	Total 1442 (10%)
3	Stops with Lighting	Total 3133 (21.7%)
4	Stops with Boarding Slab	Total 8351 (57.8%)
5	Stop with Footpath	Total 8062 (55.8%)
6	Stops with Map/Schedule (A4)	Total 7240 (50.2%)
7	Stops with Map/Schedule (A3)	Total 1535 (10.6%)

2.2 Generation of Path Set and Stop Set

2.2.1 Review of Path Generation Models

The literature on path generation methods is limited for public transit networks or multimodal transport networks compared to the road network. However, most transit path generation approaches are derived from path generation methods for road networks with some modifications to incorporate the characteristics of transit operations (Tan, 2016). A multimodal/transit network is characterised by stops, transfers and schedules or frequency of service, which make the path choice set generation problem more challenging and distinct from the path generation of the road network.

2.2.1.1 Choice set generation for road network

Prato (2009) discussed choice set generation methods for auto in details. Prato (2009) classified the literature into four broad categories 1) deterministic shortest path based methods, 2) stochastic shortest path based methods, 3) constrained enumeration methods, and 4) Probabilistic methods. This subsection discusses these methods briefly.

Deterministic shortest path based methods

In the deterministic method, the choice sets are generated by repeated shortest path searches in the network. Here, the computation process modifies the link impedance, route constraints, and the search criteria to detect the optimal paths. The most clear-cut deterministic method is the *K*-shortest path algorithm where the best *K* paths are generated from different generalised cost functions (Bekhor et al., 2006; Van der Zijpp and Catalano, 2005). Although this method generates attractive and heterogeneous paths, certain limitations are pointed out by Prato (2009). Firstly, the generated choice set for an OD pair is the same for every traveller as the method does not consider personal preferences or constraints. Secondly, for prediction purposes, the number of generated paths relies entirely on the sensitivity and the experience of the researcher.

The link elimination approach introduced by Azevedo (1993) is based on a repetitive search of the shortest path after eliminating links (partly or wholly) from previous searches. Different variants of this approach were found in the literature (Bekhor et al., 2006; Frejinger and Bierlaire, 2007; Prato and Bekhor, 2006; Tan, 2016) where the link elimination decision was executed after different iterations. The original formulation (removes all the links from the shortest path) does not allow a more attractive path to be generated because of the network disconnection problem. Conversely, the other variants can generate attractive paths, but with a very high computational time.

The link penalty approach (Bekhor et al., 2006; De La Barra et al., 1993; Prato and Bekhor, 2006; Scott et al., 1997) is similar to the link elimination approach in terms of a repetitive search for the shortest path. However, this approach imposes a penalty on all the links of the shortest path rather than eliminating link. Therefore, the approach retains the essential links in the network. However, the critical task is to determine the value of the penalty factor. The limitations of this approach are the generation of high impedance paths, dependency on the definition of the penalty factor and computational inefficiency.

The labelling approach, introduced by Ben-Akiva et al. (1984) addresses the heterogeneity of travellers by defining different path cost functions called labels. These labels reflect different objectives that travellers can have, including minimising travel time, travelling through familiar landmarks or scenic roads, and avoiding congestion. Other researchers (Prato and Bekhor, 2006; Ramming, 2002) also used this approach by defining different attribute labels ranging between four and sixteen, respectively. Although the path generation time (computation time) of this approach is faster than other methods, the generated choice set does not necessarily represent real choices (Bekhor et al., 2006; Ramming, 2002). Again, a representative choice set mostly depends on the definition of the labels and the correct guess of the objective functions, which requires a priori knowledge about travellers' preference (Prato, 2009).

Stochastic Shortest Path Based Models

This method employs repeated shortest path searches based on random values of network impedances and individual preferences. These methods use simulation techniques to randomise network attributes through different types of probability distribution including the probit distribution (Sheffi and Powell, 1982), normal distribution (Ramming, 2002), truncated normal distribution (Bierlaire and Frejinger, 2005; Nielsen, 2000; Prato and Bekhor, 2006), and Kumaraswamy distribution (Frejinger et al., 2009). Nielsen (2000) introduces the concept of the doubly stochastic path generation method, where along with the different perception of traffic network attributes, he introduces the taste heterogeneity of the drivers. This approach is also applied by Bovy and Fiorenzo-Catalano (2007) to the Rotterdam-Dordrecht corridor in the Netherlands and by Bliemer and Taale (2006) to the Dutch national main road network in the Netherlands. This method can generate a choice set of attractive and heterogeneous (from the doubly stochastic method) alternatives which is suitable for model estimation and flow prediction. However, the assumptions of the distributions can be very sensitive and can be challenging to collect.

Constrained Enumeration Methods

The underlying idea of this method is to enumerate paths according to behavioural rules rather than the minimum cost path. In this method, different constraints are set to avoid unattractive paths to be enumerated. This approach is first introduced by Friedrich (1994), which is termed as a "branch and bound" algorithm, and later used by other studies for road networks (Bekhor and Prato, 2009; Prato and Bekhor, 2006), multimodal networks (Hoogendoorn-Lanser, 2005) and transit networks (Friedrich et al., 2001). This method proved to generate most of the attractive paths (Bekhor and Prato, 2009; Prato and Bekhor, 2006). Although this method needs higher computation time, it can be applied to large networks (Bekhor and Prato, 2009). However, according to Prato (2009), the lack of knowledge about the effects of the threshold value of the behavioural constraints can make it challenging to correctly apply this method.

Probabilistic Methods

The probabilistic method was proposed by Manski (1977), which assigns a probability to each path in the generation process. This method was later improved to avoid extensive computational requirements by calculating the selection probability of all the enumerated paths. Cascetta and Papola (2001) proposed to calculate the probability of the alternatives which were only considered in the choice set rather than all alternatives. Frejinger (2007) and Frejinger et al. (2009) used the "random walk" approach, which assigns a probability to each link starting from the origin to the destination according to the Kumaraswamy distribution. The probability of the path included in the choice set is calculated from the product of the associated link probabilities. This approach also uses a correction term to account for the unequal selection probability. The study applied this method on a synthetic network which showed promising results with the inclusion of the correction term. Very recently, Zimmermann et al. (2018) applied this method in a multi-modal network of Zurich. Rasmussen et al. (2017) presented a route choice model for transit assignment called RSUET (Restricted Stochastic User Equilibrium with Threshold) where paths are generated in a doubly stochastic approach. The introduction of both preference and taste heterogeneity gives this model a strong base to reflect user behaviour. Taste heterogeneity was modelled at the OD level, and preference heterogeneity was modelled at the link level. The approach also used the error term for path commonality.

2.2.1.2 Choice Set Generation for Transit Network

On top of road network attributes, transit networks have additional network attributes such as transit links, walking links and waiting links/nodes. As such, traditional choice set generation methods can be also be used to generate transit paths from any origin to destination. However, the complexity will be to address the transfers and the schedule/frequency of travel.

In the literature, limited research on the transit path generation problem was found. Abdelghany (1999), and Abdelghany and Mahmassani (2001) both used a k-shortest path in a multimodal network. Lozano and Storchi (2001) used a k-multimodal shortest path algorithm where they considered two conditions, a viable path for the order of the used modes and the number of transfers. Bielli et al. (2006) added another condition of time constraints to generate the transit paths in a multimodal network while Ziliaskopoulos and Wardell (2000) considered delays of transit modes and links to develop an intermodal time-dependent least-time path algorithm.

A multi-objective shortest path algorithm was applied by Abdelghany and Mahmassani (1999) and Florian (2004). Horn (2003) used Dijkstra's label-setting shortest path algorithm where different generalised cost functions were computed in a multimodal (walking, public transit and taxi) network setting. Florian (2004) adopted both time and cost based on Dijkstra's label setting algorithm.

A simulation approach was presented by Benjamins et al. (2001), where a different choice set is developed for different traveller groups. This stochastic approach was further extended by Fiorenzo-Catalano et al. (2004) by simulating link costs which were enumerated after simulating the traveller's behaviour.

Friedrich et al. (2001) presented a study where the branch and bound method was used in a time-dependent public transit network. Later, Hoogendoorn-Lanser (2005) extended this approach to a multimodal network by implementing appropriate restrictions for multimodal characteristics. Again, further constraints were set to eliminate irrelevant alternatives and maintain the spatial and functional variety of the choice set.

Tan (2016) presented a comprehensive comparison among the path generation methods including the labelling approach, link elimination (N=30) approach, simulation approach (50 draws), k-shortest path (k=30) approach, nested labelling & link elimination and branch & bound approach (maximum transfer = 5). Tan (2016) found that the labelling approach took less computational time than the other methods (an average of 13.9 seconds for one OD pair) while the branch and bound approach could not finish generating paths for 200 OD pairs in two weeks. The nested labelling & link elimination and link elimination approaches seem to be the best approaches as they do not need much computation time and the choice sets have the produce highest coverages.

Anderson et al. (2017) used a doubly stochastic path generation algorithm based on a schedule-based stochastic transit assignment model developed by Nielsen (2000) and refined by Nielsen and Frederiksen (2006). Anderson et al. (2017) included error components within the utility function to capture the preference heterogeneity.

Zimmermann et al. (2018) applied the recursive logit model proposed by Frejinger et al. (2009) in a multimodal network in Zurich. The high computational time of the recursive logit model was reduced by solving the value function for only one subset of links.

Khani et al. (2014) presented an algorithm named TBSP (Trip Based Shortest Path) which aims to model three critical aspects including the complexities of a time-dependent network, realistic user behaviour, and higher efficiency in path search. This algorithm has the following characteristics:

- Schedule based transit assignment which incorporates GTFS (General Transit Feed Specification) data developed by Google.
- Uses transit network hierarchy where any deviation from a route (i.e. transfer to another route) can be done only at some specific stops.
- Develop a new trip based network structure instead of a node-based (Ahuja et al., 1993) or a link-based (Noh et al., 2012) structure.

A few studies (Malekzadeh, 2015; Nassir et al., 2015b, 2016) applied this algorithm and found an impressive level of accuracy regarding the inclusion of the chosen alternative in the choice set.

2.2.2 Adopted Path Generation Method

As discussed in the previous subsections, there are few alternatives available in the literature which can be used to generate path sets for route choice modelling. For example, the doubly stochastic algorithm developed by Nielsen (2000) or the nested labelling and link elimination approach used in Tan (2016) can be the right choice as these proved to generate paths with substantial coverage. However, developing a path generation method was out of the scope of this research. Again, a version of the Trip-Based Shortest Path (TBSP) algorithm developed by Khani (2013) was available to use for this research which also proved to generate paths with substantial accuracy (Malekzadeh, 2015; Nassir et al., 2015b, 2016). Moreover, the TBSP algorithm can generate paths with three transfers, which was also desirable for the SEQ HTS-2009 data.

Therefore, this study used a version of the Trip-Based Shortest Path (TBSP) algorithm developed by Khani (2013) and Khani et al. (2014; 2012), and later modified by Nassir et al. (2015a; 2015b). This algorithm was used to generate preliminary route sets in four steps showed in Figure 2.3. The first step involved the preparation of the data, consisting of trip information (origin and destination location, departure time), transport and walkway networks and transit schedules. The next step eliminated route segments after the generation of each route. The fourth step checked the reasonability of the route by imposing certain conditions and thus formed the route set.

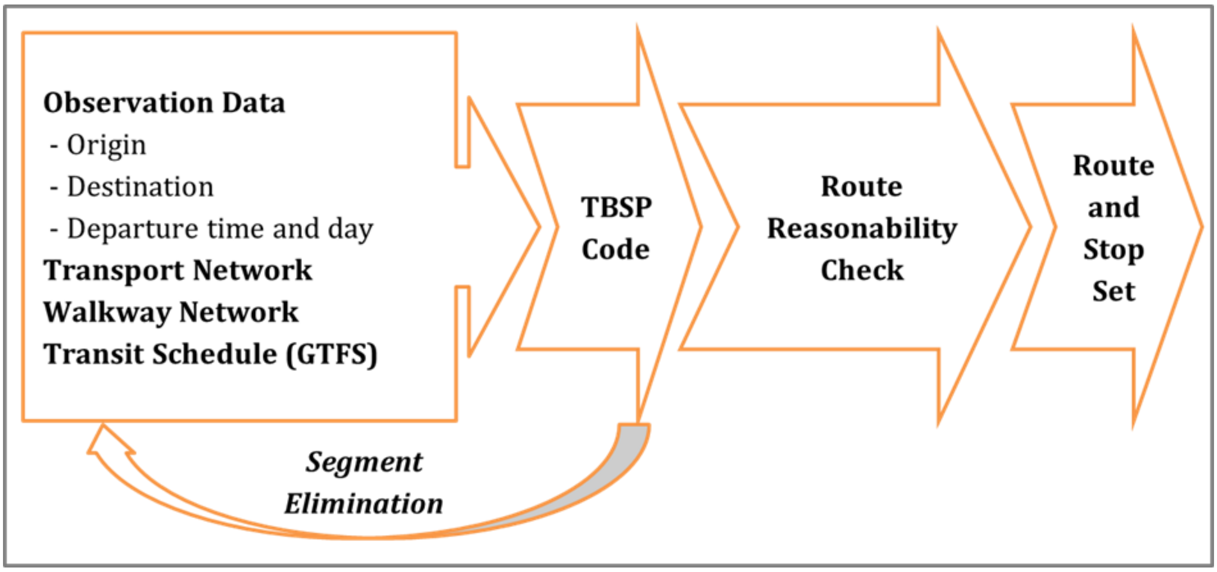

Figure 2.3 Route Set Generation Framework

2.2.2.1 TBSP Code

The version of the TBSP code used in this thesis is a transit time-dependent path generation algorithm, which aims to minimise the arrival time to the destination. Furthermore, for computational efficiency, this is modified to terminate after the destination is marked.

2.2.2.2 Segment Elimination

The "segment elimination" step is a heuristic which executes after each iteration (path generation) of the TBSP code. A segment is a combination of three elements:

boarding stop, alighting stop, and the route connecting these two stops. In each iteration, after the TBSP generates a path, the segment elimination step removes its segments so that the next path alternatives do not contain the same segments (ride on the same route from the same stops to the same stop). Alternatively, the whole route could have been eliminated, but that would also eliminate path alternatives using the same route but transferring at different points.

2.2.2.3 Route Reasonability Check

In the third step, reasonable paths are sorted out. Three reasonability conditions were embedded in the TBSP code: 1) the transfer walking distance cannot exceed 1 km, 2) the waiting time before boarding cannot exceed one hour, and 3) the access and egress walks cannot exceed 2 km (30 minutes, @4 km/hr). This 2 km threshold range was taken from the preliminary analysis of access walks from the SEQ household travel survey (HTS) data where about 17% of the observations were found to walk more than 1 km to access a transit stop (Nassir et al. 2015).

Besides the above-mentioned constraints which were embedded in the TBSP algorithm, additional constraints were needed to restrict the path generation algorithm to achieve a reasonable path set. As this algorithm can generate paths containing multiple transfers and as the algorithm itself did not have restrictions on the travel time, it would have generated paths which are not realistic. Therefore, two other reasonability checks were performed after the TBSP path generation. These two criteria were: 1) the path travel time did not exceed the shortest path travel time plus a threshold factor known as off-optimality, and 2) the number of transfers did not exceed three. The maximum off-optimality threshold was set to 20 minutes, as suggested in the literature (Nassir et al., 2015a). Again, from the HTS-2009 data, it was found that the number of transfers does not exceed three.

2.2.2.4 Route and Stop Choice Set

After the path generation process, the data were analysed and cleaned to prepare the stop choice dataset. The generated paths provide the information of access stop identity, walking time to access stop, travel route number, in-vehicle travel time, the identity of transfer stops, walking time for transfer, the identity of egress stop, egress walking time, and the correction factors for path commonality. The analysis found that the TBSP algorithm could select about 94.5% of the chosen access stops (1,599 out of 1,693) successfully. The unsuccessful choices of stops were added in the choice set manually. The impedance attributes of these stops were calculated by restricting the shortest path generation algorithm to start from these stops. However, some observations were not matched to the exact stop location. These locations were inferred by applying three matching keys: whether the distance was within a 100m threshold, the mode of the stops and the path serving that stop. Ultimately, some of the observations had to be excluded (about 26.8%) as the chosen access stop could not be located because of the observed ambiguity between the HTS data and generated paths. In the end, the maximum number of stop alternatives in a set was found to be 70 stops, though the majority of observations had less than 20 stop choices in the set (see Figure 2.4). Basic statistics of the stop choice sets are presented in Table 2.3.

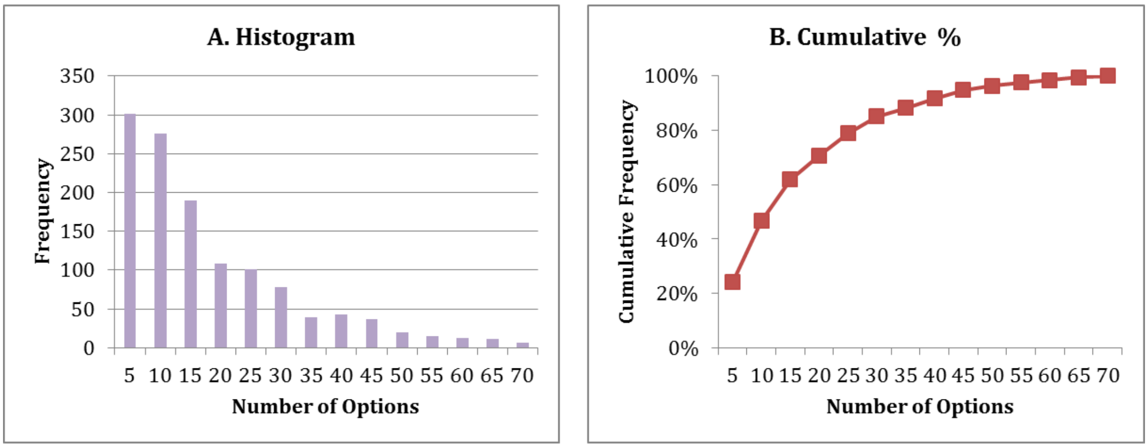

Figure 2.4 Stop Choice Set Analysis

Table 2.3 Basic Statistics of Stop Choice Set

Statistics	Stop Choice
Mean	16.19
Median	11
Mode	2
Range	68
Minimum	2
Maximum	70
Total Observations	1237

Again, the observations finalised for the stop choice modelling were taken into consideration for the route choice modelling. The generated paths from the TBSP algorithm were evaluated to see the coverage measure proposed by Ramming (2002) (see subsection 2.4.3 for details). One aspect of the HTS-2009 data is that respondents mentioned the perceived time taken for travel components, including walking time to access stops, waiting time, transfer walking time and egress time. However, the walking routes were unknown. Again, the analysis showed that passengers' reported time varies from the calculated time (@4km/hr) as also found in other studies (Anderson et al., 2017; Meng et al., 2018). Therefore, the coverage measure is limited to the access stop to egress stop and the line-level coverage is computed. It was found that the TBSP algorithm could generate about 82.5% of the chosen paths with a 100% overlap threshold. However, the coverage was increased to 98.8% with 80% overlap. Finally, the remaining 1.2% of paths (15 observations) were added in the choice sets.

Basic statistics of the generated route choice sets are shown in Table 2.4. The range of choice set size seems to be very high, ranging from 2 to 1925. Figure 2.5 shows that the majority (about 80%) of the observations have a choice set size of less than 50. However, there are observations where the choice set size is very large.

Table 2.4 Basic Statistics of Route Choice Set

Statistics	Route Choice
Mean	35.63
Median	18
Mode	4
Range	1923
Minimum	2
Maximum	1925
Total Observations	1237

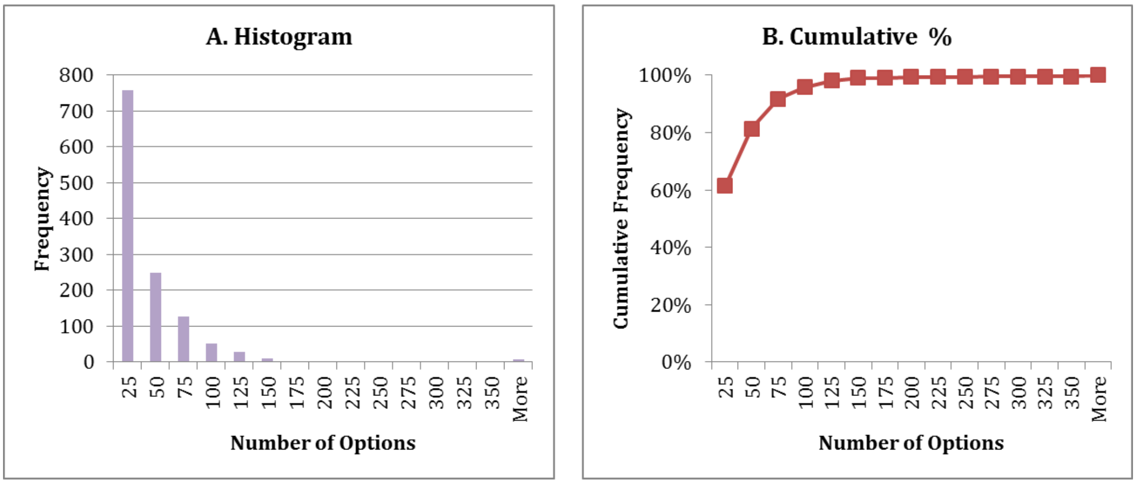

Figure 2.5 Route Choice Set Analysis

At the end of this part, two separate datasets for stop choice and route choice modelling were developed with all their relevant attributes.

2.2.2.5 Explanatory Variables

Several transit choice studies were reviewed to develop explanatory variables. Debrezion et al. (2009) mainly considered station facility attributes to construct their model. Chakour and Eluru (2013) considered socio-demographic attributes, trip characteristics, facility attributes and land-use and built environment factors. Mahmoud et al. (2014) studied facility attributes and land-use variables. Anderson et al. (2017) considered travel time components along with transfer details, frequency, and purpose of travel. Tan (2016) considered travel time components, transfer and travel cost. Vrtic and Axhausen (2003) considered in-vehicle time, transfer time, number of transfer and headway along with the purpose of the trip. Other studies (Van der Waard, 1988; Bovy and Hoogendoorn-Lancer, 2005; Raveau et al., 2011; Eluru et al., 2012) included travel time components and transfers. Nassir et al. (2015b) considered facility attributes, impedance attributes and correction attributes. Sociodemographic attributes are considered by Prato et al. (2012) and Abdel-Aty and Huang (2004). In this study, the explanatory variables can be classified into seven classes: 1) impedance attributes, 2) access stop attributes, 3) mode attributes, 4) strategy attributes, 5) trip attributes 6) user attributes and 7) correction attributes. Different models considered different attributes. A brief description and use of the variables are provided in **Appendix A**.

Impedance Attributes

Impedance attributes were the route attributes calculated from the route generation process. These included total travel time, in-vehicle travel time, travel time by train, travel time by bus, travel time by ferry, total walking time (summation of access walk, transfer walk and egress walk), other walking time (summation of transfer walk and egress walk), waiting time and the number of transfers.

Besides these two sub-groups of impedance (direct and aggregate), attributes were generated for the stop choice modelling. Direct impedance attributes included five variables: fastest travel time, minimum transfer, minimum walking, minimum fare and minimum waiting time. These attributes represented the most attractive path from the stop. For example, for a particular stop, the fastest travel time variable indicated the fastest travel time of all paths from that stop. Similarly, the minimum number of transfers of all paths from the stop was recorded for the minimum transfer variable, and so on. Aggregate impedance attributes (including averages among all reasonable paths) included seven variables among which five included the average measure (travel time, transfer, walking time, fare and waiting time) among all reasonable paths. The other two contained the total number of routes from the access stop to the destination and the total frequency of all these paths.

Access Stop Attributes

Access stop attributes were the variables of the transit access stop, such as walking time to access the stop, availability of shelter, lighting, footpath, boarding slab, and the number of maps/ schedule information at the stop. Mode attributes contained two types of binary variables: the first type considered the type of mode used in the route/stop and the second type considered the primary mode used in that route/stop. The first group of mode attributes contained three variables: only bus, only train and mixed-mode (containing at least two different modes). The second group contained three variables: mainly bus, mainly train, and mainly ferry.

Strategy Attributes

The strategy attributes were binary variables, which showed whether the routes had unique characteristics like minimum travel time, minimum access time and a minimum number of transfers. The strategy attributes for the MNL model were calculated from the sampled choice set. However, for the nested model, as the users' perception of time does not precisely match reality (Anderson et al., 2017; Malekzadeh, 2015; Meng et al., 2018), "reasonable minimum" travel times and access times were considered. Ten per cent of the differences between the maximum and minimum travel/access times (of all the paths in each case) was added to the minimum travel/access time to calculate the "reasonable minimum" travel time and access time. From the definition of "reasonable minimum", it is understood that the relationship between the original variable (e.g. minimum travel time) and the derived variable (minimum travel time strategy) is not direct or linear. The correlation coefficients of these variables are presented in Table 2.5, which shows a very weak correlation between travel time and minimum travel time strategy. A mild correlation between access walk time and minimum access time strategy is evident for the route choice data while this correlation seems to be moderate for the stop choice data. However, a considerable correlation is evident between the number of transfers and the minimum transfer strategy. Therefore, to avoid multicollinearity, these two transfer variables were not considered together in the model specifications. Again, the access time variables were not considered together in the model specifications for stop choice models. Strategy attributes were used as dummies in the MNL and Mixed MNL models and also used to form the nests for NL models.

Table 2.5 Correlation Coefficient for the Strategy Variables

Original Variable	Derived Variable	Correlation Coefficient	
		Stop Choice Data	Route Choice Data
Travel Time	Minimum Travel Time Strategy	-0.03	-0.09
Access Walk Time	Minimum Access Time Strategy	-0.49	-0.36
Number of Transfer	Minimum Transfer Strategy	-0.59	-0.65

Trip and User Attributes

Trip attributes contained variables regarding trip timing and trip purpose. User attributes contained a variety of socio-economic attributes of the user, including age, gender, household and family-related information, income, and job.

Correlation Between the Attributes

A high correlation between attributes can affect the modelling exercise with the multicollinearity issue. Although multicollinearity does not significantly affect the estimated parameters, the standard error of the coefficient will be high. As a result, precise estimations cannot be obtained. Therefore, the correlation coefficient of the route choice datasets and stop choice datasets are analysed and presented in **Appendix B.**

2.3 Model Structure

The third segment involves the modelling exercise. In this study, two major modelling approaches, RUM and RRM, are considered for modelling two types of transit choices (stop and route). Brief discussions on the modelling structures are presented in this section. Table 2.6 summarises the model structures used in the study.

Table 2.6 Models Considered for the Transit Choice Modelling Study

Modelling Approach	Model Structure	Stop Choice	Route Choice
Random Utility Maximization (RUM)	Multinomial Logit (MNL)	√	
	Mixed Multinomial Logit (Mixed MNL)	√	
	Nested Logit (NL)	√	
	Mixed Nested Logit (Mixed NL)	√	
	MNL with Sampling		√
	NL with Sampling		√
	MNL with Sampling & Path Size Correction		√
	NL with Sampling & Path Size Correction		√
Random Regret Minimization (RRM)	P-RRM	√	
	μRRM	√	
	Hybrid RUM-RRM	√	√
	RRM with Sampling		√
	Hybrid RUM-RRM with Sampling		√
	RRM with Sampling and Path Size Correction		√

2.3.1 RUM Models

2.3.1.1 Multinomial Logit (MNL)

In the MNL structure, the restricting Independence of Irrelevant Alternatives (IIA) property holds. The form of MNL can be described by equation (2.1).

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_i e^{\beta' x_{ni}}} \qquad (2.1)$$

Where, P_{ni} is the probability of selecting alternative i by individual n,

x_{ni} is the column vector associated with attributes influencing the choice, and β′ is the vector of parameters to be estimated.

2.3.1.2 Mixed Multinomial Logit (Mixed MNL)

The mixed MNL model is used to capture random taste variations among individuals. In the mixed MNL formulation, β' is treated as a random parameter to be estimated, having a probability density function of $f(\beta)$. The choice probability of the mixed MNL form can be written by the form provided in equation (2.2)

$$P_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} f(\beta) d\beta \qquad (2.2)$$

2.3.1.3 Nested Logit (NL)

Nested Logit (NL) is chosen to capture the correlation between alternatives belonging to different travel strategies. It is assumed that alternatives falling under the same strategy have some unobserved similarities between them and a nested structure might be able to capture them. Here, the strategies are considered to form the nests and the stops/routes associated with the strategy are included under that nest (details are provided in Subsection 3.2.3). In the NL formulation, the choice probability for alternative i ▯ B_k can be written as in equation (2.3).

$$P_{n\,i} = \frac{e^{\beta' x_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{\beta' x_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} e^{\beta' x_{nj}/\lambda_l}\right)^{\lambda_l}} \qquad (2.3)$$

2.3.1.4 Mixed Nested Logit (Mixed NL)

Mixed NL can capture both random taste variations and correlation among the alternatives. Recently, some studies (Antonini et al., 2004; Bajwa et al., 2008; Hammadou et al., 2008; Hess et al., 2005) reported a technique where the β' coefficients inside the nests were treated as random parameters with a function of $f(\beta)$. The nest coefficients were not assumed to have any distribution. The model can be written as in equation (2.4).

$$P_{n\,i} = \int \frac{e^{\beta' x_{ni}/\lambda_k} (\sum_{j \in B_k} e^{\beta' x_{nj}/\lambda_k})^{\lambda_k - 1}}{\sum_{l=1}^{K} (\sum_{j \in B_l} e^{\beta' x_{nj}/\lambda_l})^{\lambda_l}} f(\beta) d\beta \qquad (2.4)$$

In the mixed models, randomness is captured by assuming a log-normal distribution for the variables that show negative signs in MNL models, uniform distribution for dummy variables and a normal distribution for all the other variables (Hensher and Greene, 2001).

2.3.1.5 Correction for the Path Commonality

The Multinomial Logit (MNL) discrete choice model has a desired closed-form probability expression that can make the computation and calibration of discrete choice models much more straightforward. However, this closed-form probability expression is mainly based on the assumption that the error (or random) components in the utility of alternatives are independently and identically (and Gumbel) distributed (IID). However, in many applications, the independence of these error components is violated. To take advantage of the logit closed-form probability expression, there have been many useful extensions to the MNL model that can treat this violation. The approaches proposed in the literature can be classified into two groups. The first group treats the problem in a direct adjustment of the error distributions by directly defining the structure of correlation among alternatives. Well-known examples of these models are Nested Logit, Cross-nested Logit and Network GEV. However, capturing the correlation structure and calibrating such models can become very expensive (regarding data requirements and computational complexity) in specific applications, particularly in real-sized transportation networks. Route choice models are among such examples. The combinatorial nature of the path alternatives can make the correlation structure very complicated. Therefore, another group of models have been developed that simplifies the situation by approximating the correlation structure of alternatives from the spatial overlap of paths. C-logit (Cascetta et al., 1996; Cascetta and Papola, 2001), Path Size logit (Ben-Akiva and Bierlaire, 1999) and Path Size Correction logit models (Bovy et al., 2008) belong to this group. The correlation structures of these models are shown in equations 2.5 to 2.8 respectively.

$$corr_i = CF_i = ln \sum_{j \in C} (\frac{L_{ij}}{\sqrt{L_i L_j}}) \tag{2.5}$$

$$corr_i = CF_i = ln \left[1 + \sum_{\substack{j \in C \\ j \neq i}} (\frac{L_{ij}}{\sqrt{L_i L_j}}) \left(\frac{L_i - L_{ij}}{L_j - L_{ij}} \right) \right] \tag{2.6}$$

Where, L_i is the length of route i,

L_j is the length of alternative route j within the choice set C,

L_{ij} is the overlapping length between routes i and j.

$$corr_i = ln\ (PS_i) = ln \left(\sum_{a \in \tau_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in C} \delta_{aj}} \right) \tag{2.7}$$

$$corr_i = PSC_i = - \sum_{a \in \tau_i} \left(\frac{L_a}{L_i} ln \sum_{j \in C} \delta_{aj} \right) \tag{2.8}$$

Where, PS_i is the path size of route i,

PSC_i is the path size correction of route i,

L_i is the length of route i,

L_a is the length of link a,

C_i is the set of links belonging to route i,

δ_{aj} is the link-route incidence dummy that is equal to one if route j within the choice set C uses links a and zero otherwise.

These models discount the deterministic (systematic) component of the utility of overlapped alternatives by using a factor that estimates the amount of overlap among the alternatives. Notably, in the path size logit model, the path size factor (depending on which definition is considered) estimates a weighted measure of the overlap that a particular path has with all other alternatives cumulated along with all its segments. This measure represents what fraction of an entirely independent alternative a particular path can be. In the transit route choice literature, travel time (Tan, 2016) and travel length (Anderson et al., 2017) measures are used for the overlapping elements.

In this path size correction, a path[1] is considered as a door-to-door sequence of transit and walking links, and it can consist of one or more transit routes. For example, Path Alt1 would consist of walking from the origin to the access stop, taking Route 414, then transferring to Route 418 and walking to the destination. Equation 2.9 is used to measure the path size correction and used as an explanatory variable in the utility function.

$$PSC_i = \sum_{a \in L_i} \frac{t_{a,i}}{t_i} \frac{1}{\sum_j \delta_{aj}} \qquad (2.9)$$

Where, PS_i is path size factor for path i;

L_i is the set of all routes in path i;

a is the identifier of the route (bus, train or ferry);

$t_{a,i}$ is waiting + in-vehicle-time on route a (when traveling on path i);

t_i is door-to-door travel time on path i;

δ_{aj} is incidence parameter of route a and path j (Incidence parameter δ_{aj} equals

one if route a is on path j, and zero otherwise).

The waiting time is calculated from the difference between the scheduled time (given in GTFS) and the arrival time at the stop (computed from the given departure time from the survey and the walking time taken to reach to the stop).

Three correction factors (CfC1, CfC2, CfC3) are proposed for the stop choice modelling (equations 2.2 to 2.4) based on the Path Size Correction Logit (PSCL) formulation; however, were adjusted to meet the specifications of the access stop choice model (Nassir et al., 2015b, 2016). These correction factors aim to capture the interdependencies among the stops that result from the common routes to the destination

[1] In the route choice literature the terms path and route are used interchangeably. However, in this thesis a trip based shortest path algorithm is used which follows a trip based network rather than link/node based. Therefore, by definition it used transit route as a part of the path.

For an observation from origin location o at departure time τ to destination location d, three definitions of correction for correlation were defined for every stop s in the choice set $C_o^{d,\tau}$:

$$CfC1_s^{d,\tau} = -\sum_{i\in\Gamma_s^{d,\tau}} \frac{1}{\left|\Gamma_s^{d,\tau}\right|} ln \sum_{t\in C_o^{d,\tau}} \delta_{i,t}^{d,\tau} \tag{2.10}$$

$$CfC2_s^{d,\tau} = -\sum_{i\in\Gamma_s^{d,\tau}} \frac{f_{i,s}^{\tau}}{\sum_{j\in\Gamma_s^{d,\tau}} f_{j,s}^{\tau}} ln \sum_{t\in C_o^{d,\tau}} \delta_{i,t}^{d,\tau} \tag{2.11}$$

$$CfC3_s^{d,\tau} = -\sum_{i\in\Gamma_s^{d,\tau}} \frac{\left(T_{j,d}^{\tau}\right)^{-1}}{\sum_{j\in\Gamma_s^{d,\tau}}\left(T_{j,d}^{\tau}\right)^{-1}} ln \sum_{t\in C_o^{d,\tau}} \delta_{i,t}^{d,\tau} \tag{2.12}$$

Where, i, j: Indices of routes;

s, t: Indices of stops;

$\Gamma_s^{d,\tau}$: Set of all routes at stop s with reasonable paths to the destination d at time τ;

$f_{i,s}^{\tau}$: Frequency of route i at stop s at time

τ;$T_{j,d}^{\tau}$: Travel time of the fastest path from stop s boarding on route i to destination d at time τ;

$\delta_{i,t}^{d,\tau}$: Stop-route incidence parameter, $\delta_{i,t}^{d,\tau} = \begin{cases} 1, if\ i\ \in \Gamma_s^{d,\tau} \\ o, if\ i \notin \Gamma_s^{d,\tau} \end{cases}$

2.3.1.6 MNL with Sampling

When the number of alternatives is large, it might be computationally challenging to estimate the model, especially in the transit route choice scenario where the origin and the destination are highly accessible by transit. From the path generation stage discussed in subsection 2.2.2.4, it is found that a portion of the route choice set contains a very high number of paths. For this, a subset of the alternatives (a choice set) can be considered in the MNL model and can be consistently estimated by using the conditional maximum likelihood estimation approach (McFadden, 1977). In this approach, a sampling of alternatives can be addressed with a simple correction to the log-likelihood function (McFadden et al., 1978). In this case, the probability that an individual n chooses alternative i is then conditional on the choice set C_n and can be written as equation 2.13.

$$P(i|C_n) = \frac{e^{\mu V_{in} + \ln q(C_n|i)}}{\sum_{j \in C_n} e^{\mu V_{jn} + \ln q(C_n|j)}} \tag{2.13}$$

Where, μ is a scale parameter, and

$\ln q(C_n|j)$ is an alternative specific term that corrects for the sampling bias.

This correction term $\ln q(C_n|j)$ is based on the probability $q(C_n|j)$ of sampling C_n, given that j is in C_n.

2.3.1.7 NL with Sampling

The correction term for the choice set formation can also be incorporated in the Nested Logit (NL) formulation. In this study, it is assumed that routes (alternatives) falling under the same strategy (nests) have some unobserved similarities among them, and a nested structure might be able to capture that. Here, the strategies are considered to form the nests, and the routes associated with each strategy are included under that nest (see details of nest structure in subsection 4.2.6). Guevara and Ben-Akiva (2013b) showed that sampling of alternatives could also be done for the NL formulation by considering NL as a multivariate extreme value (MEV) model. Their formulation is an extension of McFadden's (1978) logit formulation and seems to be consistent, asymptotically normal and efficient. They (Guevara and Ben-Akiva, 2013b) suggest two different choice sets (D_n and $\widetilde{D}_n$) to be drawn from the universal set C_n when the researcher has full control of the dataset (as in this study). D_n will be used to determine the sampling correction ($\ln q(D_n|i)$) and $\widetilde{D}_n$ will be used to conduct all the other calculations. Therefore, the conditional probability of choosing alternative i, given that the sets D_n and $\widetilde{D}_n$ were drawn separately from the universal set C_n, will correspond to:

$$P(i|D_n, \widetilde{D}_n) = \frac{e^{\mu V_{in} + \ln f(\widehat{B}_{in}(\widetilde{D}_n)) + \ln q(D_n|i)}}{\sum_{j \in D_n} e^{\mu V_{in} + \ln f(\widehat{B}_{jn}(\widetilde{D}_n)) + \ln q(D_n|j)}} \tag{2.14}$$

Equation 2.14 differs from equation 2.13 by adding a term, $\ln f(\widehat{B}_{in})$, adjusting for the nesting. This term can be calculated from equation 2.15.

$$ln\, f(\hat{B}_{in}) = \left(\frac{\mu}{\mu_{m(i)}} - 1\right)\left(ln \sum_{j \in \widetilde{D}_{m(i)n}} w_{jn}\, e^{\mu_{m(i)} V_{jn}}\right) + ln\, \mu + \left(\mu_{m(i)} - 1\right) V_{in} \qquad (2.15)$$

Where, μ_m is the scale parameter of nest m, and

w_{jn} is the expansion factor which is calculated from $\widetilde{D}_n$.

The correction term (and the expansion factor, w_{jn}) can be different for different sampling protocols. However, a common form is presented (Frejinger et al., 2009; Guevara and Ben-Akiva, 2013b) as in equation 2.16.

$$ln\, q(C_n|i) = ln \frac{k_{in}}{q(i)} = ln\, k_{in} - ln\, q(i) \qquad (2.16)$$

Where, k_{in} is the number of times alternative *i* is drawn, and

$q(i)$ is the sampling probability of alternative *i*.

2.3.2 RRM Models

2.3.2.1 Pure RRM

The functional form of the first RRM model, also known as pure RRM, is given in equation 2.17.

$$R_i = max_{j \neq i}\left\{\sum_m max\left\{0, \beta_m . \left(x_{jm} - x_{im}\right)\right\}\right\} \qquad (2.17)$$

Where, R_i denotes the 'observed' regret associated with *i*

β_m denotes the estimable parameter associated with the attribute x_m

$x_{jm} - x_{im}$ denote the values associated with the attribute x_m for, respectively, the considered alternative *i* and another alternative *j.*

In pure RRM, regret is considered by focusing on the best competitive alternative. Another notable feature is that only the positive values of the regret are taken into account. Consequently, this function provides a kink around zero (see Figure 2.6A), which raises problems with estimation and derivation of elasticities, marginal effects, and willingness to pay. Again, this model failed to address some fundamental behavioural issues like the irrelevancy of ‘rejoice’ and peoples’ feelings about “regret concerning the best non-chosen alternative” (Chorus, 2012).

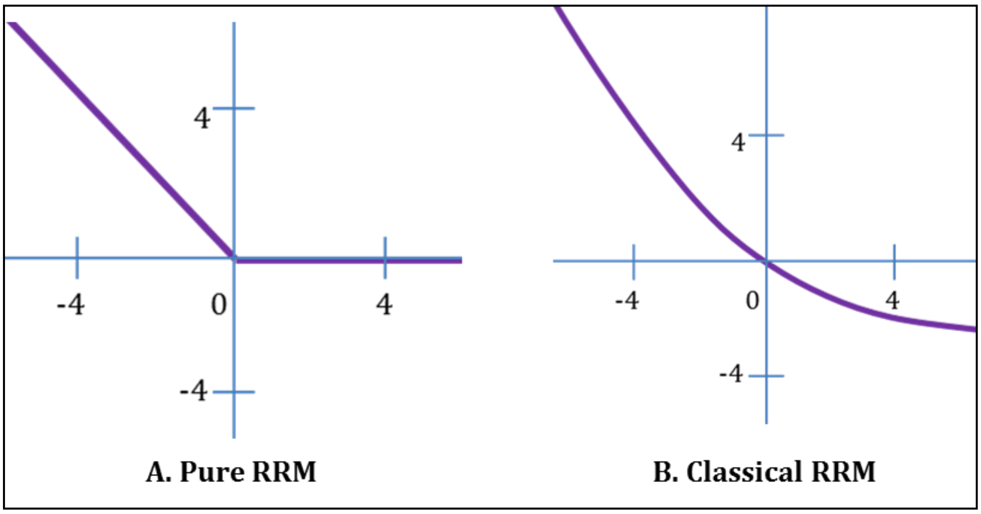

Figure 2.6 Shape of RRM Function

2.3.2.2 Classical RRM Model

Considering the econometric as well as the behavioural issues in the pure RRM model, the second version (Chorus, 2010) replaces the max operator with the logsum operator. This version is termed as the classical RRM Model, and the functional form of this model is presented in equation 2.18. Consequently, a smooth and differentiable function is achieved (see Figure 2.6B). Furthermore, the assumptions are also rectified as follows (Chorus, 2012).

1. When a considered alternative outperforms another alternative regarding a particular attribute, the comparison of the considered alternative with the other alternative on that attribute does not generate anticipated regret.
2. When a considered alternative is outperformed by another alternative regarding a particular attribute, the comparison of the considered alternative with the other alternative on that attribute generates anticipated regret.
3. Anticipated regret increases with the importance of the attribute on which a considered alternative is outperformed by another alternative.

4. Anticipated regret increases with the magnitude of the extent to which a considered alternative is outperformed by another alternative on a particular attribute.
5. Anticipated regret increases with the number of attributes on which the considered alternative is outperformed by another alternative.
6. Anticipated regret increases with the number of alternatives that outperform a considered one on a particular attribute.
7. Anticipated regret is, from the perspective of the analyst, partially 'observable' (in the sense that it can be explicitly linked to observed variables) and partially 'unobservable'.

$$RR_i^{classical} = R_i + \varepsilon_i = \sum_{j \neq i} \sum_m ln\big(1 + exp\big[\beta_m.\big(x_{jm} - x_{im}\big)\big]\big) + \varepsilon_i \qquad (2.18)$$

Where, $RR_i^{classical}$ denotes the random (or: total) regret associated with a considered alternative *i*

ε_i denotes the 'unobserved' regret associated with *i.*

2.3.2.3 G-RRM model

The generalised RRM model or G-RRM model (Chorus, 2014) generalises the Classical RRM model to achieve more flexibility to operate in different circumstances (see equation 2.19). In this model, the fixed constant term 1 is replaced by a regret-weight variable denoted γ. The effects of γ can be seen in Figure 2.7, where γ = 1 results in the Classical RRM model. As γ gradually increases the attribute level regret function becomes less convex. In the particular case in which γ = 0, the G-RRM model predicts the same choice behaviour as the linear-additive RUM model.

$$RR_i^{G-RRM} = \sum_{j \neq i} \sum_m ln\big(\gamma + exp\big[\beta_m.\big(x_{jm} - x_{im}\big)\big]\big) + \varepsilon_i \qquad (2.19)$$

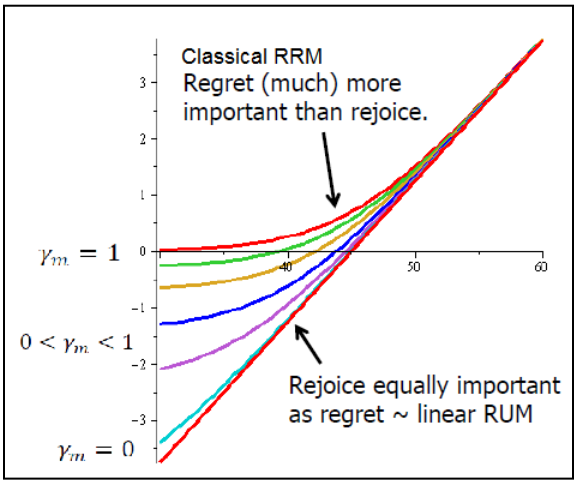

Figure 2.7 Visualization of G-RRM Model (Chorus, 2016)

2.3.2.4 µRRM Model

The µRRM model (see equation 2.20) introduces a shape parameter denoted as μ in the Classical RRM model (Van Cranenburgh et al., 2015b). This shape parameter is in fact confounded with the scale. Figure 2.8 shows the shapes of the attribute level regret function for different size of μ. For, μ=1 it took the shape of the classical RRM model. Consequently, the size of μ shows the degree of regret minimisation behaviour imposed by the µRRM model which is termed as the profundity of regret. Estimating a relatively large scale parameter (μ >10) signals a mild profundity of regret, implying fully compensatory (RUM) behaviour.

$$RR_i^{\mu RRM} = \sum_{j \neq i} \sum_m ln\left(1 + exp\left[\frac{\beta_m}{\mu}.(x_{jm} - x_{im})\right]\right) + \varepsilon_i \qquad (2.20)$$

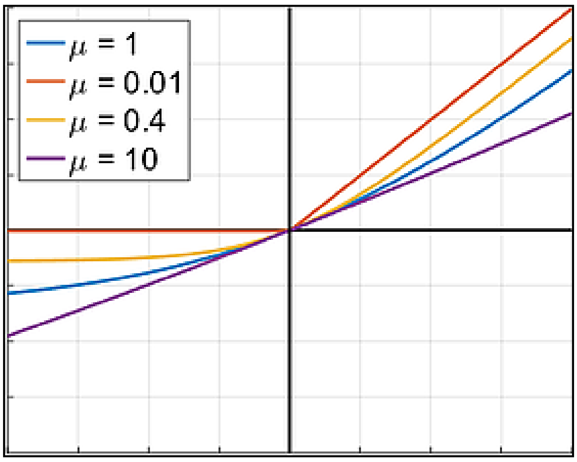

Figure 2.8 Visualization of µRRM Model
(Source: https://www.advancedrrmmodels.com/mu-rrm)

2.3.2.5 P-RRM

Recently another form named the P-RRM model was presented (Van Cranenburgh et al., 2015b) which postulates the strongest RRM behaviour possible within the RRM modelling framework. This is, in fact, one of the two special limiting cases of the μRRM model. The crucial concept behind this model is, in fact, in contrast to the classical RRM and μRRM model as it assumes that no joy will be experienced when the considered alternative outperforms a competitor alternative with regard to an attribute *m.* The model is shown in equation 2.21, where the main component of the model is to compute a so-called P-RRM x vector. The estimation of the P-RRM model is straightforward because of its linear form and the fact that the x-vector can be computed before the estimation. This also reduces the runtime significantly as compared to the previous RRM models.

$$RR_i^{P-RRM} = \sum_m \beta_m x_{im}^{P-RRM} \ \ where \ x_{im}^{P-RRM} = \begin{cases} \sum_{j\neq i} max\left(0, x_{jm} - x_{im}\right) \ if \ \beta_m > 0 \\ \sum_{j\neq i} min\left(0, x_{jm} - x_{im}\right) \ if \ \beta_m < 0 \end{cases} \tag{2.21}$$

2.3.2.6 Error Term of the RRM Models

The negative of the error term of the RRM models is assumed to be IID-type 1 Extreme Value distributed with a variance of $\pi^2/6$, resulting in the well-known and convenient closed-form logit formula. However, in the μRRM model, the error variance is $(\pi^2/6)*\mu^2$ (Van Cranenburgh et al., 2015a).

2.3.2.7 RUM-RRM Hybrid Modelling

One great advantage of the RRM formulation is that it can easily be considered in a RUM-RRM hybrid function. The behavioural interpretation of such modelling is that the researcher assumes that the decision-maker processes some of the attributes by the RUM approach and some other attributes by RRM approach. The overall utility/regret (denoted as V_i) of the hybrid model can be formulated as equation 2.22.

$$V_i = \sum_{m=1..Q} \beta_m . x_{im} - \sum_{j \neq i} \sum_{m=Q+1..M} ln\left(1 + exp\left[\beta_m . \left(x_{jm} - x_{im}\right)\right]\right) \qquad (2.22)$$

Where, Q attributes are implemented outside the regret function, and (M-Q) attributes inside the regret function.

2.3.2.8 Choice Probability in RRM Model

The choice probability of alternative *i* for individual *n* can be written as in equation 2.23. However, for μRRM model, the choice probability will be as per equation 2.24.

$$P_{in} = \frac{e^{-RR_{in}}}{\sum_j e^{-RR_{jn}}} \qquad (2.23)$$

$$P_{in}^{\mu RRM} = \frac{e^{-\mu\, RR_{in}^{\mu RRM}}}{\sum_j e^{-\mu\, RR_{jn}^{\mu RRM}}} \qquad (2.24)$$

2.3.2.9 Accounting for the Variation in Choice Set Size

In many choice situations, the choice set size varies across choice observations. In RRM models, the overall regret level of an alternative equals the sum of all pair-wise regrets arising from bilateral comparisons. This makes the overall regret level rise with the increase of the choice set size. Consequently, variations of choice set size in the dataset show deterministic choice behaviour, as a larger choice set tends to show a more significant difference in regret level. Such variation is unrealistic and hampers the performance of the RRM models in the context of varying choice set size. This problem is addressed in a straightforward but effective way by normalising/scaling the regret level with a correction factor shown in equation 2.25 (Van Cranenburgh et al., 2015b).

$$\tilde{R}_{in} = \frac{\Gamma}{J_n} R_{in} \tag{2.25}$$

Here, the overall regret level is scaled by choice set size, denoted J_n. The numerator Γ depends on the type of RRM model. For some types of RRM models, the choice of Γ is consequential, however for some models like μRRM and P-RRM models it is not. This is because μRRM or the P-RRM have some scale variance while the classical RRM and G-RRM model do not. Consequently, it is not advisable to use the classical RRM or G-RRM models when there is a varying choice set size (Chorus, 2012; Van Cranenburgh et al., 2015b).

2.3.2.10 RRM for Sampling of Alternatives

The discussion in this section is taken mainly from the work of Guevara et al. (2014) in the context of the classical RRM model (shown in equation 2.18). For sampling, let a researcher sample a subset D_n with $\tilde{J}_n$elements (cardinality of$\tilde{J}$) from the true choice set C_n(cardinality of J). For estimation purposes, in D_n the chosen alternative should be included and thus D_n is not independent of *i*. Now, defining $\pi(i|D_n)$ as the conditional probability that the researcher drew the set D_n, the conditional probability of $\pi(i|D_n)$ can be written as equation 2.26.

$$\pi(i|D_n) = \frac{e^{-R_{in} + ln\,\pi(D_n|i)}}{\sum_{j \in D_n} e^{-R_{jn} + ln\,\pi(D_n|j)}} \tag{2.26}$$

Equation 2.23 is similar to equation 2.20, except for the term $\ln \pi(D_n|i)$, which is known as the sampling correction. Again, the summation in the denominator is only over the alternatives in D_n. However, this cannot provide a practical solution for the problem as the argument of R_{in} still depends on the full choice set C_n. Consequently, Guevara et al. (2014) suggest calculating R_{in} by resampling another choice set $\widetilde{D}_n$from C_n. The following equation 2.27 is proposed in this regard where $\hat{R}_{in}$ is suggested as a feasible approximation of R_{in}.

$$\hat{R}_{in} = \sum_{J \in \widetilde{D}_n} w_{jn} \sum_{m=1}^{M} l\,n\big(1 + ex\,p\big[\beta_m \cdot \big(x_{jmn} - x_{imn}\big)\big]\big) \tag{2.27}$$

In equation 2.27, an expansion factor W_{jn} is included and can be estimated from equation 2.28. This expansion factor is similar to the expansion factor suggested in (Guevara and Ben-Akiva, 2013b) where a simple random sampling protocol is used.

$$w_{jn} = \frac{\tilde{n}_{jn}}{E(\tilde{n}_{jn})} = \frac{1}{\tilde{J}/_{J}} = \frac{J}{\tilde{J}} \quad (2.28)$$

Again, the sampling correction can correspond to equation 2.29 (Guevara et al., 2014), where the sampling protocol is as follows. In the first step, the chosen alternative is included in the choice set, and in the second step the non-chosen alternatives are randomly sampled without replacement to make a total of $\tilde{J}$. Again, to get the sampling correction, the method proposed in subsection 2.3.1.7 can also be used.

$$ln\,\pi(D_n|i) = ln(\frac{J-1}{\tilde{J}-1}) \quad (2.29)$$

2.3.2.11 Correction for path Commonalities

Prato (2014) proposed three solutions to address the path commonality issue in the RRM approach. The first solution is to use a utility-based correction factor in the regret function proposed in equation 2.30, which is a RUM-RRM hybrid structure discussed in subsection 2.3.2.7. The second solution is using the correction term within the RRM formulation, which provides a pairwise comparison of the degree of independence of alternatives (equation 2.31). The third solution uses a term which expresses a pairwise correlation between alternatives (equation 2.32).

$$V_i = -\sum_{j \neq i} \sum_m ln\big(1 + exp\big[\beta_m.\big(x_{jm} - x_{im}\big)\big]\big) + \beta_{Corr}\, Corr_i \quad (2.30)$$

$$V_i = -\sum_{j \neq i} \sum_{m} ln\big(1 + exp\big[\beta_m.\big(x_{jm} - x_{im}\big)\big]\big) - \sum_{j \neq i} ln\big(1 + exp\big[\beta_{corr}\big(PS_j - PS_i\big)\big]\big) \quad (2.31)$$

$$V_i = -\sum_{j \neq i} \sum_m ln\big(1 + exp\big[\beta_m.\big(x_{jm} - x_{im}\big)\big]\big) - \sum_{j \neq i} ln\big(1 + exp\big[\beta_{corr}\, \rho\big(\varepsilon_i, \varepsilon_j\big)\big]\big) \quad (2.32)$$

Where, *$corr_i$* is the utility-based correction term for route *I,*

β_{corr} is the correction parameter to be estimated,

PS_i is the path size of route *i*,

$\rho(\varepsilon_i,\ \varepsilon_j)$ is the correlation between two routes *i* and *j* can be expressed as follows

$$\rho(\varepsilon_i, \varepsilon_j) = \frac{L_{ij}}{\sqrt{L_i L_j}} \quad OR \ \rho(\varepsilon_i, \varepsilon_j) = \frac{L_{ij}}{\sqrt{L_i L_j}}\left(\frac{L_i - L_{ij}}{L_j - L_{ij}}\right)$$

L_i is the length of route *i*,

L_{ij} is the length of the common part between route *i* and *j.*

Prato (2014) proposed two types of correction factors: (i) a commonality factor which decreases the utility of a route because of its degree of similarity with the alternative routes (equation 2.5 and 2.6) (Cascetta et al., 1996; Cascetta and Papola, 2001); and (ii) a path size measure (PS in equation 2.7 and PSC in equation 2.8) which indicate the fraction of a route that constitutes a ‘‘full’’ alternative (Ben-Akiva and Bierlaire, 1999; Bovy et al., 2008). This research uses the path size correction factor presented in equation 2.9 (discussed in subsection 2.3.1.5).

2.4 Model Evaluation Measures

The results of the modelling exercises are discussed in this segment, which are presented in Chapters 3 to 6. Key findings are presented with the help of figures and tables. The recommendations are carried out on the light of the key findings. The following evaluation measures were considered in this study.

2.4.1 Goodness of fit measures

The adjusted ρ^2 measure is presented as the goodness of fit of the models. However, for model selection, the Bayesian Information Criteria (BIC), which includes a penalty term for the number of the parameters is used (see equation 2.33). The model with a lower BIC value is preferred.

$$BIC = ln(n) * k - 2 * ln(\hat{L}) \tag{2.33}$$

Where, $\hat{L}$ is the maximum value of the likelihood function
n is the sample size
k is the number of estimated parameters

2.4.2 Elasticity and Marginal Effect of the parameters

The stop choice and route choice models developed in this study are unlabelled. Therefore, elasticity measures or marginal effect measures cannot be performed. However, a *pseudo direct marginal effect* of the parameters is calculated from the following definition (equation 2.34).

> *"Measures the absolute change in the probability of choosing the chosen alternative in the choice set concerning a unit change in an attribute of that same alternative".*

$$\widehat{M}_{x_{ink}}^{P(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} \tag{2.34}$$

Where, $\widehat{M}_{x_{ink}}^{P(i)}$ is the pseudo direct marginal effect of the chosen alternative for attribute *k and* sample *n,*
$P_n(i)$ is the probability of the chosen alternative for the sample *n,* and
x_{ink} is the associated vector of the attribute of the chosen alternative.

The probability weighted sample enumeration (PWSE) method suggested by Hensher et al. (2005) is applied to the aggregate *pseudo direct marginal effect* as per equation 2.35.

$$\widehat{M}_{x_{ink}}^{\overline{P(i)}} = \frac{\sum_{n=1}^{N} \hat{P}_{in}(i)\ \widehat{M}_{x_{ink}}^{P(i)}}{\sum_{n=1}^{N} \hat{P}_{in}(i)} \tag{2.35}$$

Where, $\overline{P(\iota)}$ is the aggregated choice probability of the chosen alternative

$\hat{P}_{in}(i)$ is an estimated choice probability

Again, for the dummy variables, the change in the probability is calculated from the differences between probability when $x_{ink} = 0$ and probability when $x_{ink} = 1$. This method is stated in Washington et al. (2010) and applied by Chang and Mannering (1999) and Ulfarsson and Mannering (2004) in the context of elasticity calculation.

2.4.3 Route Detection and Prediction Capability Measure

The detection of the chosen route from the generated routes was calculated from the coverage measure proposed by Ramming (2002). The coverage measure is expressed in equation 2.36 (Anderson et al. 2017).

$$Coverage\ (\delta) = \frac{\sum_{n=1}^{N} I\left(O_{max(n)} \geq \delta\right)}{N} \qquad (2.36)$$

Where, $O_{max(n)}$ is the best overlap measure (overlap in length as the unit of measure) for observation n calculated from equation 2.37,

δ is the overlap threshold

N is the number of observation

$I\ (.)$ is an indicator equal to 1 when the criterion is fulfilled and 0 otherwise.

The overlap of the routes was measured from the following equation as suggested by Ramming (2002).

$$O_{nr} = \frac{L_{nr}}{L_n} \qquad (2.37)$$

Where, O_{nr} is the overlap for observation n and route r,

L_{nr} is the sum of the length of the overlapping elements at the line level

L_n is the length of the observed path

This coverage measure was not used for prediction purposes. However, to determine the prediction capability of the models. the overlap threshold was considered as 100%.

For prediction following measures are considered in this thesis.

- Rank: The choice probability of the alternatives are calculated and ranked. The rank of the chosen alternatives are evaluated as follows
    - Within the top 5
    - Within the top 10
- Hit Rate: The percentage of the highest-ranked chosen alternative is termed as "Hit Rate".

2.5 Findings and Limitations of this Chapter

This chapter discussed the methodology of this study with discussions from the existing literature. The study chooses to use a trip based shortest path (TBSP) algorithm, which finds the shortest path according to the earliest arrival to the destination. Some of the advantages of this algorithm include the ability to use precise origin-destination information to generate transit paths, use of a trip based network configuration which improves the performance of trip generation and use of a transfer stop hierarchy which eliminates unrealistic paths.

Household Travel Survey data of Southeast Queensland, Australia is used to study the transit choice behaviours. Choice sets of stop choice and route choice for 1237 respondents were generated using the TBSP algorithm. Different modelling structures used in the discrete choice modelling were also discussed.

The limitations of this chapter are as follows:

- There were many algorithms for transit path generation available in the literature. As the scope of work is not to compare the methods, this study considers only one, which is the TBSP algorithm developed by Khani (2013), and Khani et al. (2014; 2012) and later modified by Nassir et al. (2015a; 2015b).

- Detailed analysis of coverage testing for the path generation algorithm was not performed, as this was out of the scope of the study. However, necessary information on coverage was extracted and provided in the thesis.

3 MODELLING TRANSIT USERS ACCESS STOP CHOICE BEHAVIOUR

This chapter aims to study transit stop choice behaviour with a focus on how people strategise their choices. It is hypothesised that travellers treat stops differently based on various strategies. This Chapter, in particular, consider strategies including minimising travel time, access time and the number of transfers. The effectiveness of several discrete choice model specifications is examined in this chapter. The findings confirm that transit users indeed strategise while choosing transit access stops. Moreover, route and stop attributes have a significant impact on stop selection. Furthermore, users' socioeconomic characteristics along with trip timing play essential roles in choosing transit stops. The outcomes of this analysis would facilitate the recent move towards the development of behavioural route choice models using transit fare card data, which can then assist travel demand estimation models with a focus on public transport.

3.1 Introduction

In the transit demand modelling literature, two areas have been distinctly discussed: 1) transit mode choice (or even general transit ridership) and 2) transit assignment and route choice. Recently, researchers have started using transit fare card (from now on smart card) data to develop transit route choice models (Jánošíková et al., 2014; Schmöcker et al., 2013; Tan, 2016). As smart card datasets can detect repeated observations, route identification and estimation becomes more manageable. By using a transit fare card dataset, Schmöcker et al. (2013) proposed a bi-level discrete choice model where the upper level considers the choice preference of users while the lower level deals with the deterministic probabilities of boarding routes. However, as smart card datasets usually lack information about the actual origin and destination, these models can only determine route choice from the departure stop. Consequently, these models miss the link between the trip origin and the departure transit stop.

This gap was addressed by Nassir et al. (2015b) by developing a transit stop choice model. In their study, Nassir et al. (2015b) assumed that transit users select their route by selecting a stop (bus stop, train station or ferry terminal) from a desirable choice set. They argued that modelling the route choice behaviour at the stop level is more appropriate as the observed data is consistent with the choice made by the users. They proposed a nested structure where an acceptable model fit is gained by considering a bi-level train and no-train nesting structure. Moreover, the study found that the choice of stop does not only depend on the attributes of the routes (including fastest travel time and the number of transfers), but also the attributes of the stops. They showed that the presence of shelter at stops, the walk time from the origin location to the stop, travel time, number of transfers and number of routes significantly affect the choice of stops. These findings certainly add to the body of knowledge on the behavioural aspect of transit mode choice. However, their work cannot be treated as a comprehensive stop choice study due to three significant shortcomings. Firstly, they did not consider users' socioeconomic and demographic characteristics and the taste heterogeneity. Secondly, attributes related to the trip and strategy were also not considered. Thirdly, the model specification used was quite limited and restricting.

Other stop choice studies are also found in the literature, but they have focused on other issues. Debrezion et al. (2009) presented a railway station choice model for Dutch railway users to determine a measure of station accessibility. They proposed a nested logit model where access modes were modelled at the upper level, and stations were modelled at the lower level. They found that access distance had a negative effect on the accessibility indicator, while parking availability, the frequency of public transport and railway station quality had a positive effect on station choice. Chakour and Eluru (2013) modelled access modes and station choice in a different approach. This study found that a latent segmentation technique delivered better results than the nested logit approach proposed by Debrezion et al. (2009). Mahmoud et al. (2014) investigated the choice of park and ride stations for cross-regional commuter trips in the Greater Toronto and Hamilton area. The study aimed at finding aspects relevant to the design of more sustainable and attractive transit stations. They developed several multinomial logit models by using data on parking facilities, surrounding land use and station amenities. Tan (2016) presented a comprehensive study on transit route choice, which included multimodal access and egress behaviour. The author used smart card data of Singapore and presented different model specifications, including PSC Logit and Latent Class models.

The work presented in this chapter aimed to develop stop choice models by addressing the shortcomings of the model developed by Nassir et al. (2015) by introducing a strategy-based specification. Again, the study aims to model the users' preference heterogeneity by introducing different socio-demographic variables in the modelling structure. Again, this chapter aims to capture the taste heterogeneity for different attributes.

3.2 Modelling Stop Choice Strategy

The idea of considering strategies in the stop choice modelling is derived from the literature (Fonzone and Bell, 2010; Fonzone et al., 2010; Kurauchi et al., 2012; Nassir et al., 2015b). Nassir et al. (2015b) showed that transit users tend to choose stops, which minimise travel time, minimise access time and minimise the number of transfers. Kurauchi et al. (2012) found that London Oyster Card users might use different strategies for their regular commute as they do not use fixed routes. Fonzone and Bell (2010) and Fonzone et al. (2010) also reported similar findings.

In this study, it is assumed that when a transit user wants to make a trip, he/she decides what type of travel strategy is suitable for his/her current situation (see Figure 3.1). In this study, three basic strategies are considered: minimising the time of travel (MTT), minimising access time (MAT) to reach the boarding stop and minimising the number of transfers (MTR). Combinations of these three primary strategies (four combinations) are also considered. It is assumed that users choose the alternative (access stop) that best matches their desired strategy and maximises their utility. For example, if a user wants to minimise travel time (an MTT user), he/she chooses an alternative that falls under the MTT strategy. Similarly, a MAT-MTR user chooses a stop that takes less time to access and has the most direct connection to the destination (MAT-MTR strategy).

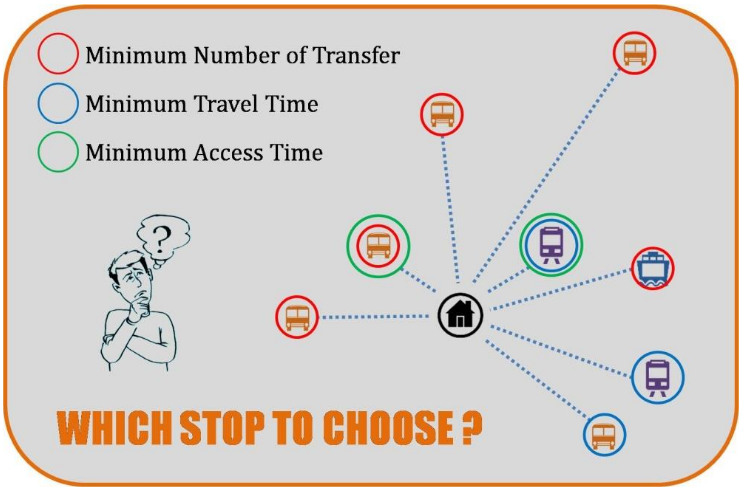

Figure 3.1 Understanding Strategies of Stop Choice

3.2.1 Visualising Stop Choice Strategy from the HTS Data

Figure 3.2 shows that most travellers choose access stops that contain some strategies. In the three unique strategy situations, MTT and MTR strategies seem to be more common (63% of users choose MTT, and 72% choose MTR) than MAT strategies (only 49% of users choose MAT strategies). If there are multiple strategies, users seem to prefer combined strategies rather than single strategies or none. For example, in TT-AT and TT-TR, the share of combined strategies are dominant (MTT&MAT=37%, MTT&MTR=49%) compared to single strategies or none. Contrastingly, in the AT-TR combination, the share of MTR (38%) is more than the combined strategies of MAT&MTR (34%). Finally, in the TT-AT-TR combination, users seem to prefer combined strategies. Very few (8%) users seem to have no preference for strategies.

3.2.2 Stop Choice Strategy as Explanatory Variables

In this study, three basic strategies (MTT, MAT and MTR) are used as dummy variables and considered only in the MNL and Mixed MNL models.

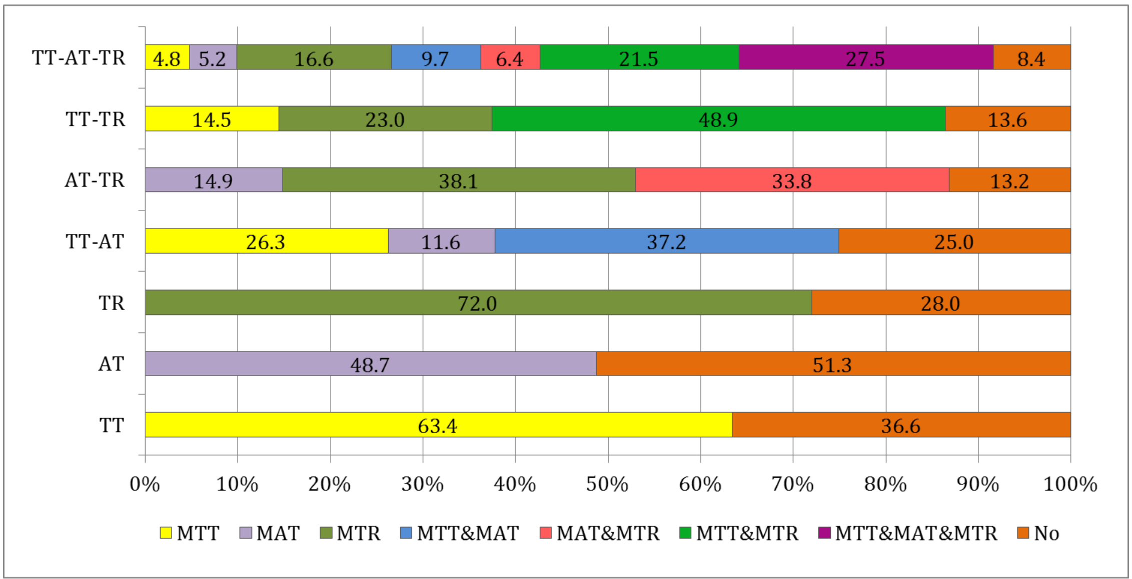

Figure 3.2 Users Preference for Strategies

3.2.3 Stop Choice Strategy as Nest Structures

This study considers seven nesting structures (see Table 3.1). These nesting structures are also used in the mixed NL models. Therefore, each group consists of two models: NL and mixed NL.

In

Table 3.1, the first nesting group is for the MTT strategy. This group consists of two nests: (1) stops that are fastest (fastest routes from the stop) are grouped in the MTT nest, and (2) the rest of the stops are grouped in the NoMTT nest. The next two groups consider the MAT and MTR strategies, respectively, similar to the first nesting group. The next three groups (4, 5 and 6) couple two strategies together. For example, in the fourth structure, both MTT and MAT are coupled. Here, there are four probable combinations of these two strategies: (1) minimising travel time only (MTT), (2) minimising access time only (MAT), (3) considering both (MTT and

MAT) and (4) considering none of them (None). The last structure considers all three strategies, with all the probable combinations (eight nests). Figure 3.3 is presented to explains the nest structure.

Table 3.1 Nest structures for the proposed NL and mixed NL models

Group	Model Name	Number of Nests	Nest Structure
1	TT, TT[M]*	2	MTT, NoMTT
2	AT, AT[M]	2	MAT, NoMAT
3	TR, TR[M]	2	MTR, NoMTR
4	TT-AT, TT-AT[M]	4	MTT, MAT, MTT&MAT, None
5	TT-TR, TT-TR[M]	4	MTT, MTR, MTT&MTR, None
6	AT-TR, AT-TR[M]	4	MAT, MTR, MAT&MTR, None
7	TT-AT-TR, TT-AT-TR[M]	8	MTT, MAT, MTR, MTT&MAT, MTT&MTR, MAT&MTR, MTT&MAT&MTR, None

** [M] means mixed NL model*

Option	Strategy Attributes			TT Model		TT-AT			
	MTT	MAT	MTR	MTT	NoMTT	MTT	MAT	MTT&MAT	None
				IF MTT=1, then 1, otherwise 0	IF MTT=0, then 1, otherwise 0	IF MTT=1 AND MAT=0, then 1, otherwise 0	IF MTT=0 AND MAT=1, then 1, otherwise 0	IF MTT=1 AND MAT=1, then 1, otherwise 0	IF MTT=0 AND MAT=0, then 1, otherwise 0
P_1	1	1	1	1	0	0	0	1	0
P_2	1	1	1	1	0	0	0	1	0
P_3	0	1	1	0	1	0	1	0	0
P_4	1	0	1	1	0	1	0	0	0
P_5	0	0	1	0	1	0	0	0	1
P_6	0	0	1	0	1	0	0	0	1
P_7	1	1	0	1	0	0	0	1	0
P_8	1	0	0	1	0	1	0	0	0
P_9	1	0	0	1	0	1	0	0	0
P_{10}	0	0	0	0	1	0	0	0	1

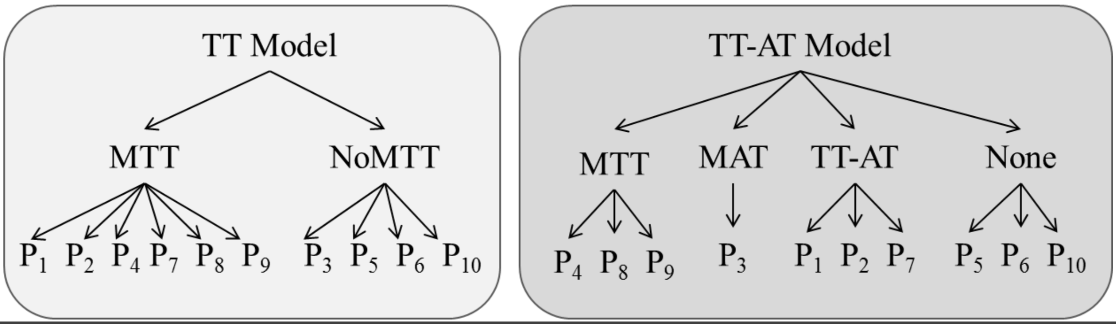

Figure 3.3 Illustrative example of Nesting Structure

Figure 3.3 shows a case where a passenger has ten alternatives, and each alternative has its strategy attributes. For example, P1 is an alternative, which satisfies all three strategies. In the TT model, this alternative is considered in the MTT nest. However, in the TT-AT model, this alternative falls under the TT-AT nest as this alternative satisfies both the strategy attributes (MTT and MAT). Again, P3 is not an MTT alternative; therefore, in the TT model, it falls in the NoMTT nest. However, in the TT-AT model, P3 represents the MAT nest.

3.3 Model Results and Discussions

The models were estimated using the discrete choice estimation package BIOGEME (Bierlaire, 2003). The stop choice model estimation results are presented in **Appendix C**.

3.3.1 Model Comparison

Table 3.2 provides a comparison between the models (MNL, NL, mixed MNL and Mixed NL) estimated in this study. Most of the models show similar goodness of fit (adjusted ρ^2) values. TT-TR[M] model show the best goodness of fit of 0.327. However, two of the nest coefficients of this model (MTT and None) are found to be insignificant (at 80% confidence interval). Among the fourteen nested models, only four models including TT, AT, TT-AT and AT[M] yield significant nest coefficients (at 80% confidence interval) for all the nests. The BIC values of these four nested models, along with the MNL and Mixed MNL models are compared to find the best model. MNL model showed the best BIC value of 4003.96 followed by the TT model (BIC = 4033.63) and Mixed MNL model (BIC =4069.31).

Table 3.2 Comparisons of Models*

	MNL	Nested Logit Models						
		TT	**AT**	**TR**	**TT-AT**	**TT-TR**	**AT-TR**	**TT-AT-TR**
No. of parameters	7	9	10	13	13	19	17	24
Final log-likelihood	-1977.056	-1984.771	-2054.072	-1971.252	-1997.259	-1942.09	-2013.01	-1966.695
Likelihood ratio test	1921.744	1906.315	1767.713	1933.353	1881.338	1991.676	1849.836	1942.263
Adjusted ρ^2	0.325	0.321	0.297	0.325	0.32	0.32	0.309	0.322
BIC	4003.96	4033.63	4179.36	4035.08	4087.09	4019.48	4147.08	4104.30
Nest Coefficients (λ)	Not Applicable	MTT=0.81 NoMTT=0.83	MAT=0.78 NoMAT=0.75	MTR=0.78 NoMTR=0.95 (0.53)	MTT=0.699 MAT=0.8 MTT&MAT= 0.588 None=0.781	MTT=0.84 MTR=0.66 MTT&MTR= 0.72 None=0.93 (0.84)	MAT=1 (0) MTR=0.61 MAT&MTR= 1 (0.01) None=0.91	MTT=0.72 MAT=0.97 (0.21) MTR=0.56 MTT&MAT=0.85 (0.72) MTT&MTR=0.41 MAT&MTR=0.7 (0.92) MTT&MAT&MTR=1(0) None=0.89 (1.35)
	Mixed MNL	**Mixed Nested Logit Models**						
		TT[M]	**AT[M]**	**TR[M]**	**TT-AT[M]**	**TT-TR[M]**	**AT-TR[M]**	**TT-AT-TR[M]**
No. of parameters	9	13	12	13	16	20	20	24
Final log-likelihood	-2002.61	-2002.2	-1994.5	-1990.11	-2006.316	-1968.77	-1978.566	-1956.693
Likelihood ratio test	1870.635	1905.09	1926.79	1955.64	1923.224	2018.315	1998.724	2042.266
Adjusted ρ2	0.315	0.315	0.318	0.32	0.314	0.327	0.325	0.328
BIC	4069.31	4097.06	4074.52	4072.79	4126.57	4079.97	4099.56	4084.3
Nest Coefficients (λ)	Not Applicable	MTT=0.89 (1.2) NoMTT=0.91 (1.24)	MAT=0.78 NoMAT=0.76	MTR=0.81 NoMTR=1 (0.01)	MTT=0.74 MAT=0.79 MTT&MAT= 0.63 None=0.81	MTT=0.85 (1.23) MTR=0.65 MTT&MTR= 0.75 None=1 (0.06)	MAT=0.96 (0.3) MTR=0.66 MAT&MTR= 0.93 (0.33) None=0.95 (0.58)	MTT=0.73 (1.27) MAT=1 (.01) MTR=0.63 MTT&MAT=0.74 (0.97) MTT&MTR=0.54 AT TR=0.3 (0.97) MTT&MAT&MTR=1 None=0.94 (0.8)

**t-test value (for the hypothesis, H_0 λ=0) are provided in parenthesis for the nest coefficients that are not significant at 80% confidence level*

3.3.2 Behavioural Interpretation

The estimated parameters of the best three models MNL, TT, and Mixed MNL models are presented in Table 3.3. Two direct impedance attributes, including *Number of Transfers* (in TT model), and *Other Walking Time,* are found to be significant. The signs of these coefficients are negative as expected, which indicates that transit users prefer to start their trip from a stop that had a more direct connection to their destination and involved less walking. One of the aggregate impedance attributes, *Number of Routes*, is found to be significant in the models, which means that transit users tend to choose access stops that have multiple route options. Facility attributes like *Walking to Stop Time,* and *Stop Lighting* are also found to be significant. The negative sign of *Walking to Stop Time* means users perceive more disutility if they have to walk more to the access stop. The positive sign of the *Stop Lighting* attribute implies that users prefer to choose stops that have lighting arrangement. The sign of the coefficient of *Train* is positive which means that transit users in SEQ are much more willing to travel by train than by other modes.

Some of the socio-economic attributes of travellers are also found to be significant, as reported in Table 3.3, and Appendix C. The TT models show that Australian born users are more likely to select MTT strategies for choosing transit stops. The other significant socio-demographics include household size, number of bedrooms, flat owner, medium-income group, bike license holder, vehicle owner, student and gender are found to be significant in other models not reported in Table 3.3 (see Appendix C). These socio-demographics depict users preference for accessing stops by differing the strategy selection. For example, in TT-AT[M] model, female students are more likely to use a combination of MTT and MAT strategies while choosing their preferred transit stop. Again, users from larger households and users living in flats tend to prefer the combination of MTT and MAT strategies when choosing transit stops.

Table 3.3 Detailed Estimation Results of the Best Models

Explanatory Variables (β)	**MNL Model**	**NL Model (TT)**	**Mixed MNL Model**
Number of Transfer	-	-0.867 (-11.15)	-
Other Walking Time	-0.034 (-3.19)	-0.0304 (-3.3)	-0.064 (-5.85) **
Number of Routes	0.059 (4.25)	0.0528 (3.85)	0.0582 (3.9)
Walking to Stop Time	-0.196 (-22.39)	-0.134 (-9.58)	-0.2115 (-30.8)**
Stop Lighting	0.393 (4.09)	0.307 (3.63)	0.537 (5.79)
Train	2.32 (18.75)	1.96 (13.16)	1.36 (15.53)
MTT Strategy	0.656 (7.58)	N/A	0.713 (7.84)
MTR Strategy	1.4 (15.34)	N/A	1.43 (13.84)
Australian Born (MTT Strategy)	-	0.695 (6.34)	-
Standard Deviation			
Other Walking Time	-	-	0.2844
Walking to Stop Time	-	-	0.086
Nest Coefficients (λ)*			
TT	-	0.81 (2.15)	
NoTT	-	0.83 (2.65)	

*Robust t-test values are within the bracket * Robust t-test is estimated for the hypothesis, H_0 λ=0   ** mean value*

Some trip attributes are also found to be significant. The Trip attribute *PM Peak Departure* is found to be significant in some models. In TT-AT[M] model, this signifies that users making a trip other than at the PM peak hour are inclined to follow the combined strategy of MTT and MAT when choosing their transit stop. Another finding is that two strategy attributes (MTT and MTR) are found to be significant in the MNL and Mixed MNL models, suggesting that travellers choose stops with minimum travel time and minimum number of transfers.

The mixed MNL and Mixed NL models assume that the users have taste variations for different travel time components and other variables. However, the results show that the SEQ users have taste heterogeneity for three variables, including access walk time, other walk time and number of transfers (in TT[M] and AT[M] models, see Appendix C). In the Mixed models, all these variables are assumed to follow a lognormal distribution. The results prove that the assumption of heterogeneous behaviour is correct as both the parameters and the variances are found to be significant at a 99% confidence level. *Other walk time* also found to be significant in all the other mixed MNL models accept AT[M]. However, the variance of access walk time variable was not found to be significant in other models at 95% confidence level. The variance of *number of transfer* found to be significant in TT[M] and AT[M] models.

The rate of substitution of the attributes for MNL and TT models is presented in Table 3.4. These models identify that users consider every minute of walking to the access stop to be 5 minutes (TT model) to 5.75 minutes (MNL model) of other types of walking (including walking for transfers and walking to the destination) involved in the travel path. However, in the mixed MNL model, this estimate is between 1.67 minutes to 5 minutes. The results are in alignment with the other stop/route choice studies in the literature (Anderson et al., 2017; Bovy and Hoogendoorn-Lanser, 2005; Fosgerau et al., 2007a, b; Nassir et al., 2015b; Raveau et al., 2011).

Again, an additional route in an access stop seems to pose the same utility as saving 2 to 5 minutes of access walk (aligned with Nassir et al. 2015b). Moreover, one additional transfer can pose the same disutility as 6.5 minutes of access walk time. However, access stop posing the minimum number of transfers (MTR strategy) can yield the same utility as saving 4.56 to 12.21 minutes of access walking time. Furthermore, a train station can yield similar utility of 11.8 minutes (MNL model) to 14.6 minutes (TT model) of access walk time-saving. However, the mixed MNL model shows a range of 4.34 and 11.62, which means SEQ passengers may consider walking extra to access a stop which is a train station or pose MTR/MTT strategy.

Table 3.4 Rate of Substitution of the Stop Choice Models

Explanatory Variables	MNL	TT	Mixed MNL *	Nassir et al. (2015b)
Walking to Stop Time	1	1	1 (1, 1)	1
Number of Transfer	-	6.5	-	4.4
Other Walking Time	0.17	0.2	0.33 (0.22, 0.6)	-
Number of Routes	-0.30	-0.39	-0.27 (-0.5, -0.19)	-0.34
Stop Lighting	-2.0	-2.3	-2.5 (-4.6, -1.72)	-
Train	-11.8	-14.6	-6.31 (-11.62, -4.34)	-
MTT Strategy	-3.3	-	-3.31 (-6.09, -2.27)	-
MTR Strategy	-7.1	-	-6.64 (-12.21, -4.56)	-
Australian Born (MTT Strategy)	-	-5.2	-	-

** average (minimum, maximum)*

3.3.3 Model Prediction Capability and Sensitivity

The choice probabilities of all the options are calculated for the MNL and TT model. It is found that the models could correctly predict the users' chosen alternatives in 49.9% (MNL) and 49.7% (TT) of cases. It could also be interpreted that, according to the MNL model, 49.9% of users choose the stop with the highest probability. Again, 86% of users (MNL model) seem to choose the access stop from a set of 5 stops with the highest probabilities; for the TT model, this is about 84.7%. The prediction capabilities of these models are shown in Figure 3.4 and Figure 3.5. Figure 3.4 presents cumulative percentages according to the rank of the choice probability of the chosen option.

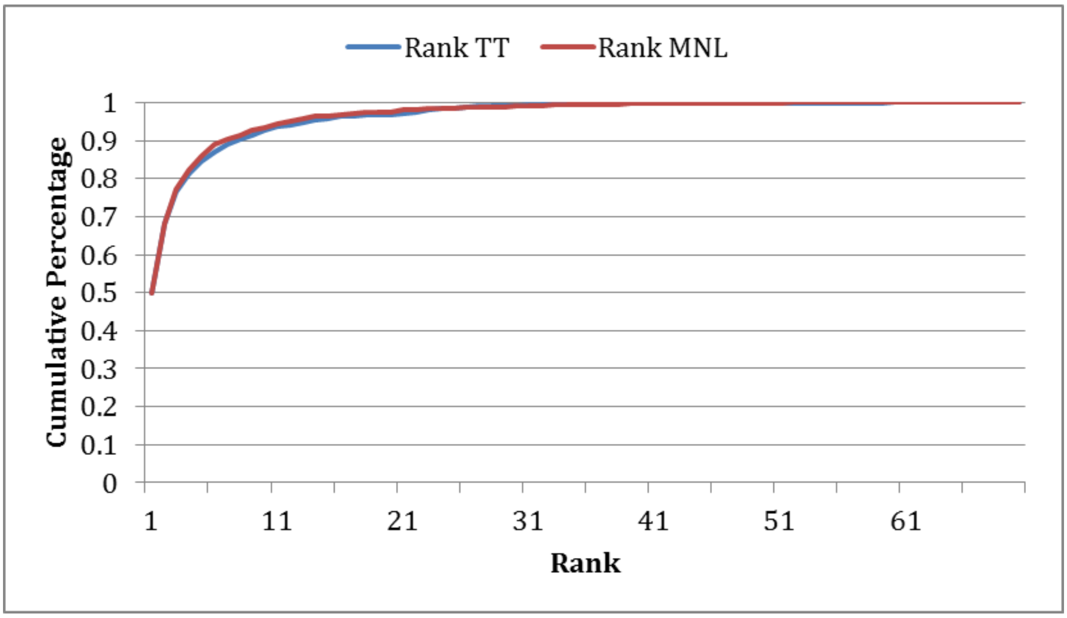

Figure 3.4 Rank of the Chosen Options in the Stop Choice Model

Figure 3.5 presents the cumulative percentage of successful prediction, with an increasing pattern for the number of options considered to include the selected option. In other words, if a *set of predicted options* is considered to include the observed option, the chance of the observed option occurring increases. Obviously, as the choice set size increases, the chance of including the observed option in the *set of predicted options* decreases. In Figure 3.5, five curves are shown representing the prediction capabilities for the observed choice in the set of predicted options where the highest probability is for curve 5. This shows that the models can predict the choices better if the choice set size is relatively small and vice versa. However, when the choice set size is larger than 40, the prediction is uncertain.

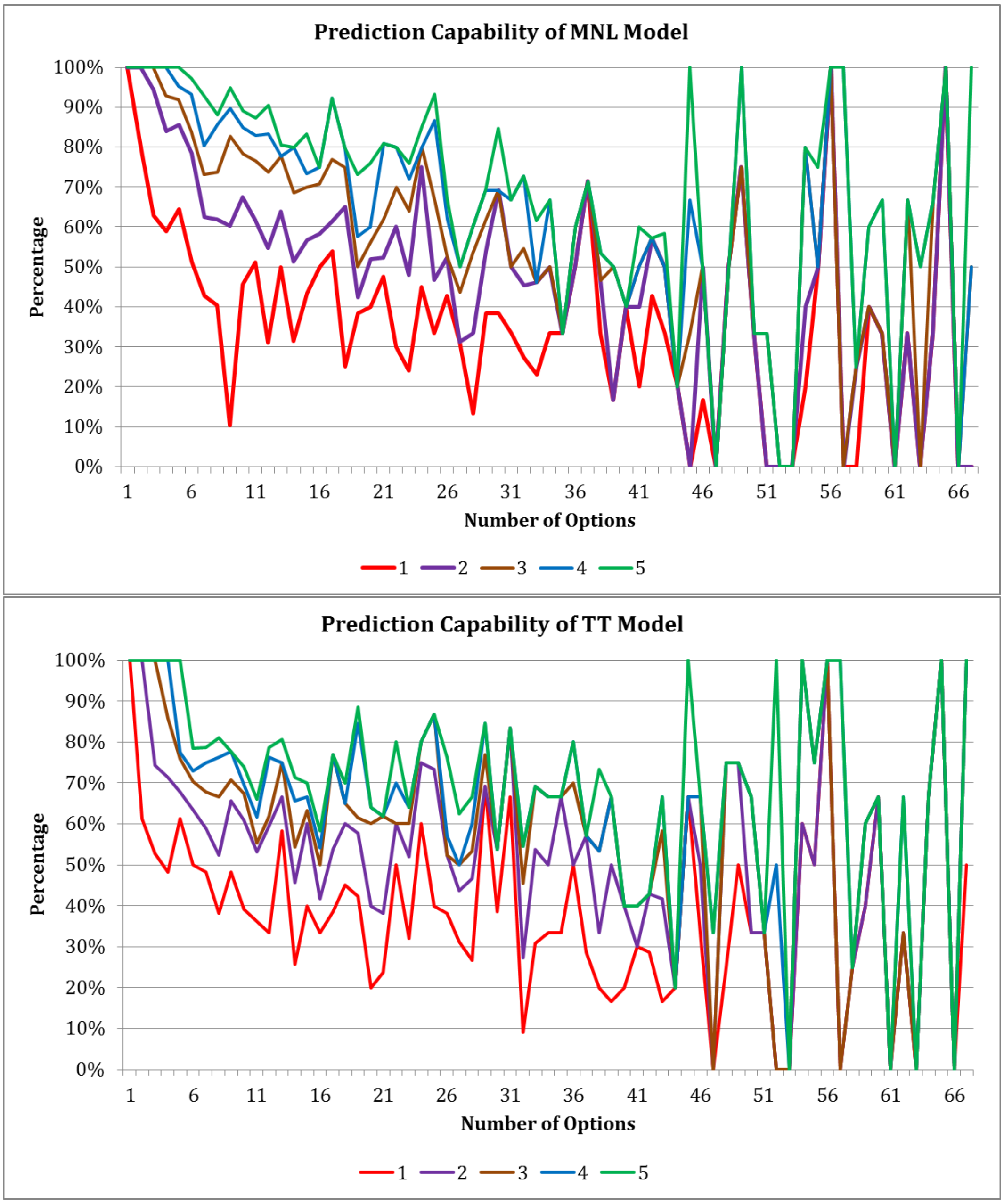

Figure 3.5 Prediction Capabilities of the Stop Choice Models

The pseudo direct marginal effects (discussed earlier in subsection 2.4.2 of the thesis) of the variables of the stop choice model are presented in Table 3.5. The effects are similar among the MNL and mixed MNL models. However, the TT model shows some differences in *stop lighting* and *train* attributes. The variable *train* displays the highest effect, followed by the *MTR Strategy*. The results of the MNL model show that if the chosen alternative changes from the train station to other stops or vice versa, the choice probability of the chosen alternative will change by 31.3%. However, in the TT model this is about 22.9%, much lower than the MNL model. Again, if the status of the chosen alternative changes from transfer minimiser (*MTR Strategy*), or vice versa, the choice probability changes by 20.5% (MNL Model). However, an increase or decrease in a *transfer* from the chosen alternative (in TT model) can change the choice probability by 14.5%. Furthermore, a minute change in *walking to stop time* and *other walking time* for the chosen alternative changes the choice probability by 2.5% and 0.5% respectively.

Table 3.5 Pseudo Direct Marginal Effect of the Stop Choice Models

Explanatory Variables	**MNL (%)**	**TT (%)**	**Mixed MNL (%)**
Walking to Stop Time	2.5	2.8	2.9
Number of Transfer	-	14.5	-
Other Walking Time	0.5	0.5	0.5
Number of Routes	0.9	0.9	0.9
Stop Lighting	6.0	4.8	6.2
Train	31.3	22.9	28.1
MTT Strategy	9.9	-	9.2
MTR Strategy	20.5	-	20.2
Australian Born (MTT Strategy)	-	6.8	-

The TT model is tested to observe the sensitivity of the nests with a change of access walk time and other walking time. The results are shown in Figure 3.6, where the effect of access walk time seems to be higher than the effect of other walking time. Figure 3.6 shows that by increasing the access time and other walking time, the probability of choosing from the *NoMTT* nest increases. This can be interpreted as the increase of access walk time or other walking time, the probability of selecting a stop that follows the MTT strategy will be decreased, and vice versa.

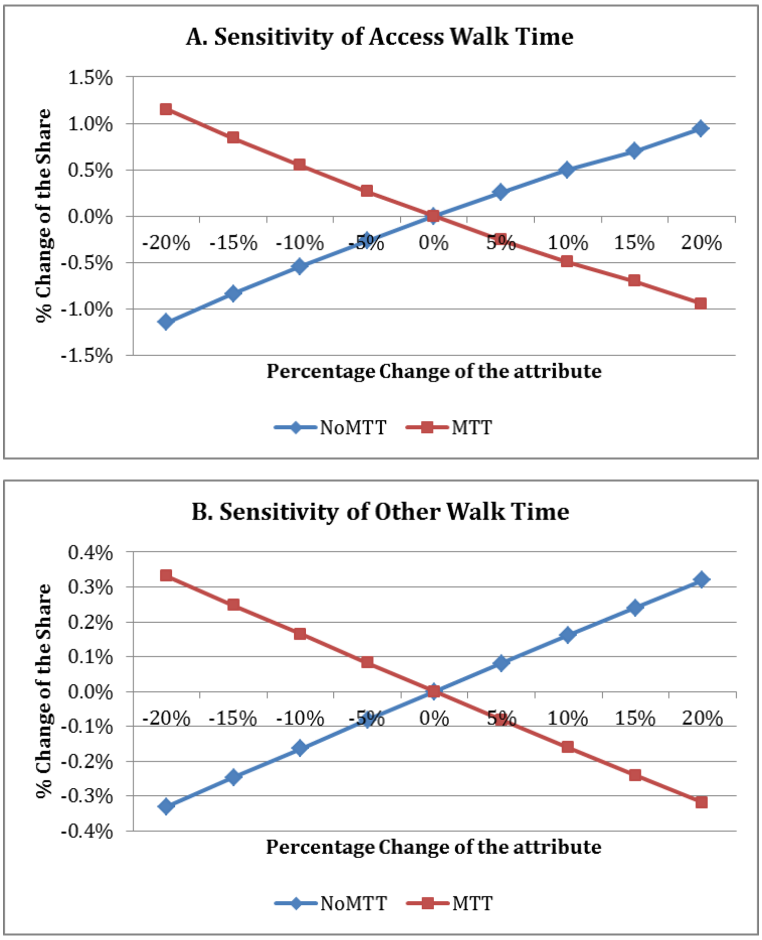

Figure 3.6 Effects of Different Variables on the Nests

3.4 Conclusions

This study provides a deeper understanding of stop choice behaviour compared to the existing literature. The study explored different modelling specifications ranging from a basic multinomial logit to a complex mixed nested logit specification. However, the basic specification of MNL seems to yield better results than the other structures. One of the contributions of this study is to examine the effects of the sociodemographic, trip and strategy variables while choosing an access stop. Especially, this study considered different nesting structures based on strategies rather than transit mode (as presented in Nassir et al. 2015b). Incorporating the strategy attributes to the model specifications is a unique contribution of this study. Consequently, the explanatory power of the presented models in this study (adjusted ρ^2: 0.325 in the MNL model) outperforms the models developed by Nassir et al. (2015b) (adjusted ρ^2: 0.287).

The findings of this chapter can be useful to transit assignment models as route choice from an access stop can be modelled more accurately with the help of smart card data. Integrating the access stop choice model with the route choice models is recently demonstrated by Nassir et al. (2018). Again, path generation algorithms can also benefit from the outcome of this study as this study demonstrates transit users preference criteria to select a transit stop, particularly the importance of different strategies. Moreover, this study also adds knowledge towards public transport accessibility literature, especially the influence of the users' socio-demographic attributes while selecting transit.

It is found that transit users can use different travel strategies while selecting access stops. The most critical strategies seem to be minimising travel time and minimising access time. From the behavioural point of view, it can be concluded that South East Queensland transit users prefer the stop alternatives that are either faster (MTT nest) or more easily accessible from the origin of the trip (MAT nest) or both (fast and nearby). The study results confirm that the SEQ transit users perceive different travel time component differently. Furthermore, SEQ users demonstrate taste variations for access walk time, other walk time and number of transfers.

This study shows that the choice of access stop is not only affected by impedance factors of the routes (number of transfers, walking time, travel time) but also affected by the attributes of the stop (like walking time to access the stop and the presence of lighting at the stop). Moreover, the presence of multiple routes from a stop shows a positive influence on the utility of stop choices. Again, some socioeconomic attributes like gender, student, place of birth, household size, and dwelling type (flat) affect the choice of the stop. Furthermore, transit users also take into account the transit mode and time of the day of the trip. One interesting point is that the developed models relate some of the impedance factors associated with routes linked to the origin and destination stops. These impedance attributes reflect the characteristics of the bundle of routes from the access stop. Again, the nested logit models developed logsum of the stop choice probability, which is grouped according to the strategies developed from the attributes of the route bundle. Therefore, the proposed approach of this thesis integrates the stop and route selection themes in a straightforward manner. However, further analysis is required to examine the opposite direction when stop attributes are included in a route choice model. This work is presented in the next chapter (Chapter 4).

The limitations of this chapter are as follows:

- This chapter studies the SEQ transit users stop choice behaviour considering walking as the only access mode. However, to get a comprehensive understanding of the stop choice behaviour, the other modes of access, including cycling and auto, can also be considered. However, this can be a good topic for future study.
- This chapter provided a deep understanding of the stop choice behaviour of the SEQ transit users by developing different model specifications, including MNL, NL, mixed MNL and mixed NL. However, as this study found an influence of socio-demographics, other specifications like the latent class model can be a good future study to capture the preference heterogeneity. Again, joint modelling of stop-route and stop-strategy and cross-nested models can also be a good topic for future study.

4 CONSIDERATION OF DIFFERENT TRAVEL STRATEGIES IN TRANSIT ROUTE CHOICE MODELLING

Route choice modelling is typically conducted by considering a subset of routes, not the universal set of all feasible routes. In this study, three travel strategies, including minimising travel time, minimising the access walk time and minimising the number of transfers are considered for sampling the generated route alternatives. For sampling the alternatives, this study develops a probabilistic importance sampling protocol where fuzzy logic is used to calculate the probability of sampling. The study uses route, stop, mode, trip, user and path size correction attributes to develop different discrete choice models. Multinomial logit and nested logit structures are tested and compared. The study also discusses the effect of choice set size on model results. The household travel survey data of South-east Queensland, Australia is used to develop these models.

4.1 Introduction

Public transportation route choice models are critical for forecasting transit travel demand and flows. Understanding transit users' route choice behaviour also helps the authorities to develop user-oriented transit facilities. However, compared to the route choice studies conducted on car drivers, public transportation route choice studies have been limited in the literature.

The earliest study that is evident in the literature was presented by Van Der Waard (1988), which showed the relative importance of transit travel time components. The author used a revealed preference (RP) survey data for 1,095 passengers to develop different MNL models. On the other hand, some literature (Fosgerau et al., 2007a, b; Nielsen et al., 2000; Vrtic and Axhausen, 2003) used stated preference (SP) data to study transit route choice behaviour. Vrtic and Axhausen (2003) used a hub-and-spoke network to study Swiss passengers regional and long-distance route choice while Fosgerau et al. (2007a, b) explored the value of time for public transport travel. Although the SP survey is relatively cheaper than the RP survey, this cannot obtain the users' actual feeling/perception as the users do not actually experience the alternatives.

Apart for Van Der Waard (1988), there were few studies, which used RP data to model transit route choice (Anderson et al., 2017; Bovy and Hoogendoorn-Lanser, 2005; Eluru et al., 2012; Raveau et al., 2011; Tan, 2016). Bovy and Hoogendoorn-Lanser (2005) developed a route choice model for train travellers in Rotterdam, Netherlands by surveying 235 respondents. Raveau et al. (2011) presented a model where the choice set was prepared from the RP data, and mostly one competing route was used along with the observed route. Anderson et al. (2017) used a doubly stochastic path generation algorithm to construct the choice set. The authors (Anderson et al., 2017) used travel survey data collected in Amsterdam and presented route choice models which accounted for path overlap and users taste heterogeneity. Tan (2016) used a combination of approaches including labelling, link elimination, simulation, k-shortest path, nested labelling and link elimination. The author (Tan 2016) used RP data along with smart card data for Singapore to develop the route choice models.

The research gap in the public transport route choice modelling literature is presented in Table 4.1. Only two studies (Anderson et al., 2017; Tan, 2016) have used RP data along with a multi-modal network. While these studies considered path overlapping and users' taste heterogeneity, these did not consider the sampling of alternatives as the choice set size of these studies were within the limit of 100 alternatives. In the current study, it is found that the choice set size varies between 2 and 1925, with many observations having a choice set size of more than 100 (see subsection 2.2.2.4 of this thesis). Therefore, different sampling of alternatives are considered in this study, along with the correction for the path overlap. Again, this study also considers the RRM approach (will be discussed in Chapter 5), which has not been considered in the previous public transport route choice literature.

Table 4.1 Research Gap in Public Transport Route Choice Modelling

Data/network	Model Specification					
	MNL/NL	**Overlap**	**Mixing**	**Mixing and Overlap**	**Sampling, MNL/ NL, Overlap**	**Sampling, RRM, Overlap**
SP/hub-and-spoke	Vrtic and Axhausen (2003)					
RP/hub-and-spoke	Van der Waard (1988) Bovy and Hoogendoorn-Lanser (2005) Raveau et al. (2011)	Bovy and Hoogendoorn-Lanser (2005)	Eluru et al. (2012)			
SP/network	Nielsen et al. (2000) Fosgerau et al. (2007a, b)		Nielsen et al. (2000) Fosgerau et al. (2007a, b)			
RP/network		Anderson et al. (2017)	Anderson et al. (2017)	Anderson et al. (2017; 2016)	**Current Study**	**Current Study**
RP / Smart Card / Network	Tan (2016)		Tan (2016)	Tan (2016)		

** This table has been adapted from Anderson et al. (2017) and updated*

The two main challenges in route choice modelling are the generation of a choice set and estimation of discrete choice models (Prato, 2009). The lack of detailed information is a significant challenge in the choice set generation as it hampers the understanding of the real choice and the considered alternatives. Another challenge in this regard is to develop a choice set that is both heterogeneous and realistic. The second challenge is to understand the vast number of alternatives to be processed in discrete choice modelling and how users perceive the characteristics of these alternatives to define their preferences.

Although route choice models are characterised by a large number of alternatives, it might sometimes become computationally challenging to estimate the parameters by considering all the alternatives (Guevara and Ben-Akiva, 2013b; Lemp and Kockelman, 2012). Some researchers even think that the generation of such substantial alternative sets with all their attributes is impractical (Dial, 1971; Guevara and Ben-Akiva, 2013b). Dial (1971) introduced the idea of enumerating fewer paths, which he termed "efficient paths" to avoid the explicit path enumeration process. Later, other researchers worked on his idea and suggested improved methodologies. The efficient path was criticised by other researchers (Akamatsu, 1996; Bell, 1995) as they found it ignoring realistic paths which were attractive to the drivers but were assigned zero flow. They proposed an alternative solution based on a more general idea of a link-based Markov decision process which did not require extensive path enumeration. In recent times, new methodologies like the recursive logit model for infinite path set (Fosgerau et al., 2013), optimal routing policy (ORP) (Ding et al., 2014; Gao and Chabini, 2006), Joint Network GEV (JNG) model (Papola and Marzano, 2013), and N-GEV (Hara and Akamatsu, 2012) have been introduced in the route choice modelling. In contrast to Dial's (1971) idea, some researchers have used extensive path generation processes by using a labelling approach (Ben-Akiva et al., 1984), a deterministic approach (Eppstein, 1998) and a probabilistic approach (Cascetta and Papola, 2001).

To address the sampling problem, many researchers have used McFadden's (McFadden, 1977) correction factor which has been applied in the logit model formulation. This formulation consistently estimates model parameters with a sample of alternatives and has been applied in many areas of choice studies including route choice (Fosgerau et al., 2013; Frejinger et al., 2009), residential location choice (Lee and Waddell, 2010; Sermons and Koppelman, 2001; Zolfaghari et al., 2012), and trip destination choice (Carrasco, 2008). Recently, researchers (Bierlaire et al., 2008; Garrow et al., 2005; Guevara and Ben-Akiva, 2013b) have shown that sampling corrections could also be performed in a nested logit (NL) formulation.

The route choice of transit passengers is somewhat different from that of drivers' (auto) route choice as the former is characterised by fixed routes. However, it adds new dimensions as it has to follow the transit schedule or frequency. Consequently, transit route choice models have been studied in two different categories, frequency-based and schedule based. Though, in both cases, transit route choice can vary from a unique elementary path between the origin-destination pair to a more complex and different sets of elementary paths. High frequency, travel time reliability, and overlapping routes introduce the theories of travel strategies and hyper-path selection (Nguyen and Pallottino, 1988; Spiess and Florian, 1989). In this context, a travel strategy refers to a set of coherent decision rules that allow the passenger to travel from the origin to the destination where an optimal strategy is to minimise the passenger's travel time/cost. In recent models, hyper-path algorithms have efficiently enumerated a set of attractive paths for each origin-destination pair (Khani, 2013; Khani et al., 2014; Nguyen et al., 1998; Noh et al., 2012) which can be later used in discrete choice models to estimate route choice parameters.

However, as discussed previously, the idea of enumerating fewer routes can result in an information gap to understand the choice set of a passenger. For example, if a passenger is sensitive to walking, he/she might avoid transfers (as this often involves walking), or select access/egress stops requiring less walking (but more travel time). Therefore, minimising travel time might not be an optimal strategy for him/her, as mentioned in Spiess and Florian (1989). Similarly, passengers can act differently (e.g. might want to minimise the travel time) in a different situation like in peak hour or for a work trip. Furthermore, passengers travelling with small children might choose other decision rules. As other travel attributes can have an impact on the choice of route, the current definition of an optimal strategy needs to be reconsidered. This issue has been discussed in the previous chapter (Chapter 3) in the context of transit access stop choice where passengers use different strategies including minimising travel time, minimising access time and minimising number of transfer while choosing transit access stops.

4.2 Description of the Study

4.2.1 Study Methodology

In this study, different strategies are investigated in the context of transit route choice scenarios. A two-stage modelling framework is devised (see Figure 4.1). In the first stage, preliminary choice sets are enumerated for each case to capture the full information of the strategy choice followed by sampling using the proposed importance sampling protocol. In the second stage, different discrete choice structures (MNL and NL with sampling) are tested by estimating the model parameters.

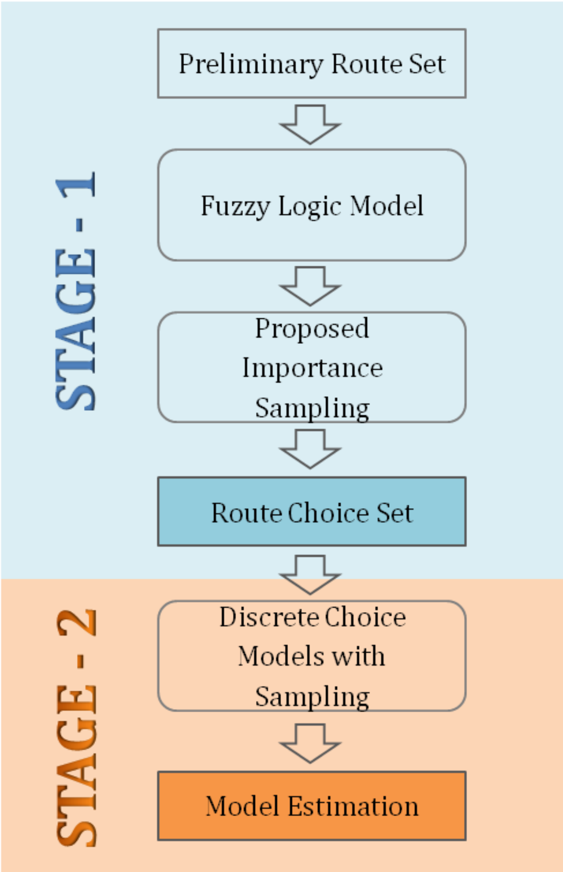

Figure 4.1 The Methodological Steps of the 2-stage Route Choice Study

4.2.2 Proposed Sampling Method for Route Choice

The value of the correction term in equation 2.10 (refer to subsection 2.3.1.7) depends on the sampling protocol used by the researcher. If all the alternatives have equal selection probabilities (uniform conditioning property; McFadden 1977), the correction for sampling bias in Equation (2.10) cancels out, and the formulation will be like equation 4.1.

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{4.1}$$

Ben-Akiva and Lerman (1985) pointed out two primary types of sampling methods: one considers uniform selection probabilities (Equation 4.1), and the other considers unequal selection probabilities or importance sampling (Equation 2.10). For many years, researchers have used the former approach along with the Simple Random Sampling (SRS) protocol (Kim et al., 2011; Pozsgay and Bhat, 2001; Simma et al., 2001; Zolfaghari et al., 2012). However, recently, a growing number of researchers have started using the latter one (Ben-Akiva and Bowman, 1998; Bhat et al., 1998; Frejinger et al., 2009; Guevara and Ben-Akiva, 2013b).

In this study, an unequal selection probability is considered. Therefore, the first part of the correction term (shown in equation 2.16), $\ln k_{in}$, depends on the sampling protocol. The second part, ln *q(i)*, is the sampling probability of alternative *i* which needs to be determined. The probability of choosing an alternative *i* by an individual *m* can be calculated using the logit formula given in equation (4.2).

$$q_m(i) = \frac{e^{\varphi A_{mi}}}{\sum_{j \in U} e^{\varphi A_{mj}}} \qquad (4.2)$$

In equation 2.13, φ is the scale factor which is considered as 1 in this study. However, for a large set of alternatives, the difference of the probabilities can be very small. As such, a large number for φ can be assumed to penalise less attractive options.

The attractiveness (A_{mi}) of the generated routes is calculated using the fuzzy logic (Zadeh, 1965) method. Fuzzy logic is a simple, rule-based, easy-to-use method which deals with reasoning or approximation rather than being fixed or exact. In contrast with the traditional logic theory, where binary sets have two-valued logic: true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Although the most popular use of fuzzy logic is in control systems (washing machine, microwave, air conditioning, traffic signal, engine throttle control, etc.), it is now used in almost every field associated with reasoning. In the transportation planning field, fuzzy logic techniques have been discussed and applied for different choice modelling studies including route choice modelling (Arslan and Khisty, 2006; Henn and Ottomanelli, 2006; Lotan, 1997; Lotan and Koutsopoulos, 1993;

Quattrone and Vitetta, 2011; Ridwan, 2004), housing search (Fatmi et al., 2017) and mode choice modelling (Kedia et al., 2015).

Fuzzy logic works in three steps. First, fuzzify all input values into fuzzy membership functions. One of the primary solutions for assigning membership functions to fuzzy variables is intuition. In the intuition approach, the membership functions are derived from the innate intelligence and understanding of modellers (Ross, 2004). Furthermore, there is no restriction on the number or shape of membership functions when forming a fuzzy set (Bezdek, 1993). The second step is fuzzy inference, which executes all applicable rules in the rule-based platform to compute the fuzzy output functions. Third, de-fuzzify the fuzzy output functions to get 'crisp' output values. Some defuzzification methods that have been used in practice to map the output of fuzzy sets into crisp numbers are: the centroid, maximum, mean of maxima, height, and the modified height de-fuzzifier (Kulkarni, 2001).

In this study, it is assumed that the attractiveness of the routes depends on three input variables: number of transfers, walking access time and total travel time. These three variables are chosen as these have been found to be significant in transit route/stop choice studies (Anderson et al., 2017; Eluru et al., 2012; Fosgerau et al., 2007a, b; Nassir et al., 2015b; Shakeel et al., 2016; Tan et al., 2016). These three variables are considered as inputs into the fuzzy logic model (FLM) while the output is the attractiveness (A_i) of the routes. In the fuzzy inference step, IF-THEN rules are used to match the inputs and the outputs. Finally, to map the fuzzy output sets, the centroid method is used to calculate and return the centre of gravity of the aggregated fuzzy set. This process is repeated four times to get a reliable estimate of the attractiveness. The core assumption for the rules is that attractiveness increases with decreasing transfers and travel time. It is also assumed that the attractiveness increases moderately for decreasing access time. 'Mamdani' inference system available in MATLAB-13 is used to run the FLMs. The final membership functions of the input variables and the defuzzification function of the output are shown in Figure 4.2, while the rules are presented in Table 4.2.

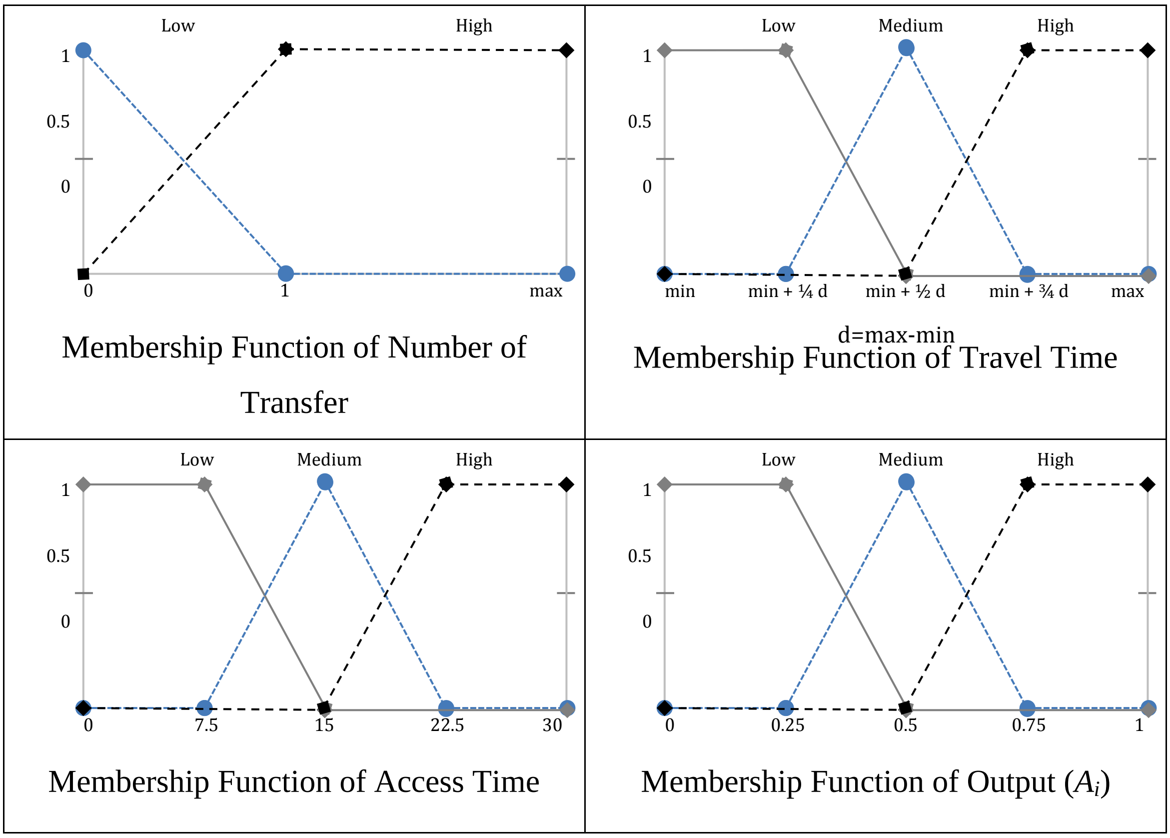

Figure 4.2 Membership Functions of Inputs and Output in the FLM

Figure 4.2 illustrates how the membership functions are set in the study. The choice behaviour of the SEQ passengers shows that in 86% of cases the chosen route has no transfer and in 13% of cases the chosen route has one transfer. As such, the membership function of the Number of Transfer has two fuzzy groups: *low* (0, 0, 1) and *high* (0, 1, max). As mentioned in subsection 2.2.2.3, maximum walk time to access the stops is considered as 30 minutes. Therefore, three membership groups for Access Time are considered with equal intervals between 0 and 30. These are *low* (0, 0, 7.5, 15) *medium* (7.5, 15, 22.5) and *high* (10, 20, 30, 30). Similarly, for Travel Time, three membership groups are defined with equal intervals according to the travel time range (the difference between the maximum and minimum travel time of the derived alternatives) for each case. For example, if a case has several route alternatives with a minimum travel time of 30 minutes and a maximum travel time of 50 minutes (range 20 minutes), the membership functions will be as: *low* (30, 30, 35, 40), *medium* (35, 40, 45) and *high* (40, 45, 50, 50). Finally, the output (attractiveness) is defined between 0 and 1 and consist of three fuzzy groups: *low* (0, 0, 0.25, 0.5), *medium* (0.25, 0.5, 0.75) and *high* (0.5, 0.75, 1, 1).

Table 4.2 Rules for the Fuzzy Logic

Rule No.	IF	AND	AND	THEN
	Number of Transfers	**Access Time**	**Travel Time**	**Attractiveness**
1	Low	Low	Low/ Medium	High
2	Low	Low	Low/ Medium /High	Medium
3	Low	Medium	High	Low
4	Low	High	Low/ Medium	High
5	Low	High	High	Medium
6	High	Low	Low/ Medium	Medium
7	High	Low	High	Low
8	High	Medium	Low/ Medium /High	Low
9	High	High	Low	Medium
10	High	High	Medium/ High	Low

4.2.3 Proposed Sampling Protocol

Using the crisp attractiveness values from the FLM, the probability of choosing an option $q_m(i)$ is calculated using equation 4.2. The options are then listed as cumulative probabilities (starting from 0 to 1) so that each option poses a range. At this stage, either the WR (with replacement) protocol or WOR (without replacement) protocol can be used for sampling. WR protocol has been used by Frejinger et al. (2009). The analysis presented in this chapter has been done using the WOR protocol; however, in Chapter 6, the sensitivity of both of these protocols is examined.

In WR protocol, route choice set C_n is generated randomly by drawing R_n routes from the universal set of routes U with replacement and then adding the chosen route to it ($|\tilde{C}_n|$ = R_n + 1). The outcome of this protocol is (k_{1n},k_{2n}, ……., k_{in}) where

k_{in} is the number of times alternative i is drawn. The protocol is demonstrated with the help of an example shown in Figure 4.3. In this example, a case is presented where there are 10 options and the choice set size is 5. The 1st column of Table A is the ID of the option where the chosen option is marked with the Asterisk. The probabilities $q(i)$ are shown in the 2nd column and are calculated from their attractiveness from the FLM. The 3rd and the 4th columns show the cumulative probabilities and the range of cumulative probabilities respectively. The figure attached beside the Table A is a graphical version of the probability range, which indicates that the chance of picking up (randomly) an option is actually the probability ($q(i)$) calculated from the fuzzy logic output. Table B shows the calculations of the correction factor. The 1st column represents the serial number while the 2nd column shows the random number. As a choice set of 5 is considered in this case, 4 random numbers are generated (from 0 to 1), and finally, the chosen option is added. These numbers are evaluated with the range of the cumulative probabilities and eventually matched options are selected and written in the 3rd column. The 4th column shows the unique options of the choice set. As such, any option (including the chosen option) can be chosen more than once, but in the choice set, it will only appear once. Therefore, the choice set size might be less than 5 (as it appears in this example). The 4th column shows the options in the choice set, and the 5th column shows the number of times the options appear during random generation. Finally, in the 6th column, correction factors are calculated.

In WOR protocol, at first, the chosen alternative is added to the route choice set C_n and then R_n routes are generated randomly from the universal set of routes U without replacement. As the number of routes in U changes in each draw, the probability of choices has to be adjusted accordingly (see Figure 4.3, Table C). Unlike the WR protocol, the WOR protocol always produces a choice set equal to the intended choice set size.

In the SRS protocol, the choice set is developed by simply picking the alternatives randomly from the universal choice set U.

Table A

Option Number	Probability $q(i)$	Cumulative Probability	Range
1	0.081	0.081	0 - 0.081
2	0.072	0.153	0.082 - 0.153
3	0.053	0.206	0.154 - 0.206
4*	0.212	0.418	0.207 - 0.418
5	0.091	0.509	0.419 - 0.509
6	0.028	0.537	0.510 - 0.537
7	0.104	0.641	0.538 - 0.641
8	0.087	0.728	0.642 - 0.728
9	0.081	0.809	0.729 - 0.809
10	0.191	1	0.810 - 1

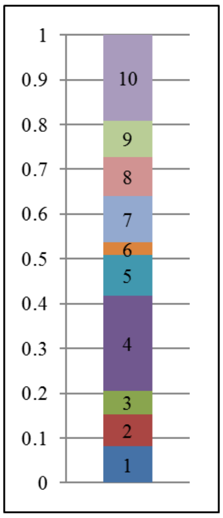

Table B: Correction Factors for WR Protocol

Serial	Random Numbers	Corresponding Option	Choice set	k_{in}	Probability $q(i)$	Correction Factor $\ln \frac{k_{in}}{q(i)}$
1	0.4801	5	4	2	0.212	2.244
2	0.2593	4	5	1	0.091	2.397
3	0.5751	7	7	1	0.104	2.263
4	0.6744	8	8	1	0.087	2.442
5	Chosen	4	-	-	-	-

Table C: Correction Factors for WOR Protocol

Serial	Random Numbers	Corresponding Option	Choice set	k_{in}	Adjusted Probability $q(i)$	Correction Factor $\ln \frac{k_{in}}{q(i)}$
1	Chosen	4	4	1	0.212	1.551
2	0.4801	7	7	1	0.132	2.025
3	0.2593	3	3	1	0.077	2.558
4	0.5751	9	9	1	0.128	2.053
5	0.6744	10	10	1	0.347	1.058

Figure 4.3 Formation of Route Choice Set from the Proposed Sampling Protocol

4.2.4 Observed Choice Behaviour

After enumerating all the routes for each of the observations, it is found that the available routes for each case vary between 2 and 1925. A summary of the number of available options is shown in Figure 4.4. It seems that for most of the cases

(61%), the options are less than 25 routes. For a significant percentage of cases (19%) there are more than 50 options, although, for 4% of the cases, there are more than 100 options. These statistics suggest a need to reduce the options to a manageable and logical size.

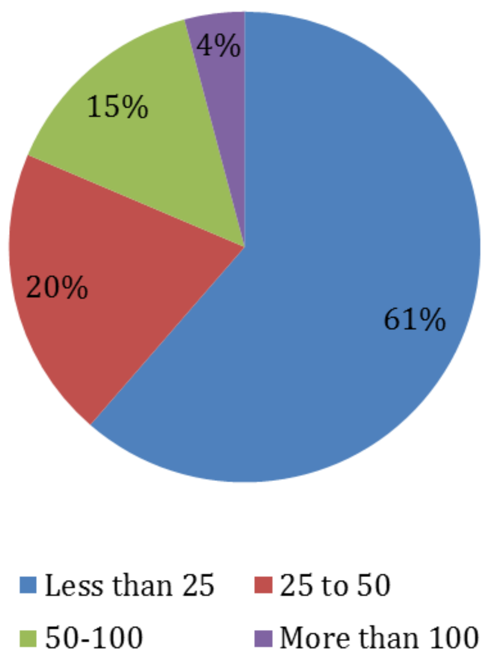

Figure 4.4 Number of Available Options

The choice of modes shown in Figure 4.5. The figure suggests that the transit users of SEQ choose only buses (69%) or only trains (26%) as their mode of transport. In a few instances (5%), they choose mixed modes. People seem to use the ferry on some occasions; however, in all the instances, it was not the only mode they choose for that journey.

The preference for the strategy, shown in Figure 4.6, seems to be quite interesting. In most of the instances (98%) people choose routes which consist of either: minimum travel time (MTT), minimum access time (MAT), minimum number of transfers (MTR) or a combination of these. However, in very few instances (2%) people choose routes that do not conform with any of the strategies mentioned above. The preference for MTR seems to be the most apparent feature. In about 92% of cases (29%+39%+6+18%) people choose routes that contain MTR. The second important strategy is MTT (63%). However, MAT strategy seems to be less favourable (27%). Another feature is that people seem to favour a multi-strategy (65%) over a single strategy or no strategy (35%).

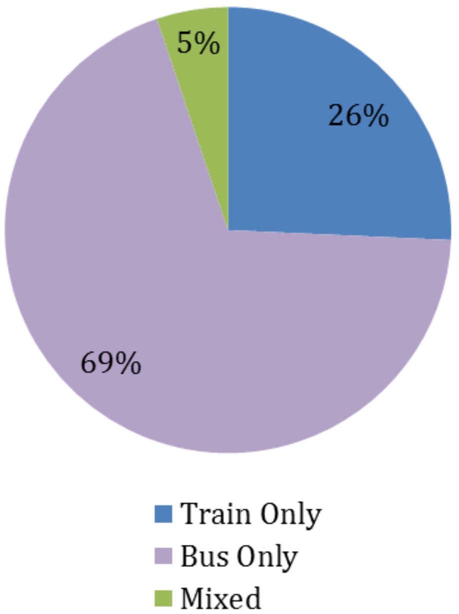

Figure 4.5 SEQ Transit Users Revealed Mode Choice Behaviour

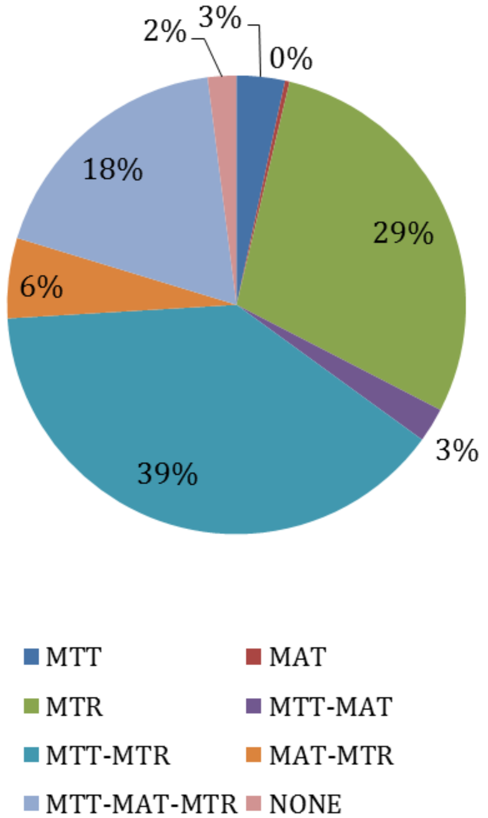

Figure 4.6 SEQ Transit Users Revealed Strategy for Route Choice

4.2.5 Route Choice Strategy in the MNL Model

Three basic strategies (MTT, MAT and MTR) are used as dummy variables and considered in the MNL model.

4.2.6 Route Choice Strategy in the NL Models

Nested logit models can account for the relationships within the alternatives by combining similar alternatives into nests. Previous route/stop choice studies addressed different sources of correlation between the alternatives by developing Nested Logit models. For example, correlation at different parts of the trip (home end, activity end) (Bovy and Hoogendoorn-Lanser, 2005), correlation between the mode of transit stop (bus stop, train station, ferry terminal) (Nassir et al., 2015b) and correlation between travel schemes (strategy or objective of travel) (Hassan et al., 2016). All these studies show that NL models can be a better way to explain the choice behaviour than MNL models. Moreover, significant improvement (goodness of fit) can be achieved through the nesting.

As mentioned in subsection 2.2.2.5, three travel strategies are considered in this study. Nesting is done in a way where the strategies are considered on the upper level (branch), and the routes associated with each strategy are considered as the twigs. Similarly, another NL model is developed which considers modes as nests. Eventually, seven NL model groups are tested in this study. The nest specifications of these models are shown in Table 4.3. The formation of the nests and the nesting structure is already illustrated in subsection 3.2.3.

4.3 Discussion of the Model Results

The models are developed with about 70% of the total observations (867 out of 1237). The rest 30% (370 observations) is kept for the validation process. The model parameters are estimated using the Maximum Likelihood Estimation process. As a sampling method is used (for the observations which have more alternatives than the desired choice set size), the estimated parameters can be different for different runs. Therefore, at first, a full MNL model (MNL-FL; choice set size of 25) is developed where only significant variables (at 95% confidence level) are included. In the second step, this model (with the same variables) is replicated 100 times. This process gives a total of 100 sets of estimation of the parameters for one particular choice set size. After that, the second step mentioned above is replicated to estimate parameters for nine other choice set sizes. This process is followed to estimate all the NL route choice models, MNL-NSA (No Strategy Attribute) and MNL-SRS models. Finally, each model group contains ten different sets of models (depending on the size of the choice set), and each of these sets comprises 100 sets of estimation results. All the discussion and analysis, presented in this section, is based on these 100 sets of results for each set of model. However, in some analysis including goodness of fit, estimation time, prediction capability and behavioural interpretation averages of these 100 sets are presented and discussed.

Table 4.3 Nest Structures for Proposed NL Models

Number	Model Name	Number of Nests	Nest Names
1	TT	2	MTT, NoMTT
2	AT	2	MAT, NoMAT
3	TR	2	MTR, NoMTR
4	TT-AT	4	MTT, MAT, MTT&MAT, None
5	TT-TR	4	MTT, MTR, MTT&MTR, None
6	AT-TR	4	MAT, MTR, MAT&MTR, None
7	Mode	3	TrainOnly, BusOnly, Mixed

4.3.1 Goodness of Fit of the Models

The goodness of fit (ρ^2) of the models for different choice set sizes is presented in Figure 4.7. As each model group consists of different choice set sizes (hence different initial log-likelihood values), the models cannot be compared. However, in general, if the comparison is done within the same choice set size, the MNL-SRS and MNL-FL models appear to have better goodness of fit than the MNL-NSA model. Similarly, within the NL models, TT-AT and TT models appear to have better goodness of fit than the other NL models.

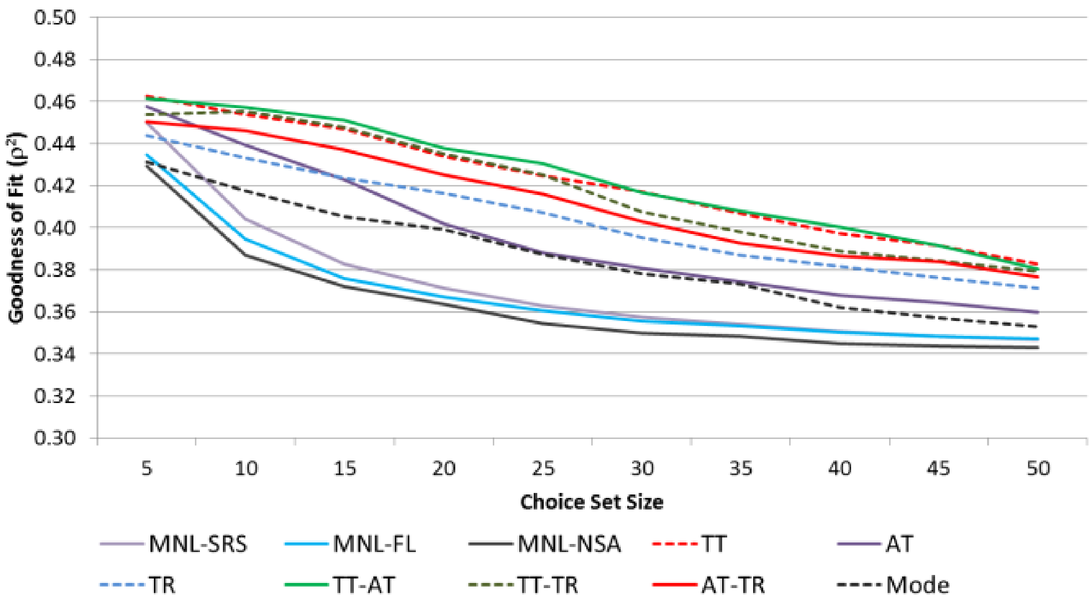

Figure 4.7 Goodness of Fit of the Route Choice Models

4.3.2 Estimation Time of the Models

The model estimation times are recorded and presented in Figure 4.8. In general, the estimation time for all the models seems to be acceptable. However, MNL models, especially MNL-SRS models, need less time than the NL models (e.g. TT, AT, etc.) to execute one single run. This is due to the simplicity of the MNL model structure. The average time needed to execute each run tends to increase with an increase in choice set size up to 25 or 30. After that, the computation time seems to gradually decrease. This might happen because most of the cases (61%) have alternatives less than 25 (See Figure 4.4). Therefore, for a higher choice set size (more than 25) these cases do not need to run the sampling segment, which saves computational time. Therefore, for the NL model, the computational time seems to decrease after a choice set size of 25 or 30.

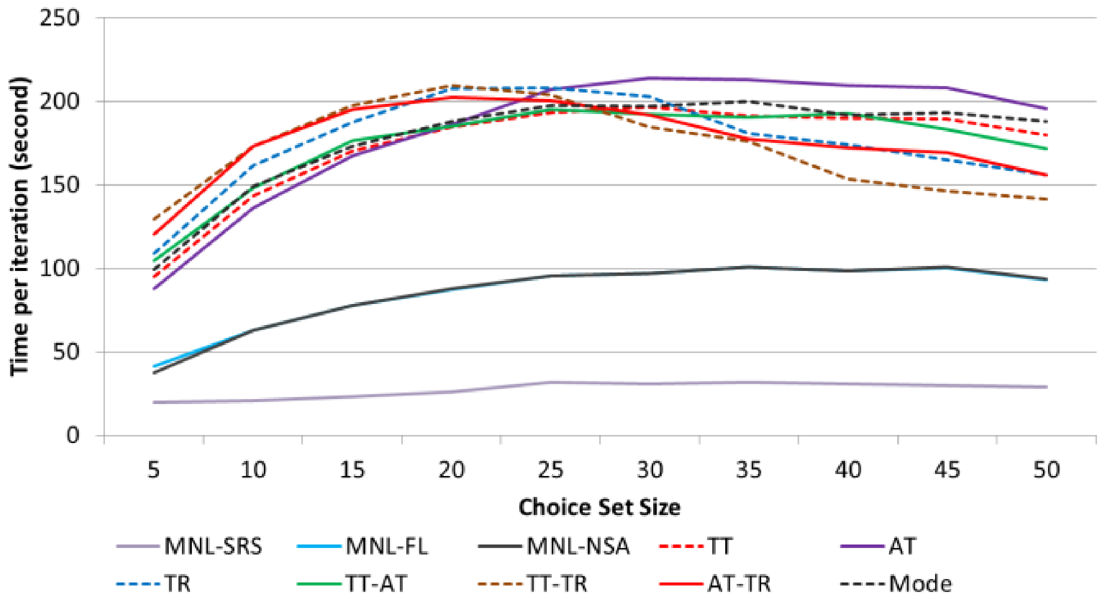

Figure 4.8 Computational Times of the Route Choice Models

4.3.3 Parameter Estimation

The estimated parameters (of the 100 estimates) of the MNL-FL model are presented as box plots in Figure 4.9. Figure 4.9 shows that the estimated parameter values are a bit different for the different choice set sizes. The variability of the estimated parameters seems to be more in the lower choice set size models as the whiskers seem to be more dispersed than the higher choice set size models. Therefore, the parameter estimates seem to be more consistent in the higher choice set size models. A similar pattern is also evident from the estimated parameters of the other models in the study and in the literature (Guevara and Ben-Akiva, 2013b; Guevara et al., 2014)

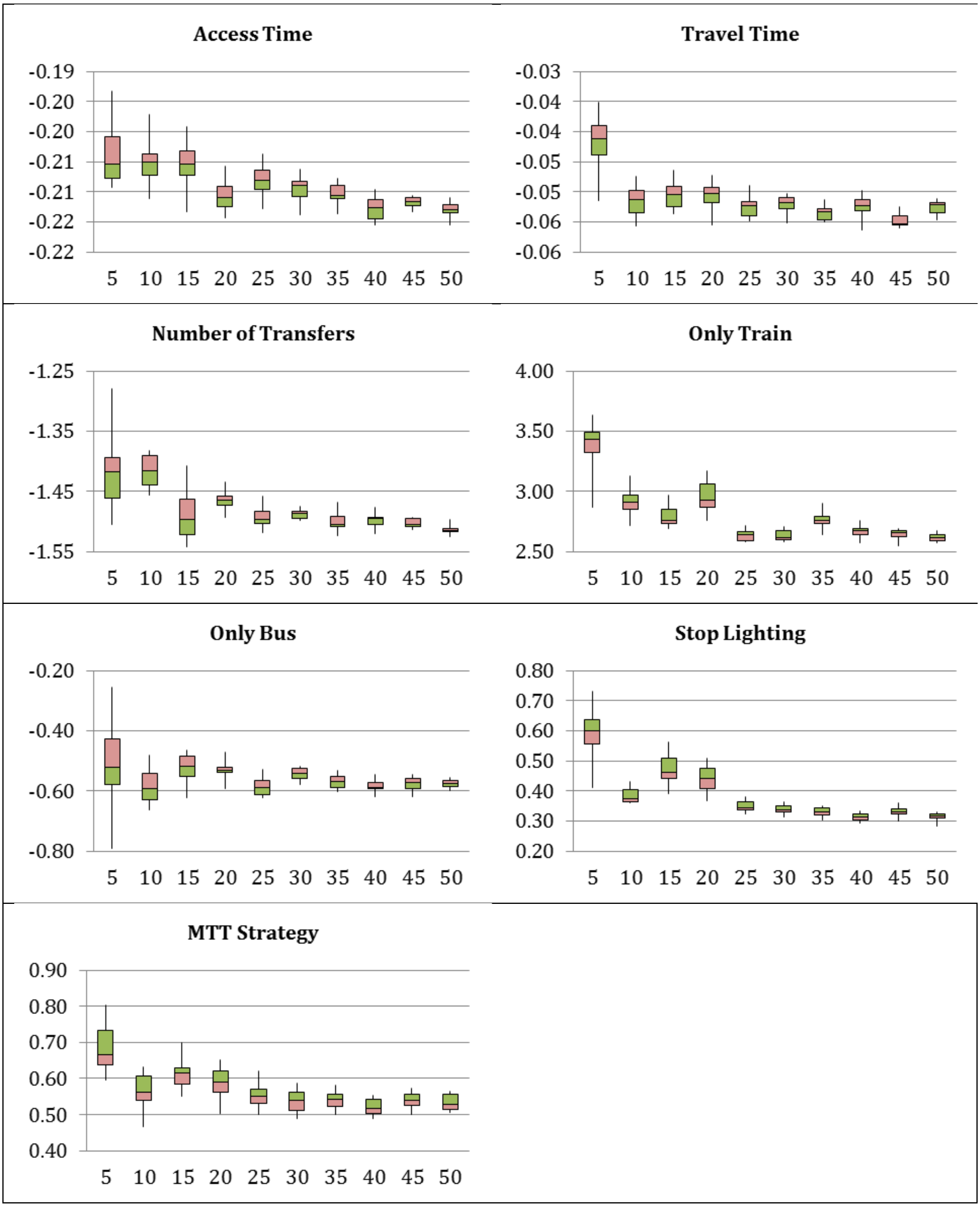

Figure 4.9 Boxplot of the Estimated Parameters of the MNL Model

4.3.4 Nest Coefficients of the NL Models

The nest coefficients for the NL models are tested for significance at 95% confidence level. T-tests are performed against the null hypothesis, i.e. the nest coefficient value of zero and one. Considering the average of all the iterations, nest coefficients of TT, AT, TR and AT-TR models seem to be significant in both the tests. All the nests of Models TT-AT and TT-TR found significant against the value of 0. However, two nests of TT-AT and one nest of TT-TR model are found to be insignificant against the value of 1. These nest coefficients are also found to be significant at 90% confidence level. The nest structure of the Mode model is found to be insignificant in both the tests, even at 90% confidence level and thus it is not used for further analysis. Consequently, it can be concluded that the nest structures of all the NL models (accept Mode) can be acceptable.

4.3.5 Prediction Capability of the Models

The mean parameter values are used to see the prediction capability of the models. The validation dataset is used for this purpose. The sampling protocol is altered slightly to capture the prediction capability. As such, the chosen alternative is not included separately and the number of samples is kept the same as the choice set size. Therefore, it is easy to find whether or not the chosen alternative is picked by the sampling protocol to form the choice set. Similarly, if the chosen alternative is picked in the choice set, the relevant discrete choice model is applied. Next, based on the choice probability, the chosen alternative is ranked. This whole process is replicated 100 times and the detection percentages (% of correctly predicted cases) are summarised in Figure 4.10 to Figure 4.12. The prediction capability of the models are presented according to three criteria: 1) whether the chosen route is detected in the first stage based on sampling (Figure 4.10), 2) whether the chosen route appears in the top 5 based on the logit probability (Figure 4.11), and 3) whether the chosen route appears in the top 10 based on the logit probability (Figure 4.12).

Figure 4.10 shows that the TT-TR models outperformes the other models by detecting the chosen alternative on more than 80% instances (in the choice set size 45 and 50). The MNL-FL models seem to be the second-best while the MNL-SRS models perform moderately. However, detecting the chosen alternative (by using the choice model) seems to be better by using the MNL-FL models, mainly when the choice set size is 30 or more. The TT-TR models show higher prediction capability (in the discrete choice model) than the MNL models when the choice set size is 25 or less. Although TT and TT-AT models demonstrate better goodness of fit, interestingly, the prediction capability of these models is not that good. Another interesting finding is that all the models seem to detect the chosen alternative in the choice set better when the choice set size increases. However, most of the models (except the MNL models), irrespective of choice set size, seem to be indifferent while detecting the chosen alternative by applying the choice model. The MNL-FL and MNL-SRS models seem to show an improvement when the choice set size is 30 or more. However, the MNL-FL models show better prediction capability than the MNL-SRS models.

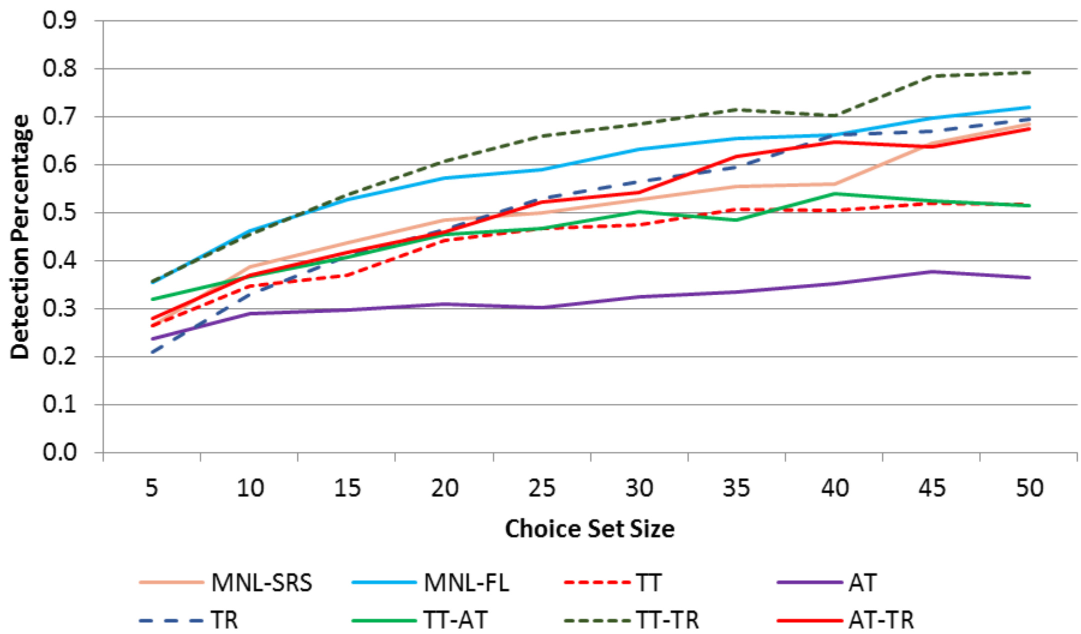

Figure 4.10 Prediction Capabilities of the Route Choice Models in the Choice Set Formation Stage

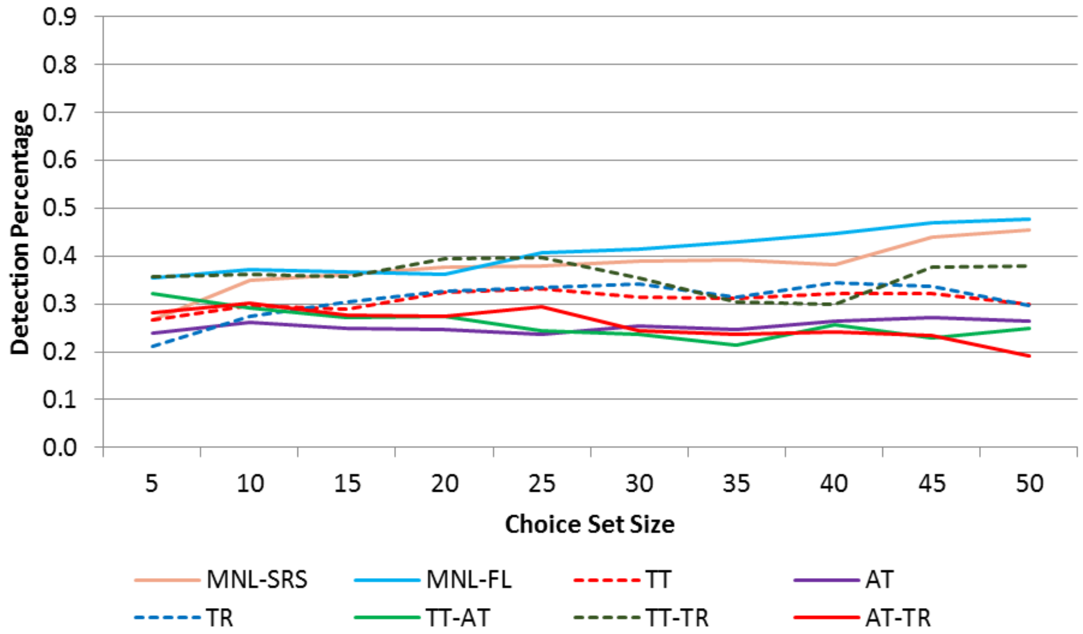

Figure 4.11 Prediction Capabilities (within top 5) of the Route Choice Models in the Discrete Choice Modelling Stage

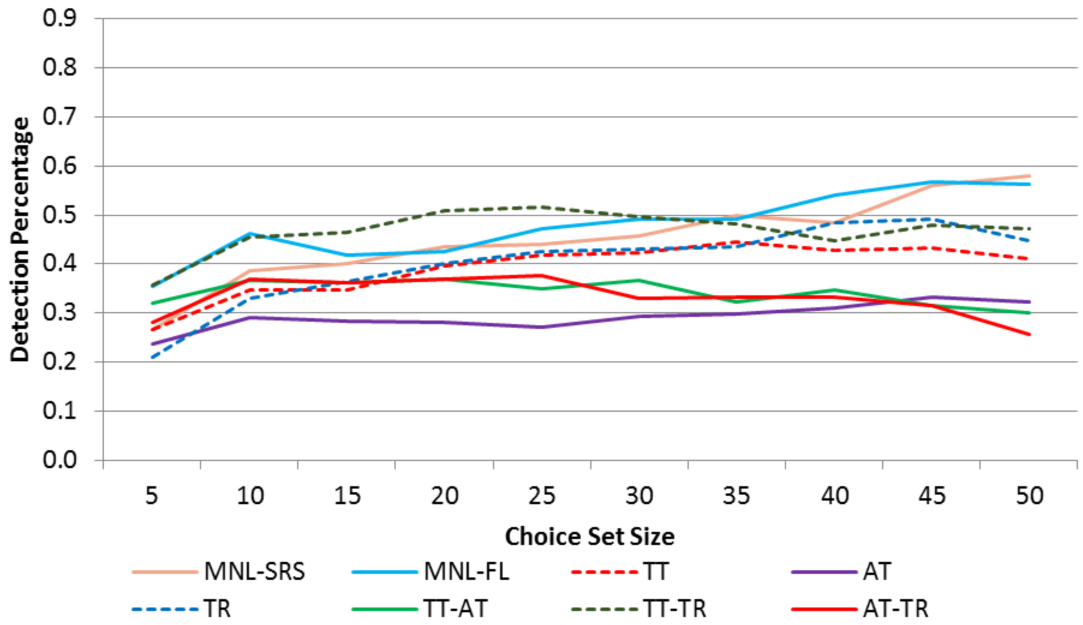

Figure 4.12 Prediction Capabilities (within top 10) of the Route Choice Models in the Discrete Choice Modelling Stage

4.3.6 Behavioural Interpretation

Considering prediction potentials, the MNL models seem to be better for a choice set size of 30 and more. However, for a lower choice set size, TT-TR model seems to be comparatively better. As these findings are obtained from a specific dataset rather than a controlled or simulated dataset, the findings cannot be generalised. However, this is an interesting issue and can be a topic of future study. The estimated parameters and the model statistics of these two models are presented in Table 4.4.

Table 4.4 Estimated Parameters for the Best Discrete Choice Models

Explanatory Variables (β)	Mean of β value		t-statistics (min, max)	
	MNL-50	TT-TR-25	MNL-50	TT-TR-25
Access Time	-0.216	-0.08	-13.2, -2.9	-6.7, -2.1
Travel Time	-0.054	-0.045	-8.4, - 1.8	-4.8, -2.1
Number Of Transfer	-1.51	-	-5.4, -1.6	-
Only Train	2.524	1.213	4.3, 10.1	2.5, 6.1
Only Bus	-0.585	-0.194	-3.4, -1.7	-7.4, -1.2
Stop Lighting	0.270	-	1.7, 3.4	-
Travel Time Strategy	0.468	-	1.8, 4.5	-
Weekday **X** Walking Time	-	-0.011	-	1.6, 3.9
Nest Coefficients (λ)				
MTT	-	0.355	-	1.7, 3.8
MTR	-	0.357	-	3.2, 6.3
MTT&MTR	-	0.415	-	1.7*, 2.9
None	-	0.343	-	3.4, 6.9
Model Statistics				
Number of Estimated Variables	7	9	-	-
Initial log-likelihood	-2347	-2782	-	-
Final log-likelihood	-1541	-1607	-	-
Goodness of Fit (ρ^2)	0.345	0.423	-	-

All the parameters are significant at 95% confidence level
X *Indicates interaction* **Significant at 90% confidence level*

Table 4.4 shows that the signs of the parameters are reasonable. From the group of impedance variables, *travel time* and *number of transfers* are found to be significant. The signs of these parameters suggest that the utility of a route will be high when the travel time and the number of transfers are less. One of the Stop variables, *access time to the stop* seems to be significant with a negative sign. However, the time parameters suggest that the users do not consider travel time and access time in the same manner; the access time poses more disutility than travel time. This finding aligns with the previous studies.

Table 4.5 shows a comparison of the rate of substitution with the previous studies. In the current study, the access time of one minute is considered as about 0.25 minutes (TT-TR-25 model) to 0.56 minutes (MNL-50 Model) of travel time. Again, the SEQ passengers pose high disutility for transfers. Table 4.5 shows that one additional transfer poses the same disutility as walking for 7 minutes (MNL-50) to an access stop. This indicates that the SEQ transit passengers prefer to take the most direct route as observed from their choice behaviour (86% cases the chosen route has no transfer and 13% cases the chosen route has one transfer). From Table 4.5, the substitution patterns of the proposed models are consistent with most of the previous studies.

Another stop specific variable, the presence of lighting at the access stop, is also found to be significant in MNL-50 model with a positive sign. The substitution pattern shows that passengers can walk around 1.25 (0.270/0.216) minutes to access a stop if it has lighting facilities. The mode parameters suggest that *only train* has a higher utility than *only bus* and *mixed mode* (here mixed mode is the base case). However, the *only bus* seems to pose disutility as the sign is found to be negative. Therefore, the SEQ transit users prefer train over bus as found in the previous studies (Anderson et al., 2017; Fosgerau et al., 2007a; Nassir et al., 2015b).

Table 4.5 Comparison of Rate of Substitution

Study	Access Time	Travel Time / In-Vehicle Travel Time	Number Of Transfer
Current (MNL-50 Model)	1	0.25	7.0
Current (TT-TR-25 Model)	1	0.56	-
Anderson et al. (2017)	1	0.1-0.4	5.9
Bovy and Hoogendoorn-Lanser (2005)	1	1.25-2.7	3.2-14.25
Fosgerau et al (2007a)	1	0.56 **	-
Fosgerau et al (2007a)	1	0.77 *	-
Raveau et al. (2011)	-	1 *	3.8
Eluru et al. (2012) (st. dev.)	-	1 **	9.3 (3.6)
Nassir et al. (2015b)	1	0.14-0.15	4.1-4.4
Tan (2016)	1	0.28	7.37

** Metro in-vehicle time ** Bus in-vehicle time*

In the MNL-50 model a strategy variable, *travel time strategy,* shows a positive sign which means passengers get higher utility if a route minimises their travel time. This can also be interpreted as: users prefer minimum travel time strategy while choosing transit routes. Similarly, during weekdays, the SEQ passengers seem to prefer less walking while selecting a route (in TT-TR-25 model). Furthermore, all the nest coefficients of the TT-TR-25 model, which are based on travel strategy, are found to be significant. The value of the nest coefficients suggests that there are moderate to high correlation in unobserved factors within each nest justifying the necessity of the NL models.

The correction factor for the path commonalities is found to be insignificant (even at 80% confidence level) in all the models, and no significant improvement is observed regarding the goodness of fit. Therefore, this is not reported. However, these corrections prove to be significant (regarding parameter estimation and goodness of fit) in the general route choice setting. Several researchers have also reported this as insignificant (Nassir et al., 2018), counter-intuitive (Anderson et al., 2017; Tan et al., 2016) or at best marginally significant (Nassir et al., 2015b, 2016).

The pseudo direct marginal effects of the route choice attributes are presented in Table 4.6. The effects of the attributes are similar between the two presented models. A train route alternative seems to pose the highest effect of 0.32 (MNL model), which means that the probability of the chosen alternative will change by 32% if the attribute changes from train to not-train or vice versa. However, for a bus route alternative the effect is 10%, which is one-third of the *only train* attribute. *Number of transfers* shows a strong effect of 0.19, which means if number of transfer for the chosen alternative is increased or decreased by one, the choice probability will change by 19%. *Travel time strategy* dummy shows an effect of 0.07, which means when this attribute changes (from MTT strategy to No-MTT strategy or vice versa) for the chosen alternative, the probability of the chosen alternative would change by 7%. The effects of *access time* and *travel time* are reported as 0.031 and 0.018 (MNL Model) respectively. This means that one minute change in access time and travel time for the chosen alternative would change the probability of the chosen alternative by 3.1% and 1.8% respectively.

Table 4.6 Pseudo Direct Marginal Utility of selected Route Choice Models

Explanatory Variables	**MNL-50**	**TT-TR-25**
Access Time	0.031	0.027
Travel Time	0.018	0.021
Number Of Transfer	0.19	-
Only Train	0.32	0.29
Only Bus	0.10	0.10
Stop Lighting	0.04	-
Travel Time Strategy	0.07	-
Weekday **X** Walking Time	-	0.02

4.4 Conclusions

This study shows how the choice set size plays a role in route choice modelling. The increase in the choice set size can increase the chance of the chosen alternative being included in the choice set. However, a choice set with a high number of alternatives can affect the prediction capability as the study found that the predictive ability of the NL models did not improve when the choice set size is more than 25. Moreover, in some cases, the prediction capability appears to be decreasing. However, the MNL model seems to be improving with the increase in the choice set size. Therefore, choosing an optimal choice set size for a particular model seems to be very important.

This study presented an importance sampling protocol, which samples the alternatives from a probability profile of the universal set of alternatives. Therefore, the probability of including attractive alternatives is higher than the simple random sampling protocol. The complexity added by the fuzzy-logic in the sampling stage proved to be efficient as some of these models show higher prediction capability when compared to the traditional simple random sampling models (MNL_SRS). Therefore, the combination of this econometric-rule-based approach of reducing the size of the choice set while keeping the likely alternatives is the most significant contribution of this study.

Similarly, the study also has some noteworthy findings from a behavioural perspective. Two access stop variables, walking time to the stop, and lighting at the access stop are found to be significant. Similarly, walking time to the stop is found to be more valuable than the total travel time. Furthermore, users seem to prefer trains over buses. Purpose of the trip is also found to be significant.

Another significant contribution of this study is to understand and quantify how transit travel strategies can be valuable in transit route choice modelling. Transit travel strategies show a significant effect on the route choice as it improves the explanatory power of the models, especially in the NL structures. All the NL structures of the strategy group show significant nest structure. These suggest that the routes can be categorised according to the travel strategy as the travel strategy nests show medium to high correlation in unobserved factors within each nest. Similarly, the assessment of MNL-NSA model shows that the omitted strategy attributes significantly reduce the explanatory power of the MNL model. Furthermore, considering the strategy variables in the modelling structure significantly improves the model's prediction capability and thus enable better transit demand estimation.

Furthermore, the results from this study can be used in the transit path generation algorithms to generate sensible alternatives by considering the strategy variables which can eventually reduce the computational burden and increase the accuracy of the algorithm.

The study has the following limitations:

- This study assumes the scale factor φ (equation 4.2) as 1. However, as different values of φ can change the values of attractiveness, it would be worth to investigate the potential impact of this scale parameter on the choice set formation process.
- Some of the travel time components including, waiting time and in-vehicle time for train bus or ferry, are not found to be significant (within an 80% significance level) in any of the models. Therefore, the substitution pattern for all the travel time components cannot be obtained. This can be considered as a limitation of this study.

The correction factor of the path commonality is found to be insignificant in the study (even at 80% confidence level). Even a few previous studies have also reported this issue. However, the correction factor for the path commonality is found to be significant in the RUM-RRM hybrid model discussed in Chapter 5.

One of the future directions of the route choice modelling can be to investigate the effect of these strategies in other modelling structures including Cross-Nested

Logit, Mixed Logit and Latent Class Model. Similarly, joint modelling for stop-route and strategy-route can also be investigated.

5 RRM APPROACH FOR MODELLING TRANSIT STOP AND ROUTE CHOICE

This chapter investigates the problem of "how people make transit stop or route choice decision". Several RRM specifications are tested and the results are compared with the results from the RUM specifications presented in Chapter 3 and 4. Generally, it seems that people value both the utility and the regret simultaneously as hybrid specifications show a better outcome than the individual specifications. The stop choice modelling results show that RUM specifications yield better goodness of fit, although the RUM-RRM hybrid structure proves to be better in prediction. The regret based route choice is modelled similarly like utility-based models by using the importance sampling mechanism proposed in Chapter 4. The proposed sampling technique proves to be considerably better than the conventional simple random sampling technique. The route choice modelling results show that the hybrid of RRM and RUM specifications performs comparatively better than the RUM specifications, especially when the choice set size is less than 30. However, the RUM specification seems to be better in the higher choice set size models.

5.1 Introduction

In the domain of discrete choice modelling, there are two major approaches or decision rules. The first one is Random Utility Maximisation or RUM, introduced by McFadden (1973), where the philosophy is that the decision-makers will try to maximise their utility while choosing an alternative. When choosing between different alternatives, this decision rule assumes that the decision-makers assign a utility to each alternative and choose the one that has the highest utility. The utility usually consists of a function of the attributes of the options and associated parameters (or decision weights) along with an error term.

The previous chapters modelled the transit choice behaviour by using the RUM approach, where the philosophy of modelling is to see users as utility maximisers. This chapter studies the transit choice behaviour by exploring another approach called the Random Utility Minimization or RRM, introduced by Chorus (2008), where the philosophy is to minimise the regret while choosing alternatives. The RRM approach is relatively new compared to the RUM approach and starting to gain attention in the literature just a decade ago. The RRM approach is a non-utilitarian discrete choice model inspired by the regret theory (Bell, 1982; Fishburn, 1982; Loomes and Sugden, 1982). RRM is based on the concept of regret, where people anticipate and aim to minimise regret rather than maximise utility. In this approach, people compare the attributes of each alternative and regret arises when one or more non-chosen alternatives perform better than the chosen one.

RRM is fundamentally different from the RUM approach as it compares pairs of alternatives according to their attributes. As such, this approach needs more computational time, especially when the choice set size is large. Again, this approach can be useful in many social studies where regret in choice behaviour is more relevant (Chorus, 2012). Moreover, this approach provides an alternative behavioural interpretation and can provide useful insights for policy and planning. Furthermore, this approach is appealing, since many studies reported better (or comparable) results than the RUM approach. Besides, there is a significant gap in the literature, especially using the RRM modelling approach in transit choice modelling (see Table 5.1). This chapter aims to fill this gap by modelling transit stop choice and transit route choice through the RRM approach.

The RRM model has been studied in several choice contexts of travel demand management such as destination choice (Boeri et al., 2012; Chorus, 2010), departure time choice (Chorus and De Jong, 2011; Shabanpour et al., 2017), mode choice (Boeri and Masiero, 2014; Hensher et al., 2016; Hess and Stathopoulos, 2013; Leong and Hensher, 2015) and route choice (Bekhor et al., 2012; Chorus and Bierlaire, 2012; Chorus et al., 2013; Leong and Hensher, 2015; Prato, 2014; Prato et al., 2012). An empirical comparison between RUM and RRM shows promising insights to use the RRM approach in parallel to the RUM approach. However, to date, no study seems to have considered the RRM approach in transit route choice or transit stop choice modelling.

5.2 Literature Review

There is not much literature which discusses the transit choice behaviour in an RRM setting. Table 5.1 compiles the existing literature in travel mode and travel route choice modelling that used the RRM approach. Table 5.1 also shows how this study contributes to fill the knowledge gap of the RRM modelling approach in transit route and stop choice.

Bekhor et al. (2012) presented a static stochastic user equilibrium (SUE) model where RRM is used in the route choice part. The method was demonstrated through a small synthesised grid network and the real road network of Winnipeg, Manitoba, Canada. They compared the results of the model with an RUM based (MNL) route choice model. The results showed that the link flows can be very different for these two modelling approaches depending on the network topology and the number of generated routes. However, the study suggested doing more research to understand this new approach (RRM).

Li and Huang (2017) presented an RRM route choice model for the SUE problem. They used three different networks to test the performance of the model in different conditions. This study found more consistent flow pattern from the RRM-based SUE than the RUM-based SUE model.

Table 5.1 Characteristics of the RRM Literature

Field	Data type	Study	RRM Specifications
Car Route Choice	EMME-2 Network	Bekhor et al. (2012)	Classical RRM
	Simulation	Li and Huang (2017)	Classical RRM
	SP	Chorus and Bierlaire (2012)	Classical RRM
	SP	Hess and Chorus (2015)	Latent class G-RRM
	Simulation	Prato (2014)	Correction for Path Commonality, VOT
	RP and real road network	Prato (2014)	
Mode Choice	SP	Leong and Hensher (2015)	RRM, Hybrid RUM-RRM
	SP	Hensher et al. (2016)	Mixed RRM, Elasticity
	Simulation	Van Cranenburgh and Prato (2016)	P-RRM
Parking Lot Choice	RP	Guevara et al. (2014)	Correction for Sampling
Transit Route Choice	RP and real transit network	Current Study	Correction for Sampling, Correction for Path Commonalities, classical RRM, hybrid RUM-RRM
Transit Stop Choice	RP and real transit network	Current Study	P-RRM, μRRM, Hybrid RUM-RRM

Chorus and Bierlaire (2012) tested the compromise effect in the context of route choice behaviour. In this study, they considered three approaches: (i) introducing a variable in the logit model which represents the extent of the compromise, (ii) using the classical RRM approach, and (iii) developing a contextual convexity model. Data from a stated preference route choice survey was used to develop the models. The results showed that the contextual convexity and the RRM model both perform better in terms of the model fit and prediction capability.

Leong and Hensher (2015) proposed a new modelling approach "relative advantage maximising" or RAM, which is similar to the RRM approach. This new approach captures the parsimony and choice set dependence of the RRM approach. However, the difference between these two models is that RAM considers both the disadvantages and advantages of an alternative with different functions. The study tested these models on a few stated preference datasets on mode choice. The results were compared with the model specifications including RUM, RRM and hybrid RUM-RRM. They found competitive results for all these approaches.

Prato (2014) introduced corrections for the similarity of the alternatives in route choice modelling. He also presented an analysis of the choice set composition effect and comparison between RRM and advanced RUM models. In this study, Prato discussed three different model specifications to incorporate the corrections for path commonalities including a utility-based correction (RUM-RRM hybrid), RRM based correction, and RRM based pairwise correlation (see subsection 2.3.2.11). For the correction term, the author examined C-logit, PS-logit and PSC-logit. The RUM-RRM hybrid model showed better results in the "overlapping network", but showed poor results in the "switching routes network". However, when the disjoint portion of the routes was weighted for the correlation, C-RRM provided less error than PS-RRM and PSC-RRM. Again, C-RRM was found to differentiate better between relevant and irrelevant routes. However, as these experiments were performed on dummy networks the results cannot be generalised. In this study (Prato, 2014), a revealed preference data was used to model the route choice behaviour. The RRM-MNL model was found to provide similar goodness of fit like the RUM-MNL models. However, the value of time analysis showed that the RRM-based model generally provided lower estimations than the RUM based models.

Hensher et al. (2016) introduced preference heterogeneity in the RRM modelling structure by proposing RRM-mixed model for mode choice. In this modelling structure, the attributes were assumed to have a distribution like in the mixed logit models to capture the heterogeneous behaviour of the users. They tested different model specifications including, MNL, mixed MNL, RRM and RRM-mixed with the help of a stated preference data of Sydney, Australia. They applied the Vuong test (Vuong, 1989) to see if there was any improvement in the overall statistical fit when migrating from one model to another. The study found no significant improvements in terms of the model fit between the RUM and RRM modelling approaches. However, interesting behavioural contrasts were evident in the elasticity measures and choice probability moments. Finally, the study concluded that both the RUM and RRM approaches performed well as there was no conclusive evidence found to establish the superiority of one over the other.

Van Cranenburgh and Prato (2016) explored the robustness of the P-RRM models towards the omitted attributes. They found that in the labelled data, alternate specific attributes can be used to capture the unobserved average effects of the omitted attributes. However, if the choice set varies for different observations, alternative specific attributes need to be estimated for each unique choice set group to get consistent parameter estimates. By using a Monte-Carlo simulation method for the unlabelled alternatives, they found that the RRM model is somewhat robust towards omitted attributes, however not as robust as the RUM model.

Hess and Chorus (2015) used a stated preference dataset to study the route choice behaviour of the Dutch car commuters. In this study, they introduced a latent class modelling structure for the G-RRM model. The G-RRM specification allowed the latent classes to demonstrate a full RRM or full RUM behaviour. Therefore, this structure enabled to capture the decision heterogeneity and taste heterogeneity of the respondents at the same time.

5.3 RRM Transit Stop Choice Model

As there are a wide variety of model formulations available to test the choice behaviour, three RRM formulations are tested for the stop choice model discussed in Chapter 3. As the classical RRM and G-RRM models are not suitable for the varying choice set (Van Cranenburgh et al., 2015b) (see subsection 2.3.2.9), P-RRM and μRRM models are chosen for the stop choice modelling.

- Model-1: P- RRM (equation 5.1)
- Model-2: μRRM (equation 5.2)
- Model-3: Hybrid RUM-RRM (equation 5.3)

After testing different model specifications for the above mentioned three models, specifications are finalised on the basis of the parameters that are found to be significant at 80% confidence level and are provided in equation 5.1 to 5.3. As the choice set is variable, equation 2.25 is applied. The discrete choice estimation package BIOGEME (Bierlaire, 2003) is used to estimate the models.

$$RR_i^{Model\,1} = \frac{\Gamma}{J_n}\left(\beta_{AccessTime}P_{AccessTime} + \beta_{Train}P_{Train} + \beta_{MTT}P_{MTT} + \beta_{StopLight}P_{StopLight}\right) \tag{5.1}$$

Where, $P_{AccessTime} = \sum_{j\neq i} min\left(0, AccessTime_j - AccessTime_i\right)$

$P_{Train} = \sum_{j\neq i} min\left(0, Train_j - Train_i\right)$

$P_{MTT} = \sum_{j\neq i} min\left(0, MTT_j - MTT_i\right)$

$P_{StopLight} = \sum_{j\neq i} min\left(0, StopLight_j - StopLight_i\right)$

$$RR_i^{Model\,2} = \frac{\boldsymbol{\Gamma}}{\boldsymbol{J_n}} \sum_{j\neq i}\Bigg(ln\left(1 + exp\left[\frac{\beta_{NumOfRoutes}}{\mu}\left(NumOfRoutes_j - NumOfRoutes_i\right)\right]\right) + ln\left(1 + exp\left[\frac{\beta_{AccessWalk}}{\mu}\left(AccessWalk_j - AccessWalk_i\right)\right]\right) + ln\left(1 + exp\left[\frac{\beta_{Train}}{\mu}\left(Train_j - Train_i\right)\right]\right) + ln\left(1 + exp\left[\frac{\beta_{StopLight}}{\mu}\left(StopLight_{\,j} - \right.\right.\right.$$

$$StopLight_i)\Big]\Big) + ln\left(1 + exp\left[\frac{\beta_{MTR}}{\mu}\left(MTR_j - MTR_i\right)\right]\right) + ln\left(1 + exp\left[\frac{\beta_{MTT}}{\mu}\left(MTT_j - MTT_i\right)\right]\right)\Bigg) \quad (5.2)$$

$$RR_i^{Model\,3} = \beta_{OtherWalk}.OtherWalk_i + \beta_{AccessWalk}.AccessWalk_i + \beta_{StopLight}.StopLight_i + \beta_{Train}.Train_i + \beta_{MTT}.MTT_i + \beta_{MTR}.MTR_i - \frac{\boldsymbol{\Gamma}}{\boldsymbol{J_n}} \sum_{j \neq i} ln\left(1 + exp\left[\beta_{NumOfRoutes}\left(NumOfRoutes_j - NumOfRoutes_i\right)\right]\right)_i \quad (5.3)$$

5.3.1 RRM Transit Stop Choice Models Estimation Results

The estimated parameters of the RRM models are presented in

Table 5.2. Four variables, including, *access walk time, stop lighting*, *Train,* and *MTT Strategy* are found to be significant in model-1. All these variables are also significant in the model-2. Moreover, two variables including, *stop lighting,* and *MTR Strategy* are found to be significant in model-2. The scale parameter μ in model-2 is very small indicating the presence of a high profundity of regret. However, both these models show poor values of the adjusted ρ^2, final log-likelihood and BIC. Therefore, model-3, which is a hybrid model of RUM and RRM, is developed. A hybrid structure can capture users' behaviour of evaluating an option based on the utility of a group of attributes and also regret level for another group of attributes. Consequently, model-3 shows a better model fit and significant parameter estimates. However, only one variable, *number of routes,* is found to be significant in the RRM setting and six variables in the RUM setting.

Table 5.2 Comparison of Estimated Parameters of RRM and RUM Transit Stop Choice Models

Explanatory Variables (β)	**RRM Model 1 P-RRM**	**RRM Model 2 μ RRM**	**RRM Model 3 (Hybrid)**
Other Walking Time	-	-	-0.038 (-3.59)
Number of Routes	-	0.004 (2.12)	0.0005 (2.68)*
Walking to Stop Time	-0.0004 (-12.72)	-0.034 (-10.35)	-0.198 (-21)
Stop Lighting	0.0063 (3.98)	0.01 (2.04)	0.444 (4.77)
Train	0.0778 (7.49)	0.005 (1.94)	2.3 (18.71)
MTT Strategy	0.0064 (4.11)	0.005 (1.42) **	0.736 (8.47)
MTR Strategy	-	0.014 (2.32)	1.46 (15.08)
μ	-	0.066 (3.29)	-
Model Statistics			
No. of Parameters	4	7	7
Initial Log-likelihood	-2938	-2938	-2938
Final Log-likelihood	-2629	-2802	-1982
Adjusted ρ^2	0.104	0.044	0.323
BIC	5287	5655	4014

t-statistics are provided within the paranthesis
estimated from RRM formation* * significant at 80% confidence level*

Table 5.3 presents a comparison between the MNL model (reported in Chapter 3) and the RUM-RRM hybrid model. The rate of substitution seems to be similar. However, some variables, including, *other walk time, stop lighting, MTT Strategy* and *MTR Strategy,* show a slightly higher rate of substitution. The pseudo marginal effect suggests an increased importance of *number of routes* variable as this poses a higher value in the RUM-RRM hybrid model. The other variables show similar effects on these two models.

Table 5.3 Rate of Substitution of the Stop Choice Models

Explanatory Variables	Rate of Substitution		Pseudo Direct Marginal Effect	
	MNL	RRM Model-3 (Hybrid)	MNL	RRM Model-3 (Hybrid)
Walking to Stop Time	1	1	0.025	0.026
Other Walking Time	0.17	0.19	0.005	0.005
Number of Routes	-0.30	*	0.009	0.013
Stop Lighting	-2.0	-2.2	0.060	0.061
Train	-11.8	-11.6	0.313	0.308
MTT Strategy	-3.3	-3.7	0.099	0.103
MTR Strategy	-7.1	-7.4	0.205	0.207

** Cannot be comparable as it is in the RRM settings*

5.3.2 RRM Stop Choice Prediction

An in-sample prediction is performed for the RRM model by calculating the choice probabilities of all the options. These results are compared with the results found using RUM modelling (see Chapter 3) and presented in Table 5.3. The RRM model shows an impressive hit rate[2] of 51.2%, significantly higher than the two RUM models (MNL and TT-AT). Moreover, the RRM model can identify about 87.3% of the chosen options within the top 5 predicted options, which is also significantly higher than the RUM models. However, if the top 10 options are considered, MNL model shows similar prediction like the RRM model. Similarly, the mean ranks and the standard deviations of the ranks are lower in the RRM model than the RUM models. Therefore, for the stop choice models for the SEQ passengers, the RRM model shows the best prediction followed by the MNL and TT models.

[2] Hit rate is defined in subsection 2.4.3

Table 5.4 Comparison of the Prediction Capabilities of the Stop Choice Models

Criteria	TT	MNL	RRM (Hybrid)
Total Cases	1237	1237	1237
Hit	616	618	638
Hit Rate (Rank 1)	49.8%	49.9%	51.6%
Within Rank 5 (%)	84.7%	86.0%	87.3%
Within Rank 10 (%)	92.7%	93.5%	93.9%
Mean Rank	3.53	3.36	3.21
Standard Deviation of the Ranks	5.48	5.32	5.14
Range	59	51	51
Minimum	1	1	1
Maximum	60	52	52

The hit rates are further investigated according to the choice set size, which is presented in Table 5.5. The RRM model shows better prediction potential when compared to the other models, especially when the choice set size is smaller than 50. The TT model shows better prediction capability when the choice set size is between 10 and 25. However, the MNL model shows better prediction capability when the choice set size is larger than 50. The prediction capability of the RRM model seems to be worse when the choice set size is larger than 50.

Table 5.5 Comparison of the Hit Rate According to the Choice Set Size

Choice Set Size	TT	MNL	RRM (Hybrid)
>50	33.3%	35.6%	32.8%
25-50	35.5%	35.0%	37.7%
10-25	45.9%	44.9%	45.2%
<10	59.1%	60.1%	62.9%
Total	49.8%	49.9%	51.6%

5.4 RRM Transit Route Choice Model

The RRM transit route choice models are also developed as a two-stage model, discussed in Subsection 4.2.1. Similar simulation methodology like the RUM route choice models is used for the RRM models where model parameters are estimated using 100 simulations. Similar to the RUM route choice models, ten different choice sets ranging from 5 to 50 are considered to see the variations across different choice sets. Similarly, alike RUM route choice models, estimation dataset (comprises 867 observations) is used for model estimation, and the prediction dataset (comprises 370 observations) is used to study the prediction capability of the models. Moreover, the predictions are made by using the same two-stage methodology, and for each prediction, 100 simulations are used. All these processes are performed using the programming package MATLAB 13.

For the route choice modelling, a sampling method has to be used since the number of alternatives for some observations from the path generation process is large. Therefore, the method proposed by Guevara et al. (2014), discussed in subsection 2.3.2.10 is used. However, other model specifications, including the RUM-RRM hybrid model, and path size correction models to account for the path commonality are also tested. Details of these modelling exercises are discussed in the following subsections.

5.4.1 Model 1: RRM with Sampling Corrections

In this model, the sampling corrections discussed in subsection 2.3.2.10 are used. Again, corrections for variable choice set size are also applied. The utility (regret) function of this model is given in equation 5.4.

$$\widehat{RR}_i^{Model\,1} = V_i = ln\,\pi(D_n|i) - \frac{\Gamma}{J_n}\sum_{j\neq i} w_{jn}\Big(ln\big(1 + exp\big[\beta_{TravelTime}\big(TravelTime_j - TravelTime_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{AccessWalk}\big(AccessTime_j - AccessWalk_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{Transfer}\big(Transfer_j - Transfer_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{Train}\big(Train_{\,j} - Train_{\,i}\big)\big]\big) + ln\big(1 + exp\big[\beta_{Bus}\big(Bus_j - Bus_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{MTT}\big(MTT_{\,j} - MTT_i\big)\big]\big)\Big) \quad (5.4)$$

Where, $\frac{\Gamma}{J_n}$ is the correction factor for unequal choice set size (see equation 2.22),

w_{jn} is the expansion factor (see equation 2.2.4), and

$ln\,\pi(D_n|i)$ is the correction factor for sampling.

The correction factor for sampling depends on the sampling protocol used in the calculation process. For this model, two sampling protocols are tested. The first one is termed as GS or Guevara sampling protocol (Guevara et al., 2014) discussed in subsection 2.3.2.10 (equation 2.26). The second one is the sampling protocol proposed in this study (termed as PS) for sampling correction (see 4.2.2 and 4.2.3). The detailed estimation results of this model are provided in **Appendix D**. The estimated parameter values (from the 100 iterations) of these two RRM models are presented as box plots in Figure 5.1.

Figure 5.1 shows that the parameters of the RRM (PS) models (compared in each choice set size) are less scattered than the parameters of the RRM (GS) models. Similarly, the variability in the estimation is more in the lower choice set size models than the higher choice set size models. Consequently, this suggests that the RRM (PS) models of higher choice set size provide consistent parameter estimates. The fluctuation in the parameter estimates for different choice set size is also reported by Guevara et al. (2014). Upon comparing the prediction capabilities (see subsection 5.4.4), it seems that the prediction capability did not improve much if the choice set size is more than 25. Therefore, the estimated parameters of the models for a choice set size 25 are presented in Table 5.6.

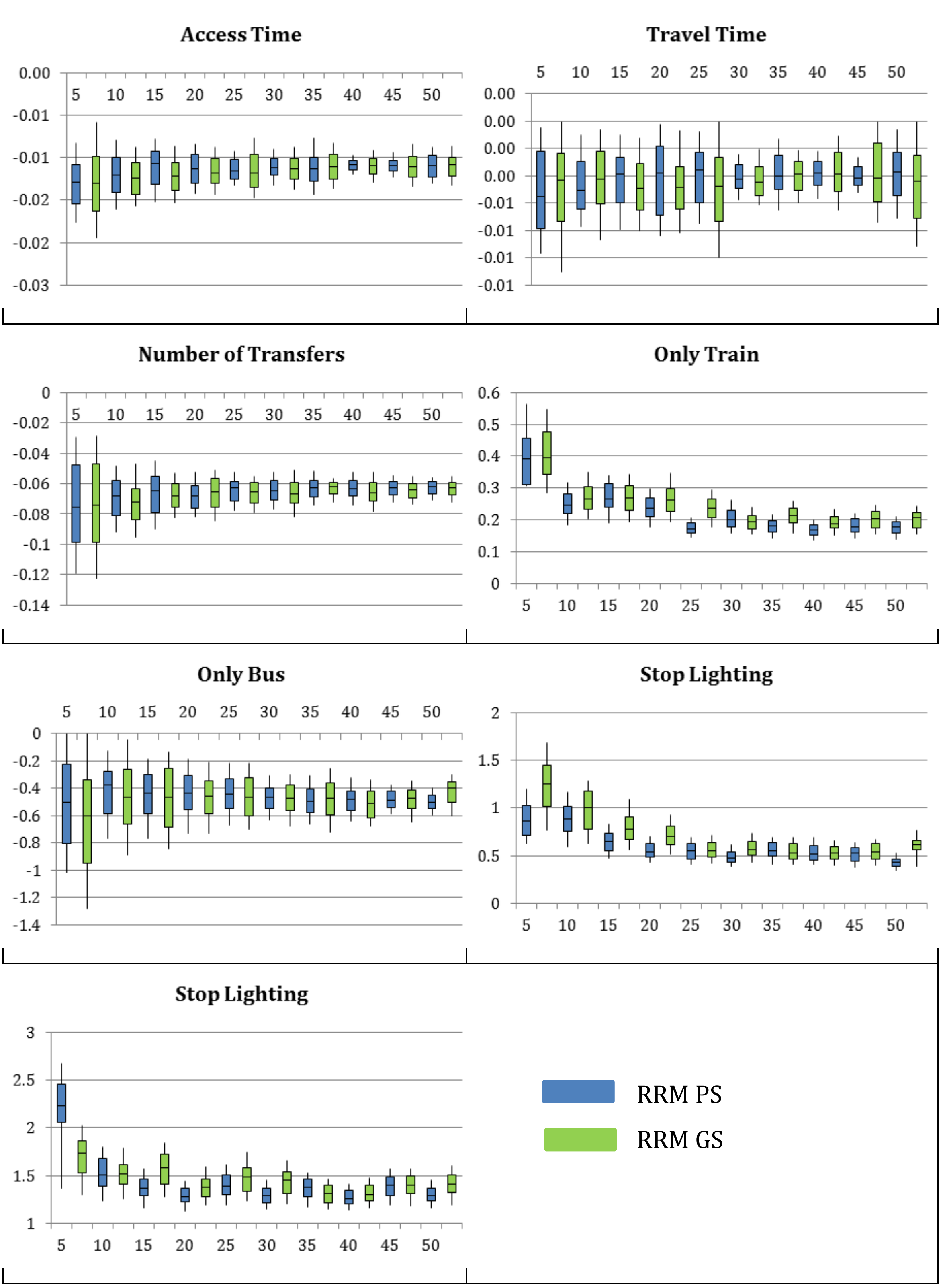

Figure 5.1 Boxplot of the Estimated Parameters of the RRM Models

Table 5.6 shows that not all the variables are significant at 95% confidence level. The t-statistics of *stop lighting* in the GS-25 model and *only bus* in the PS-25 model are found to be insignificant at 95% confidence level. However, these two variables are found to be significant in the RUM setting (see Table 4.4). Therefore, it can be inferred that these two variables can be best perceived as a utility maximiser. The BIC value suggests that the PS-25 model is superior to GS-25.

Table 5.6 Estimated Parameters of RRM Model 1

Explanatory Variables (β)	RRM (GS-25)		RRM (PS-25)	
	Mean of Coefficient	***Mean of t-test***	***Mean of Coefficient***	***Mean of t-test***
Access Time	-0.011	-14.1	-0.011	-12.0
Travel Time	-0.004	-6.0	-0.004	-12.1
Number of Transfer	-0.07	-13.6	-0.07	-11.7
Only Train	0.19	9.4	0.12	6.4
Only Bus	-0.02	-2.4	-0.02	-1.7*
Stop Lighting	0.008	1.3*	0.007	2.3
MTT Strategy	0.02	2.2	0.02	3.7
Model Statistics				
No. of Parameters	7		7	
Initial Log-likelihood	-2198		-2118	
Final Log-likelihood	-1640		-1558	
Adjusted ρ^2	0.254		0.264	
BIC	3328		3164	

** Significant at 80% confidence level*

5.4.2 Model 2: RUM-RRM Hybrid with Sampling Corrections

Model 1 shows that transit users do not consider all the variables in the RRM context. Therefore, the RUM-RRM hybrid models are developed to capture both the behaviours of utility maximisation and regret minimisation. In this stage few model specifications are initially tested. Finally, it is found that apart from *only bus* and *stop lighting*, if *MTT strategy* is modelled in an RUM setting, the explanatory power can be improved. This specification is presented in equation 5.5. In this model, the sampling protocols used in Model 1 are implemented. The detailed results of this exercise are reported in Appendix D. Table 5.7 presents two models with a choice set size of 25.

$$\widehat{RR}_i^{Model\ 2} = V_i = ln\,\pi(D_n|i) + \beta_{StopLight}.StopLight_i + \beta_{Bus}.Bus_i + \beta_{MTT}.MTT_i - \frac{\Gamma}{J_n}\sum_{j\neq i} w_{jn}\Big(ln\big(1 + exp\big[\beta_{TravelTime}\big(TravelTime_j - TravelTime_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{AccessWalk}\big(AccessTime_j - AccessWalk_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{Transfer}\big(Transfer_j - Transfer_i\big)\big]\big) + ln\big(1 + exp\big[\beta_{Train}\big(Train_{\,j} - Train_{\,i}\big)\big]\big)\Big) \tag{5.5}$$

Table 5.7 Estimated Parameters of RRM Model 2

Explanatory Variables (β)	RUM-RRM Hybrid (GS-25)		RUM-RRM Hybrid (PS-25)	
	Mean of Coefficient	***Mean of t-test***	***Mean of Coefficient***	***Mean of t-test***
Access Time	-0.01	-13.3	-0.011	-14.2
Travel Time	-0.0024	-8.3	-0.0025	-11.8
Number of Transfer	-0.06	-11.6	-0.07	-12.5
Only Train	0.11	8.8	0.12	9.2
Only Bus	-0.45	-4.0	-0.46	-2.5
Stop Lighting	0.26	2.5	0.28	2.4
MTT Strategy	0.98	8.1	0.99	9.3
Model Statistics				
No. of Parameters	7		7	
Initial Log-likelihood	-2210		-2133	
Final Log-likelihood	-1612		-1517	
Adjusted ρ^2	0.271		0.288	
BIC	3270		3082	

Modelled in RUM part

All the estimated parameters of Model 2 are found to be significant at a confidence level of 98%. Model 2 shows significant improvements from Model 1 in the BIC and adjusted ρ^2. As seen in Model 1, Model 2 also shows better BIC in PS-25 than GS-25.

Therefore, it can be inferred that the proposed sampling protocol constructs robust choice sets which eventually produce better estimates. Consequently, the proposed sampling protocol is used for further analysis.

5.4.3 Incorporating Path Size Corrections in the RRM Route Choice Model

The path size correction factor, $Corr_i$ is generated using equation 2.9 which is discussed in subsection 2.3.1.5.

$$\widehat{RR}_i^{Model\,3} = V_i = ln\,\pi(D_n|i) + \beta_{StopLight}.StopLight_i + \beta_{MTT}.MTT_i + \beta_{MTR}.MTR_i - \frac{\Gamma}{J_n}\sum_{j\neq i} w_{jn}\Big(ln\big(1+exp\big[\beta_{TravelTime}\big(TravelTime_j - TravelTime_i\big)\big]\big) + ln\big(1+exp\big[\beta_{AccessWalk}\big(AccessTime_j - AccessWalk_i\big)\big]\big) + ln\big(1+exp\big[\beta_{Transfer}\big(Transfer_j - Transfer_i\big)\big]\big) + ln\big(1+exp\big[\beta_{Train}\big(Train_{\,j} - Train_{\,i}\big)\big]\big) + ln\big(1+exp\big[\beta_{Corr}\big(Corr_j - Corr_i\big)\big]\big)\Big) \tag{5.6}$$

$$\widehat{RR}_i^{Model\,4} = V_i = ln\,\pi(D_n|i) + \beta_{Corr}.Corr_i + \beta_{StopLight}.StopLight_i + \beta_{MTT}.MTT_i + \beta_{MTR}.MTR_i - \frac{\Gamma}{J_n}\sum_{j\neq i} w_{jn}\Big(ln\big(1+exp\big[\beta_{TravelTime}\big(TravelTime_j - TravelTime_i\big)\big]\big) + ln\big(1+exp\big[\beta_{AccessWalk}\big(AccessTime_j - AccessWalk_i\big)\big]\big) + ln\big(1+exp\big[\beta_{Transfer}\big(Transfer_j - Transfer_i\big)\big]\big) + ln\big(1+exp\big[\beta_{Train}\big(Train_{\,j} - Train_{\,i}\big)\big]\big)\Big) \tag{5.7}$$

Table 5.8 shows that the path size correction factor in the PSC-RRM-$_{\text{RRM}}$ model is not significant at 95% confidence level. However, the path size correction factor is found to be significant in the PSC-RUM-RRM model. The sign of the correction factor is negative which means that the passengers prefer the alternatives which highly overlap. Usually, the PSC factor is positive in the route choice models for the auto as the travellers do not consider a route to be unique when there is a high overlap between routes. However, PSC factor in public transport route choice can be negative as found in other studies (Anderson et al., 2017; Tan et al., 2016). The adjusted ρ^2 value and the BIC value are also found to be improved in the PSC-RUM-RRM model. Therefore, from the model estimation results, the PSC-RUM-RRM model is found to the best RRM specification.

Table 5.8 Estimated Parameters of RRM Model 3 and 4

Explanatory Variables (β)	PSC-RRM-RRM (PS-25)		PSC-RRM-RUM (PS-25)	
	Mean of Coefficient	*Mean of t-test*	*Mean of Coefficient*	*Mean of t-test*
Access Time	-0.011	-12.4	-0.01	-12.7
Travel Time	-0.0025	-4.8	-0.0023	-4.2
Number of Transfer	-0.06	-11.2	-0.06	-11.6
Only Train	0.11	8.8	0.12	8.8
Only Bus	-0.52	-3.2	-0.56	-2.8
Stop Lighting	0.29	2.4	0.29	2.4
MTT Strategy	1.01	7.6	1.03	7.6
Path Size Correction	-0.02	-1.3 *	-0.53	-2.4
Model Statistics				
No. of Parameters	8		8	
Initial Log-likelihood	-2127		-2130	
Final Log-likelihood	-1516		-1514	
Adjusted ρ^2	0.287		0.289	
BIC	3087		3081	

Modelled in RUM part   ** Not Significant at 95% Confidence Level*

5.4.4 RRM Transit Route Choice Models Prediction Results

The prediction data of 370 observations is used to see the prediction capability of the RRM models. The prediction results are presented from Figure 5.2 to 5.4. All three figures clearly show that the models which used the proposed sampling protocol demonstrated higher prediction potentials. This means that the proposed importance sampling mechanism is better than the other protocol which considers simple random sampling in this stage. Figure 5.3 and Figure 5.4 also show the prediction capabilities of the choice models to detect the chosen alternative within the top 5 and top 10 alternatives respectively. The prediction capabilities for the two groups using the same sampling protocol are found to be similar, and the group of models using the proposed sampling protocol seems to be doing better.

However, among all considered models, the PSC-RUM-RRM model shows a slightly better prediction capability.

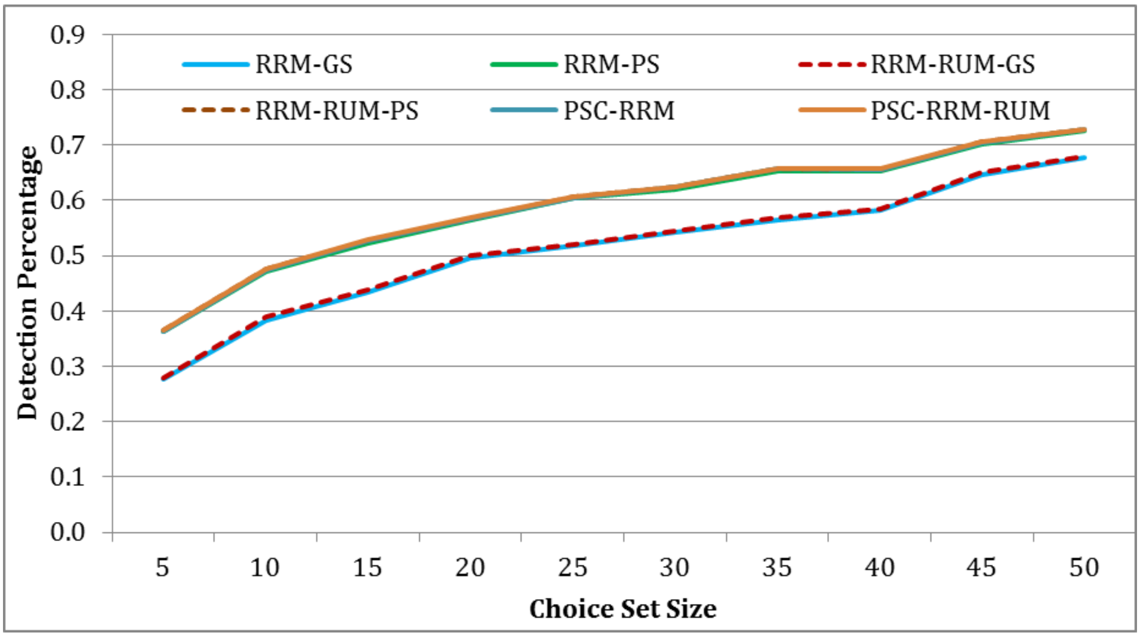

Figure 5.2 Prediction Capabilities of the RRM Models in the Choice Set Formation Stage

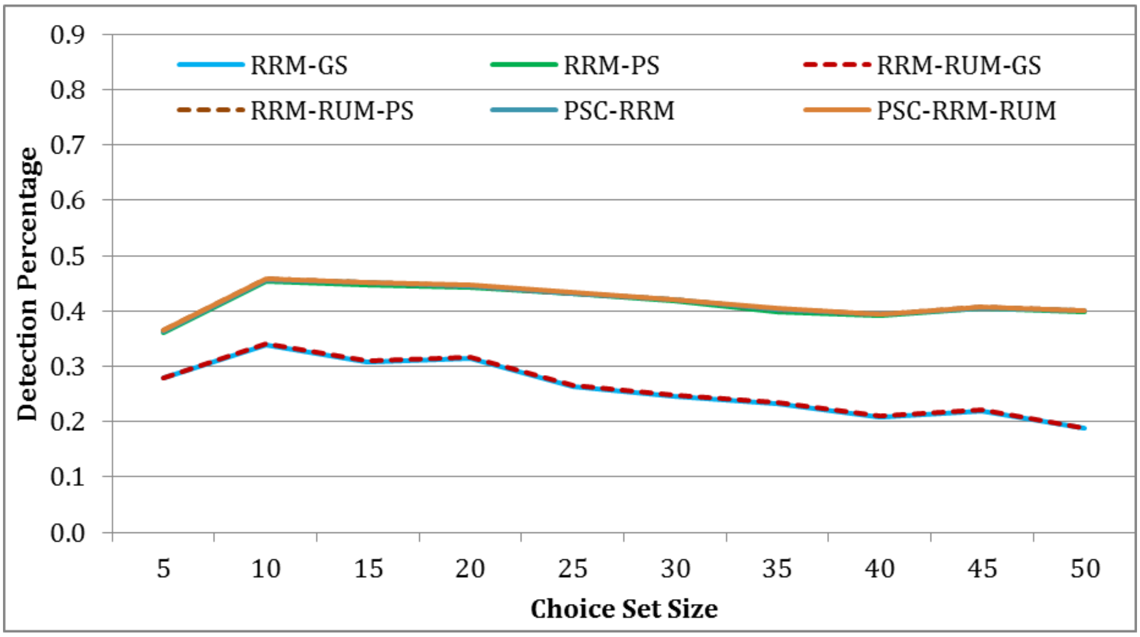

Figure 5.3 Prediction Capabilities (within the top 5) of the RRM Models in the Discrete Choice Modelling Stage

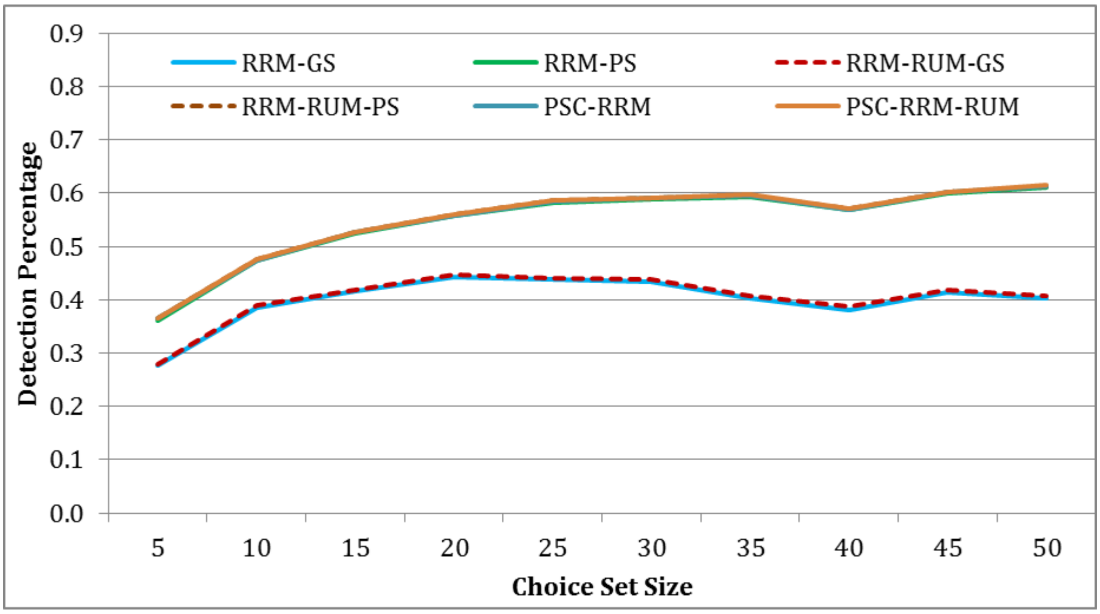

Figure 5.4 Prediction Capabilities (within the top 10) of the RRM Models in the Discrete Choice Modelling Stage

An interesting finding is that the detection capability (in the first stage, where the chosen alternative is to be included in the choice sets; see Figure 5.2) increases with the increase of choice set size. However, the prediction capability of the models (second stage, discrete choice models) seem to be decreasing (see Figure 5.3) or not significantly improving (see Figure 5.4) when the choice set size increases. Therefore, in this particular dataset, the optimal choice set size for the RRM models is somewhere between 20 and 30.

5.4.5 Comparison between Route Choice Models

The comparison of the rate of substitution across different model specifications is reported in Table 5.9. Model 1 from the RRM specifications is selected as all the variables in that model are estimated in an RRM setting. Similarly, the variables of the RRM part of Model 4 are also included. The results show that the passengers evaluate the variables differently across the models. The TT-TR-25 model has a nested structure depending on the strategy, which might have an effect on the rate of substitution as the substitution rates of *travel time* and *only train* are found to be higher when compared to the other two models.

Table 5.9 Comparison of the Rate of Substitution of the Selected Route Choice Models

Explanatory Variables	MNL-50	TT-TR-25	RRM-PS-25	PSC-RRM-RUM (PS-25)
Access Time	1	1	1	1
Travel Time	0.25	0.56	0.35	0.23
Number of Transfer	7.01	-	6.40	6
Only Train	-11.71	-15.47	-10.72	-12
Only Bus	2.71	2.37	2.11	*
Stop Lighting	-1.26	-	-0.60	*
MTT Strategy	-2.17	-	-2.04	*
Weekday **X** Walking Time	-	0.14	-	*

** Cannot be compared as these variables were in RUM part*

Comparing with the MNL-50 and the PSC-RRM-RUM (PS-25) models, the RRM-PS-25 model shows a higher importance on *travel time* in the RRM setting. In RRM-PS-25, passengers consider three minutes of travel time equivalent to one minute of access time, even though they consider four minutes of travel time equivalent to one minute of access time as per the other two models. Another variable, *only bus* is also found to be different in the RRM setting than in the RUM setting as the passengers try to avoid only bus routes more in the RRM setting than in the RUM setting. However, *number of transfers* and *stop lighting* are somewhat more critical in the RUM setting than in the RRM setting. Passengers are willing to walk extra when modelled using the RUM setting than in the RRM setting to avoid a transfer (around 12 seconds to 1 minute extra) or to access a stop with lighting (40 seconds extra). To avail an *only train* option, passengers act differently as Table 5.9 shows a wide range between 10.72 and 15.47. However, the MNL-50 and PSC-RRM-RUM (PS-25) models show a similar rate of substitution of 12 minutes of the access walk time. The rate of substitution found according to the RRM setting is comparable with the literature presented in Table 4.5.

The comparison of the pseudo direct marginal effect across different model specifications is reported in Table 5.10. The pseudo direct marginal effects of the route choice models indicate an interesting perspective. The RUM models show similar effects among the variables. However, some of the variables in the RRM model show significantly different effects than the RUM models. When compared to the RUM models, the variables in the RRM setting shows a lesser effect on the probability changes of the chosen alternatives. Out of the seven variables in the RRM model, only one variable, *number of transfer*, shows a slightly higher effect than the MNL model. However, the *only train* variable shows much lower (about one third) effect in the RRM setting.

Table 5.10 Comparison of Pseudo Direct Marginal Effect of the Selected Route Choice Models

Explanatory Variables	**MNL-50**	**TT-TR-25**	**RRM-PS-25**
Access Time	0.03	0.027	0.02
Travel Time	0.018	0.021	0.01
Number of Transfer	0.19	-	0.21
Only Train	0.32	0.29	0.12
Only Bus	0.10	0.10	0.07
Stop Lighting	0.04	-	0.02
MTT Strategy	0.07	-	0.06
Weekday **X** Walking Time	-	0.02	-

The prediction capabilities of the best RR M model (PSC-RUM-RRM) are compared with those of the best RUM model (MNL and TT-TR) presented in Chapter 4. Figure 5.5 shows the prediction capabilities of different choice set size models for the choice set generation stage. In this stage, the PSC-RUM-RRM and MNL models show similar results as both of them use the same sampling mechanism (proposed in Subsection 4.2.2 and 4.2.3). The detection capability of the NL models is better than all the other models as this considers different nests of the same choice set size allowing more options to be incorporated in the choice set compared with the other three models.

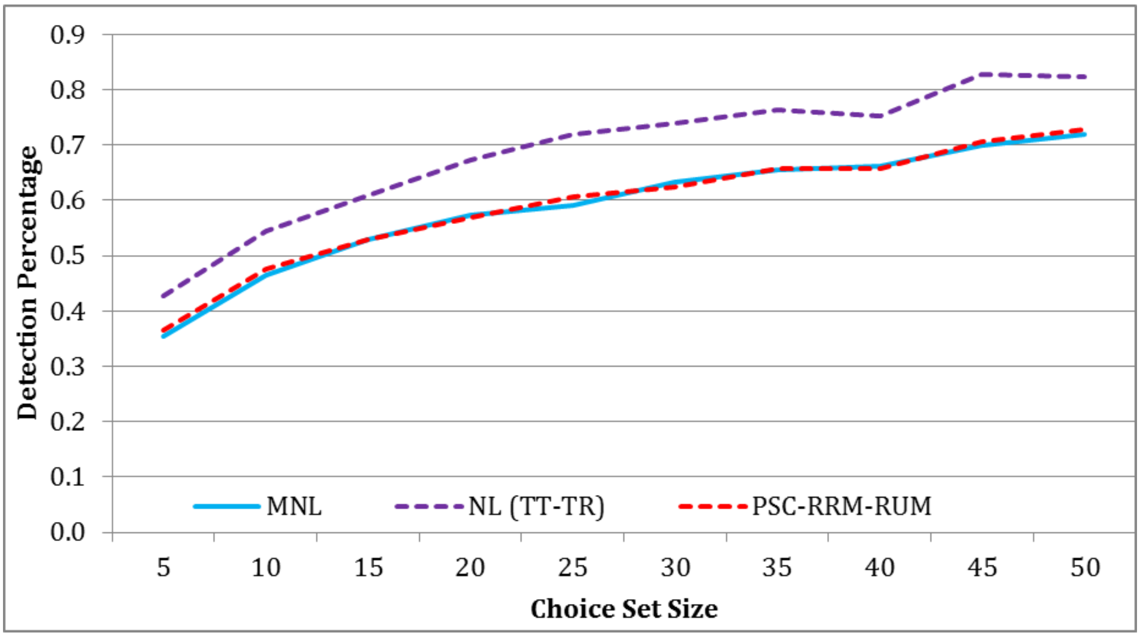

Figure 5.5 Prediction Capabilities of the Route Choice Models in the Choice Set Formation Stage

Figure 5.6 and Figure 5.7 show the prediction capabilities of the models to detect the chosen alternative within the top 5 and top 10 alternatives respectively. In this stage, the prediction capability of the discrete choice models can be assessed by calculating the choice probability of the options detected in the choice set formation stage. Figure 5.6 shows that the MNL models offer superior detection rates (within the top 5) when the choice set size is 30 or more. On the other hand, the PSC-RRM-RUM demonstrates a higher detection rate in the models where choice set size is 30 or lower. Similarly, if the prediction standard is lowered (considering the chosen option to be within the top 10 ranks), the PSC-RUM-RRM models demonstrate higher detection rates in all choice set sizes compared to the other route choice models.

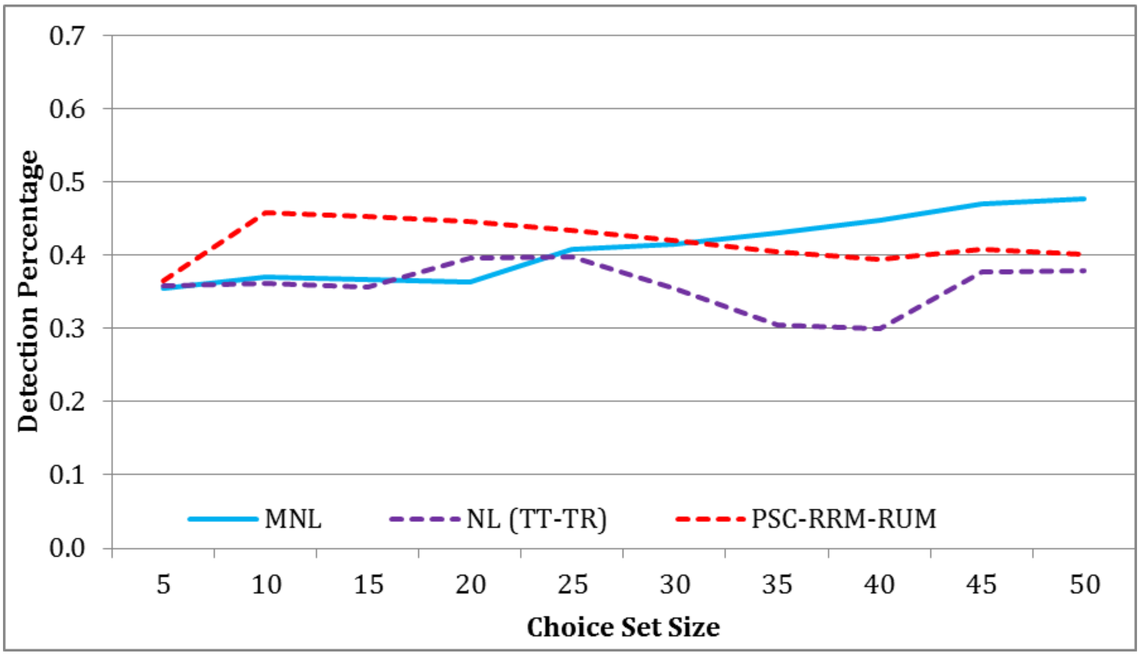

Figure 5.6 Prediction Capabilities (within top 5) of the Route Choice Models in the Discrete Choice Modelling Stage

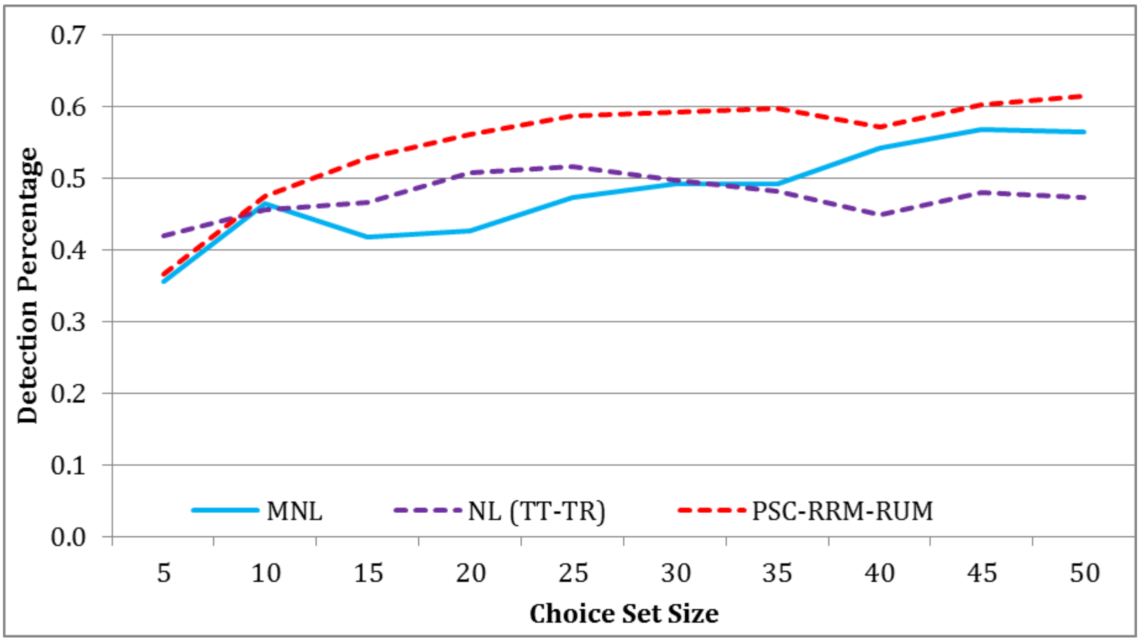

Figure 5.7 Prediction Capabilities (within top 10) of the Route Choice Models in the Discrete Choice Modelling Stage

5.5 Conclusions

In this chapter, different RRM model specifications are tested to model transit stop and route choice of the SEQ transit users. Moreover, the comparison between the

RUM and RRM models is also discussed. The route choice model addresses the sampling issue by developing sampling corrections and introducing a new sampling protocol. The path commonality issue is also tackled by introducing a path size correction factor. Furthermore, the correction for varying sampling size is also introduced. This study uses precise origin-destination data and a trip based shortest path algorithm to generate paths from a multimodal transit network to model stop and route choice behaviour. This is probably the first study which considers the RRM approach to model transit stop and transit route choice behaviour for a full-scale multimodal transit network using revealed preference data. Therefore, this study offers new insights into the RRM literature.

The findings of this chapter show that the RRM approach can be a prospective alternative to the commonly used RUM approach. However, a hybrid specification of RUM-RRM seems to predict better than the other model specifications, particularly in stop choice modelling context. Additionally, the prediction potential of the RUM-RRM specification is significantly better than the other RUM specifications. However, in the literature, the RRM-based models have been found to be equally or marginally better than the RUM-based models. One notable aspect may be that most of the models reported in the literature had a smaller choice set size compared to the choice set size considered in the current study.

The route choice models in this study use a sampling mechanism to construct the choice set for each observation. Consequently, a series of models are developed which also capture the influence of the choice set size. The results show that the RRM-based models can predict better (chosen alternative within the top 5 ranked alternatives according to the choice probability) than the RUM-based models when the choice set size is less than 30. The study also found similar outcomes for the stop choice model. However, the RUM-based models have a better prediction when the choice set size is larger than 30. Therefore, it is crucial to detect an optimal choice set size for a specific model specification. The effect of choice set size was also studied by Guevara et al. (2014). However, the universal choice set considered in their study was small (only 14 alternatives) when compared to this study where the choice set varies between 2 and 1925 (average size 35.6) alternatives.

Another important finding is that the proposed importance sampling mechanism of choice set formation proves to be a better solution than the simple random sampling mechanism proposed by Guevara et al. (2014). However, this is a critical issue which is also noted in Chapter 4 and will be addressed in the next chapter (Chapter 6).

The RRM method is a relatively new method which needs more research to enhance the literature. Although this study tested a wide variety of RRM specifications, there are the following limitations:

- The RRM models tested in this thesis do not capture the taste variation or preference heterogeneity of the respondents, which can be a potential future research.
- The study found that the size of the choice set affects the prediction potential of the RRM models, especially when the choice set size is larger than 30. However, this finding cannot be generalised and more comprehensive study should be conducted to investigate this issue.

6 EFFECTS OF SAMPLING IN CHOICE SET FORMATION

The choice modelling of transport routes, travel destinations and housing locations is characterised by high numbers of alternatives and therefore sampling is required to form choice sets. For many years, researchers have been using McFadden's uniform conditioning property with the help of a simple random sampling (SRS). However, researchers also employed other solutions of importance sampling where different sampling protocols were used. However, very few attempts were seen in the literature to investigate the strength of these sampling protocols and methods as well as their prediction capability. This study presents a comprehensive analysis of different sampling protocols and methods to understand the choice set composition and how well they can predict the observed choice. The results show that the choice set formation mechanism can significantly affect the models' prediction capability due to the composition of the choice set. The results show the models developed from the choice sets which consist of the relevant options can predict better than the models developed from the choice sets that consist of relatively less important options. This study can help modellers in choosing an appropriate sampling mechanism.

6.1 Introduction

Transit route choice is characterised by a high number of alternatives in many transit intensive urban areas. To model the route choice behaviour of these urban centres, modellers have to sample the alternatives that form the choice set. Researchers have mainly used McFadden's (1978) positive conditioning property of the MNL model discussed in equation 2.13. In this setting, a correction term is included in the utility function. However, if all the alternatives have equal selection probabilities (uniform conditioning property) (McFadden, 1977), the estimation on the choice set is performed in the same way as the full set of alternatives

(McFadden et al., 1978). Therefore, the correction for sampling bias in equation 2.13 cancels out, and the form will be like the simple MNL of equation 2.1.

Consequently, Ben-Akiva et al. (1985) pointed out two primary types of sampling: one considers uniform selection probabilities, and the other considers unequal selection probabilities or importance sampling. For many years, researchers (Kim et al., 2011; Pozsgay and Bhat, 2001; Simma et al., 2001) used the former with the simple random sampling (SRS) protocol. Recently, a number of studies (Ben-Akiva and Bowman, 1998; Bhat et al., 1998; Frejinger et al., 2009; Guevara and Ben-Akiva, 2013b) began using the latter sampling method.

The use of the SRS method is prominent for it being straightforward and not needing any correction term to be included in the model. However, the constructed choice set seldom reflects the users' original choice set. Furthermore, the choice set formed by this method may not contain the crucial alternatives (Daly et al., 2014). In other words, it is not an efficient technique as most alternatives for a given trip may have small choice probabilities (Ben-Akiva et al., 1985; Li et al., 2005). Furthermore, in many practical cases, it might not provide consistent parameter estimates (Daly et al., 2014) or the efficiency might suffer (Lemp and Kockelman, 2012).

A type of importance sampling known as stratified importance sampling is used by Park et al. (2012), Auld and Mohammadian (2011) and Li et al. (2005), where sampling is done in a different stratum and weighted according to certain criteria. Therefore, the probability of picking a reasonable choice set increases compared to the SRS method. However, to draw alternatives in each stratum, SRS is typically applied. This technique is criticized for being inefficient (Daly et al., 2014).

Typically, some alternatives are much more important than others and thus are most likely to be in the choice set. Schüssler and Axhausen (2009) showed that omitting relevant options in the choice set can lead to biased parameter estimates. Bliemer and Bovy (2008) showed that irrelevant alternatives in a choice set could affect the model quality and choice probability as the estimated parameters will be biased. Thus, a general approach with unequal q_j values (probability of sampling as in equation 2.13) is more efficient. To estimate q_j, prior information of the attractiveness of the options can be used. A simple deterministic model can be developed to estimate the attractiveness (Ben-Akiva et al., 1985). However, the actual functional relationship between the dependent variable (attractiveness) and the independent variables has to be established. Moreover, to build these relationships, extensive model calibration is needed.

Frejinger et al. (2009) provided a solution (see equation 2.16) to this problem by developing a random walk algorithm which calculates the probability of drawing the impedance of each link in the network. From a node, the algorithm selects the next node randomly from a Kumaraswamy distribution. In the selection process, the weight of the link is used which is calculated from the cost ratio between the shortest route and the shortest route that includes the considered link. This process iterates from the origin node to the destination node and extracts the probability of selecting each link. The probability of selecting a route corresponds with the product of the probability of the links associated with the route. Later, Zimmermann et al. (2018) used this method in a transit route choice scenario. This process was criticised by Vacca et al. (2015) for various reasons including a high computational time for the large-scale network, the creation of loops in the generated routes, and the destination node not being reached within a reasonable number of iterations.

Vacca et al. (2015) proposed a solution and applied it on a large-scale network where the paths were generated stochastically by using the Dijkstra shortest path algorithm. The proposed solution was simple to execute as it needed only one random number generator to get the link cost. The sampling probability was calculated from the difference between the cumulative distributions associated with the reference path travel time which is increased and decreased by an interval dx function. The dx function was calculated from the standard deviation of the travel times of the choice set, which was multiplied by a scale value (they tested this value ranging between 0.1 and 10^{-8}).

The solution of Vacca et al. (2015) is impressive and can be applied in a transit route choice context. However, this also requires substantial effort to calibrate the distribution and scale parameters. Lemp and Kockelman (2012) proposed an iterative method where the first iteration uses SRS among all available alternatives. In the next iteration, the choice probability is calculated from the estimates completed in the previous iteration and used as a priori information for the sampling probability (q_j). This study (Lemp and Kockelman, 2012) used a simulated dataset to see how the efficiency of parameter estimates changes in the MNL and Mixed MNL models.

Nerella and Bhat (2004) were the first to examine the effect of the sample size of alternatives on model performance. They designed some numerical experiments to observe the performance of the MNL and Mixed MNL models for which they used SRS. For sampling with MNL, their study recommended a minimum sample size of an eighth of the full choice set, and a desired sample size of a fourth of the full choice set. However, for sampling with inconsistent models like Mixed MNL, they suggest using one half or more samples. Daly et al. (2014) argued the effectiveness of real datasets instead of simulated data and the use of importance sampling rather than SRS as used in other research (Lemp and Kockelman, 2012; Nerella and Bhat, 2004). In this study (Daly et al., 2014), the efficiency and effectiveness of the Multivariate Extreme Value (MEV) models were evaluated by exploring various sampling protocols in an attempt to minimise the estimation error for a given computational burden. However, like the other study (Nerella and Bhat, 2004), they also used simulations.

From the literature, a few studies were found which investigated the effect of choice set size on model quality and performance. Table 6.1 provided a summary of work in the literature which investigates the effect of choice set size. Table 6.1 identifies three major fields where the effect of the choice set size has been examined. Bekhor et al. (2008) studied the composition of the choice set generated from the different path generation algorithms. They tested the choice set formation effect by generating route alternatives using a combination of link elimination (Azevedo, 1993) and link penalty (De La Barra et al., 1993) approaches for traffic assignment. They used the Sioux Falls and Winnipeg networks to study the performance of MNL-SUE and CNL-SUE models. The study found that the Sioux falls network, which is comparatively small, needed a few routes (about 10 per OD pair) to achieve convergence. However, the Winnipeg network, which is relatively large, needed around 50 routes per OD pair to converge. The study results suggested considering more realistic models like CNL-SUE instead of MNL-SUE for route choice modelling.

Table 6.1 Studies on the Effects of Choice Set Size

Network/Data	Path Generation Algorithm	Model	Sampling Protocols
Real	Bekhor et al. (2008)	Prato and Bekhor (2007)	This Study
Hypothetical		Bliemer and Bovy (2008)	
Simulation		Nerella and Bhat (2004) Guevara and Ben-Akiva (2013a, 2013b); Guevara et al. (2014)	

Prato and Bekhor (2007) examined the effect of the choice set composition by designing an experimental analysis. Their study found that the MNL modifications (C-Logit, PSL) showed more robustness in parameter estimates. Therefore, they suggested using choice situations with a large number of alternatives. They also suggested using nested or logit kernels for instances where the number of alternatives was small. Bliemer and Bovy (2008) studied the role of choice set composition in determining route choice probability. With the help of a simple hypothetical network, they tested the robustness of different model specifications including MNL, C-Logit, PSL, PSCL, PCL, and CNL towards choice set size. They found that all these models were sensitive towards the irrelevant alternatives in the choice set and the route overlap (MNL is not sensitive to route overlap). However, the CNL model showed somewhat robust estimates. The study suggested using the same size of the choice set for prediction as used for estimation. Finally, the study advised considering only relevant alternatives in the choice set while modelling. Guevara and Ben-Akiva (2013b) and Guevara and Ben-Akiva (2013a) tested the effects of the choice set size while sampling the MEV and logit mixture models respectively. Guevara et al. (2014) examined the effects of choice set size while sampling RRM models.

From the literature review, there has been no study which investigated the effects of different sampling protocols (like SRS and importance sampling) on the composition of the choice set. Consequently, this chapter aims to investigate this issue of sampling and choice set formation with the following objectives.

- Study the effect of sampling protocols on the choice set formation,
- Study the effect of sampling methods on the choice set formation and overall model performance,
- Study the effect of choice set size using different sampling protocols and methods, and
- Study the effects of prior information on model formation and performance.

6.2 Study Description

A total of five models are developed for the experiment as shown in Table 6.2. These models comprise two different groups depending on the sampling method. The first group considers the uniform selection probability method and uses a simple random sampling protocol to develop a model called SRS. The second group considers the importance sampling method and is divided into two sub-groups according to the availability of prior information. The first sub-group does not possess prior information and thus uses Fuzzy Logic (FL) to construct the choice set. The other sub-group uses prior information (in this case, an MNL model) to construct the choice set. Both sub-groups consist of two models depending on the sampling protocol (WR or WOR). The models of the first sub-group are called FL-WR and FL-WOR, and the other sub-group models are called MNL-WR and MNL-WOR. The MNL models use the estimated parameters of the FL-WR model as the prior information.

Table 6.2 Models Used in the Assessment

<table>
<tr><th>No.</th><th>Sampling Method</th><th>Modelling Scenario</th><th>Sampling Protocol</th><th>Name of the Model Group</th></tr>
<tr><td>1</td><td>Uniform Selection probability</td><td>N/A</td><td>SRS</td><td>SRS</td></tr>
<tr><td>2</td><td rowspan="4">Unequal Selection Probability / Proposed Importance Sampling</td><td rowspan="2">Prior information is not available</td><td>WR</td><td>FL-WR</td></tr>
<tr><td>3</td><td>WOR</td><td>FL-WOR</td></tr>
<tr><td>4</td><td rowspan="2">Prior information is available</td><td>WR</td><td>MNL-WR</td></tr>
<tr><td>5</td><td>WOR</td><td>MNL-WOR</td></tr>
</table>

6.2.1 Sampling Protocols

This study investigates three sampling protocols: simple random sampling (SRS), sampling without replacement (WOR), and sampling with replacement (WR). In the SRS protocol, the choice set is developed by picking n alternatives (size of the choice set) randomly from the universal choice set. In the WOR protocol, first the chosen alternative is included in the choice set; later, (n-1) unchosen alternatives j are included in the choice set with a sampling probability q_j by making a separate draw for each alternative. In the WR protocol, initially (n-1) unchosen alternatives are sampled with replacements giving each alternative a probability of q_j and then the chosen alternative is included. Finally, the duplicate sampled alternatives are deleted. Details of this process are discussed in subsections 4.2.2 and 4.2.3.

6.2.2 Sampling Methods

The study tested three sampling methods. The first one is the uniform selection probability where the SRS protocol is used. The second method is based on the scenario where no a priori information is known about the attractiveness of the alternatives. The sampling method proposed in subsections 4.2.2 and 4.2.3 is used in this regard. The third method is based on the scenario where there is pre-existing information about the attractiveness. The assumption of this scenario is: there is a pre-existing basic model that can predict the choice probability of the alternatives. This scenario is tested to see whether an existing model can be improved by the proposed sampling methods as is shown in the literature (Lemp and Kockelman, 2012). For this, the MNL model developed in Chapter 4 is used. The choice probability (or attractiveness) of the sampling of the alternatives is calculated using the choice probability from the model. Two models of the last two methods are developed from each using two sampling protocols. Consequently, a total of five models are developed. A summary of these models is given in Table 6.2.

6.2.3 Assessment Process

The study has two main parts: model estimation and prediction. A portion (70%) of the data is used to estimate the models while the remaining portion (30%) is used for the prediction exercise. As the process requires random sampling, it needs multiple runs to achieve parameter stability. Therefore, to estimate the parameters, a total of 100 runs are performed for each model and each choice set size. The choice sets are stored separately in each run to investigate their composition. The mean value of the estimated parameters is stored and used to determine the ranks by calculating the choice probability of each alternative for each case. After determining the ranks, the composition of the choice sets (stored in the estimation process) is evaluated. The prediction part intends to assess the capability of the models to predict the chosen alternative correctly. Therefore, in the choice set formation process (discussed in subsections 4.2.2 and 4.2.3) of the prediction part, the chosen alternative is not included separately. Consequently, the total number of samples is equal to the choice set size.

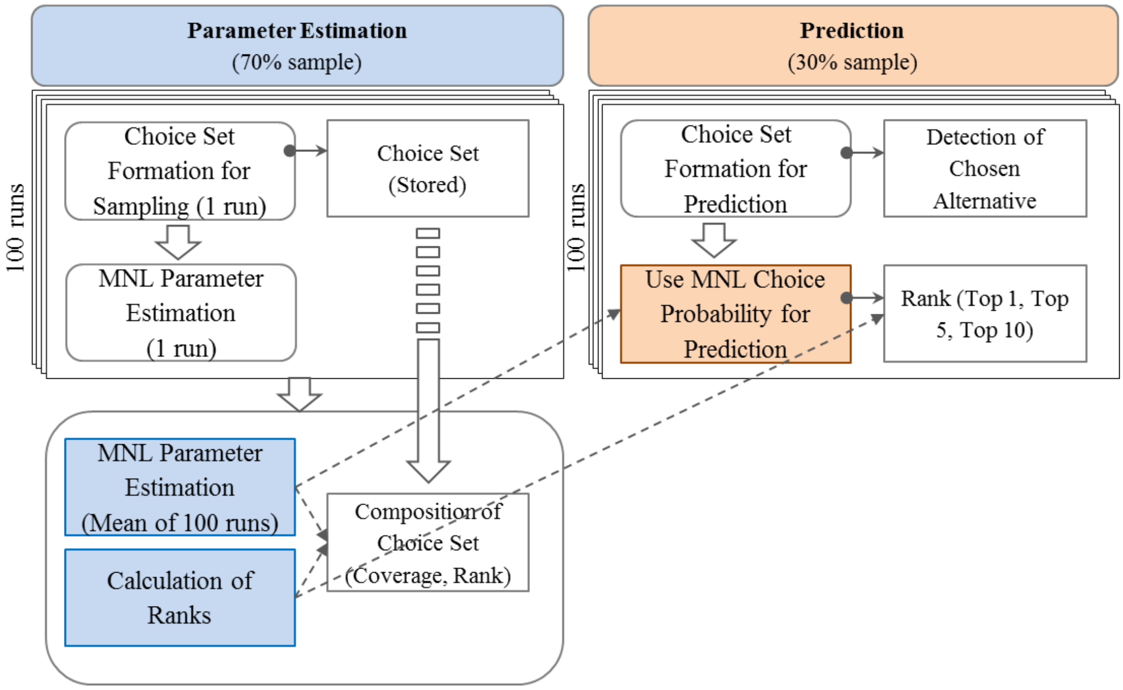

Figure 6.1 Assessment Process of the Study

6.2.4 Assessment Criteria

The overall performance[3] of the models was assessed based on two main aspects: 1) the composition of the choice set, and 2) the prediction capability of the associated model. The composition of the choice set was evaluated based on two main criteria: *probability coverage*[4], and *rank*. *Probability coverage* is the expected cumulative choice probability covered by the sampled alternatives (Daly et al., 2014). A sub-criterion, *coverage per alternative*, is assessed to offset the effect of the number of samples in the choice set. The *rank* criterion has three sub-criteria: *mean of the ranks*, *detection of the top 5 alternatives,* and *detection of the top 10 alternatives*. *Mean of the rank* refers to the mean value of the ranks of all the

[3] Usually, model performance is based on the computation time taken to estimate the parameters. However, in this chapter performance is assessed according to the other aspects.
[4] This "*probability coverage*" measure is presented by Daly et al. (2014) where they termed this as "coverage". Since the term "coverage" overlaps with the "coverage" measure discussed in subsection 2.4.3, proposed by Ramming (2002), the "coverage" measure of Daly et al. (2014) is renamed as "*probability coverage*".

alternatives in the choice set. Detection of the top 5 (or 10) alternatives refers to the percentage of times each model can detect all the top 5 (or 10) ranked alternatives in the choice set.

The prediction capabilities of the models are assessed in two ways: 1) the rate of detection of the chosen option in the choice set formation stage, and 2) the rank of the chosen alternative. The first aspect intends to assess the capability of the sampling methods and protocols to detect the chosen option in the choice set formation stage. The second aspect evaluates the strength of the MNL models and eventually the sampling methods and protocols as the MNL models are developed from choice sets generated using specific sampling methods and protocols.

6.3 Discussion of the Findings

6.3.1 Composition of the Choice Set

The comparison of the probability coverage, shown in Figure 6.2, demonstrates that the choice set of an MNL-WOR model covers the relevant alternatives better than the other models. The MNL-WR model is the second-best in this regard, followed by the FL-WOR model. The SRS model shows the worst probability coverage and is far behind the other models. All models show a gradual increase with an increase in the choice set size. However, the rate of increase seems to be higher in the models that have the WOR protocol. Similarly, the overall performance (of probability coverage) of the WOR protocol seems better than the WR protocol. The reason behind this is a big difference in the choice set size between these protocols.

Consequently, having more alternatives in the choice set derived by WOR protocols certainly increases the probability coverage of these models. However, if the effect of the average choice set size is normalised, as we can see in Figure 4B, the WR protocol performs better than the other protocols (also see Table 6.3). Another interesting finding from Figure 6.3 is that the probability coverage of all the models, except the SRS, seems to be relatively higher for smaller choice set sizes than larger choice set sizes. From Figure 6.2, it is also clear that the MNL models outperform the other models irrespective of the choice set size. Further, from Table 6.3, it is confirmed that choice sets developed by the MNL model groups provide the highest number of competitive alternatives.

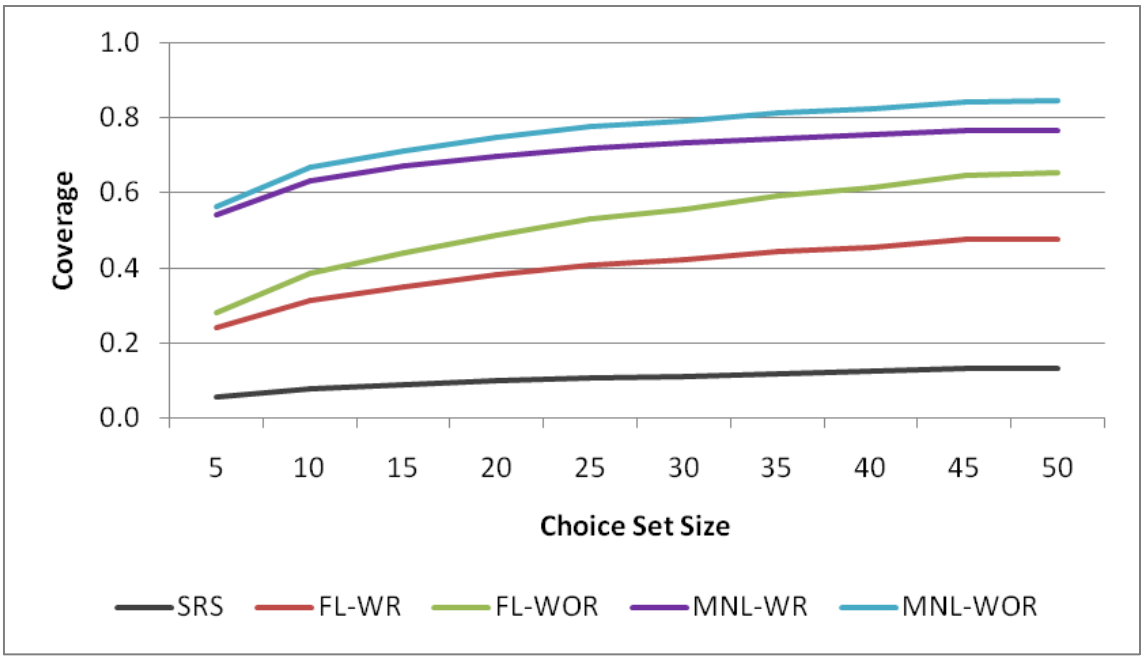

Figure 6.2 Comparison of Probability coverage for Different Sizes of Choice Set

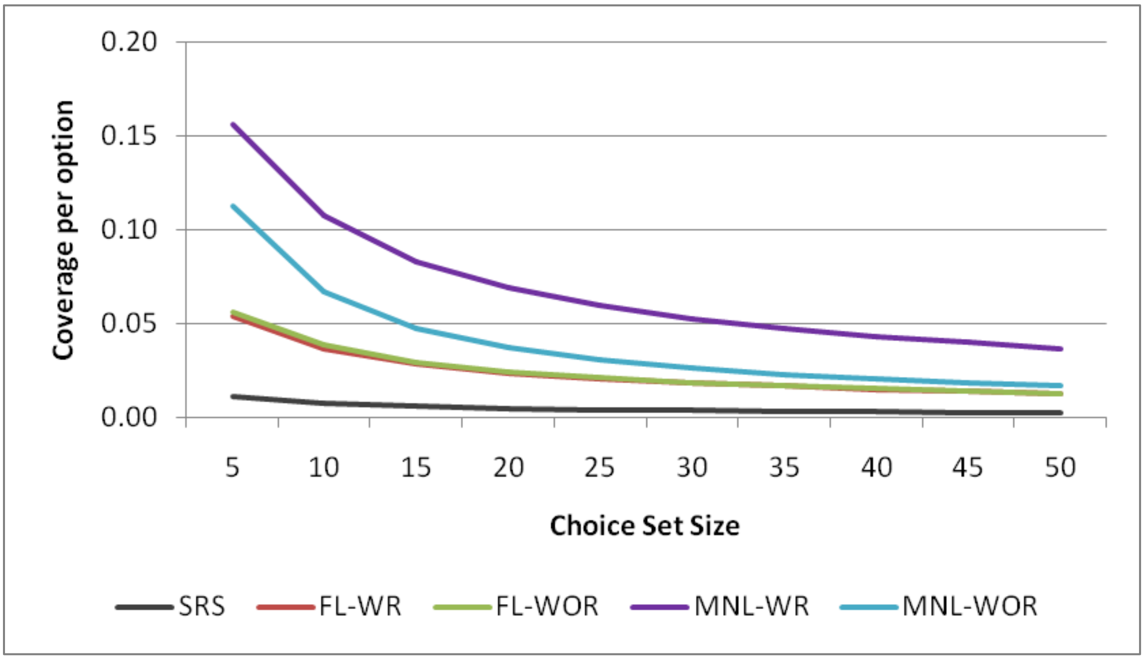

Figure 6.3 Comparison of Probability coverage per Alternative for Different Sizes of Choice Set

Table 6.3 Summary of Results of the Composition of Choice Set

Models	Average Size of the Choice Set	Probability coverage	Coverage per Alternative	Mean of the Ranks	Detection of the Top 5 Alternative	Detection of the Top 10 Alternative
SRS	27.50	10.4%	0.38%	36.86	46.8%	51.3%
FL-WR	21.04	39.6%	1.88%	38.31	40.0%	41.8%
FL-WOR	27.50	51.8%	1.88%	35.46	50.4%	54.8%
MNL-WR	12.67	70.2%	5.54%	14.26	77.8%	63.4%
MNL-WOR	27.50	75.8%	2.76%	18.47	88.6%	89.1%

The ranks of the alternatives are determined as discussed in subsection 6.2.3, where the MNL-WR model with a choice set size of 50 is used as a base for the comparison. Figure 6.4 and Table 6.3 show the results of these comparisons. Figure 6.4 shows the mean of ranks of the choice sets for the different models, where it identifies two different groups of models. One group (MNL models) is characterised by lower (better) means as opposed to the other group (SRS and FL models, which shows a higher mean of ranks. For all models, the mean of rank increases with an increase in the choice set size, while in the MNL models the rate of increase is lower than in other models. The outstanding performance of the MNL models can be attributed to the methodology (importance sampling and prior information) adopted for these models. However, it is difficult to identify a better sampling protocol in this regard as the WOR protocol provides better results with the FL method while WR seems to be better with the MNL models.

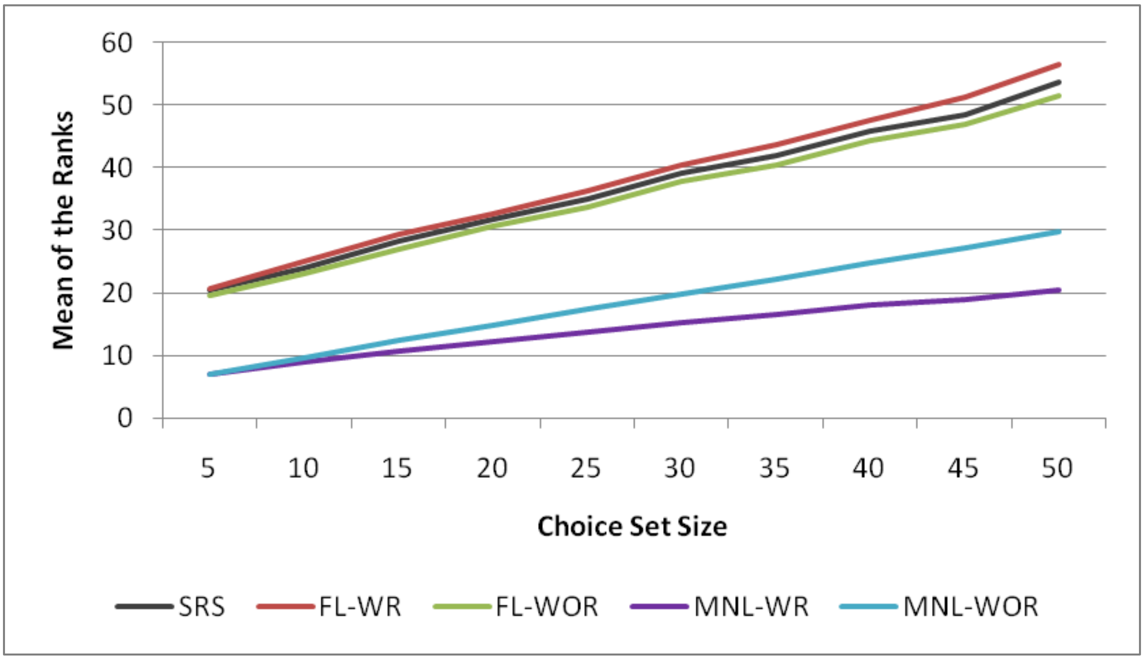

Figure 6.4 Comparison of Mean of the Ranks

An analysis of the presence of top-ranked alternatives in the choice sets is presented in Figures 6.5 and 6.6 and Table 6.3. Figure 6.5 shows the percentage of the top 5 ranked alternatives, while Figure 6.6 shows the percentage of the top 10 ranked alternatives in the choice sets. From Figures 6.5 and 6.6, the MNL-WOR model shows superior performance in both measurement criteria. The second-best model seems to be the MNL-WR model, which confirms the superiority of the MNL model groups in this regard. Another observation from Table 6.3 is that the SRS models yield better results than the FL-WR models regarding the presence of top alternatives in the choice set. However, FL-WOR models yield better results than SRS models. Furthermore, from Table 6.3, models with the WOR protocol seems to be better than the models with the WR protocol.

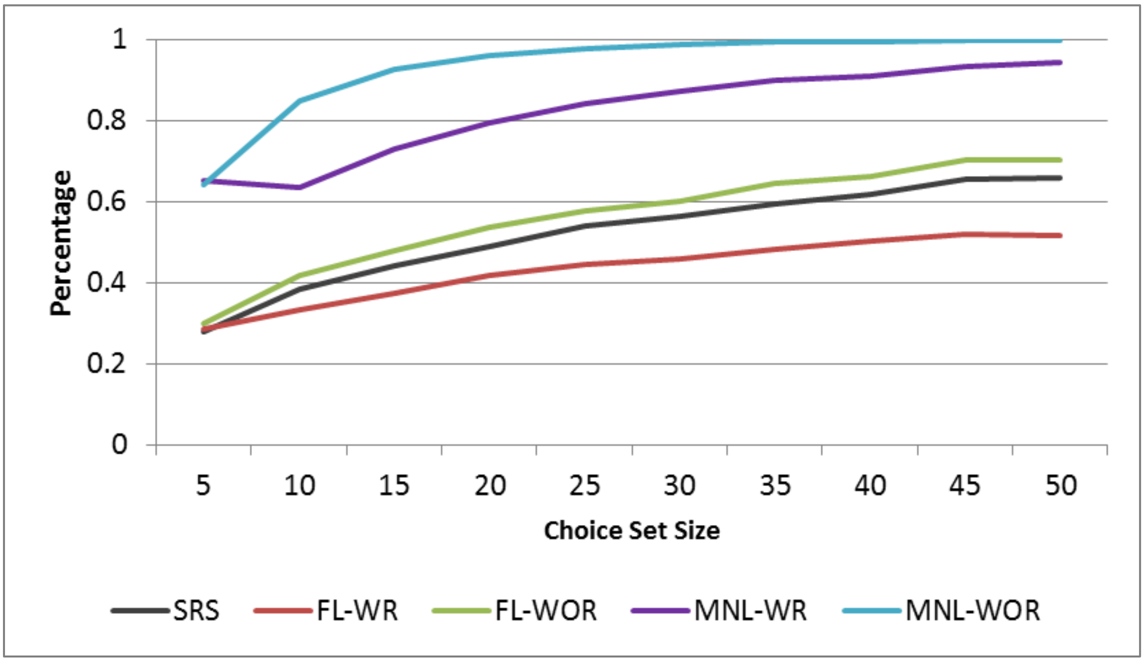

Figure 6.5 Presence of Top 5 Alternatives in the Choice Set

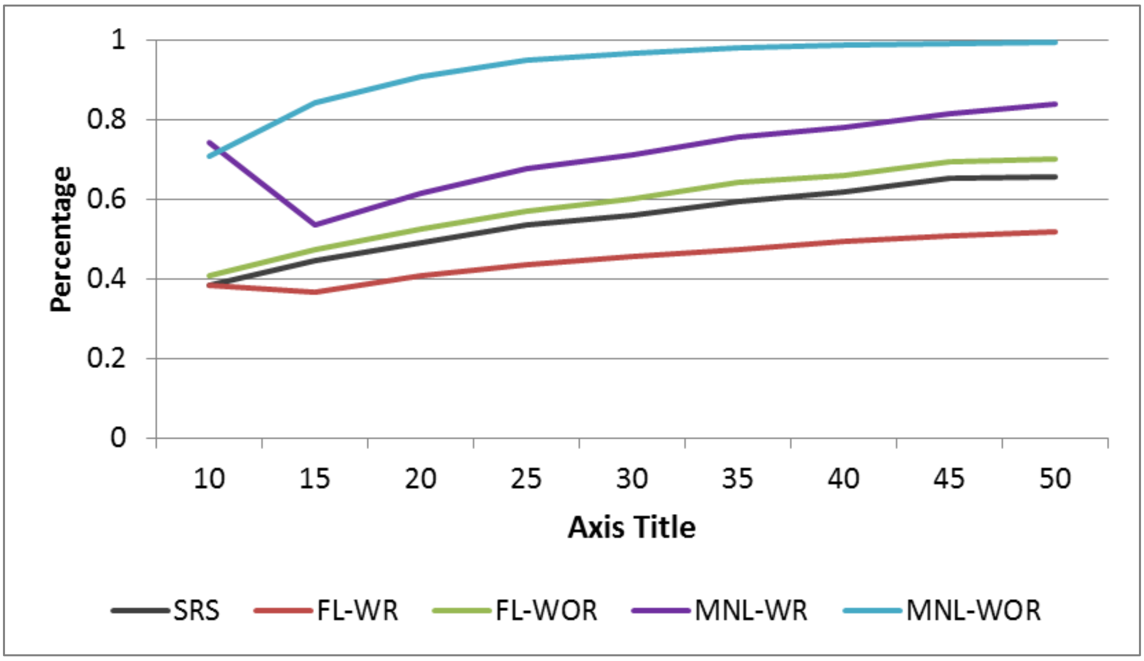

Figure 6.6 Presence of Top 10 Alternatives in the Choice Set

Finally, the analysis of this section confirms that as a modelling method the proposed importance sampling provides the choice set with more competitive alternatives compared to the uniform selection method, and the choice set will be better if prior information can be used.

6.3.2 Prediction Capabilities

The prediction is performed using the remaining 30% observations. Here, similar to the estimation process, the prediction is made in two stages (see Figure 6.1). The first stage uses the choice set formation mechanism and the second stage uses the estimated model parameters to predict the choice probabilities of the alternatives.

6.3.2.1 Prediction in the Choice Set Formation Stage

An overall comparison of the detection of the chosen alternative in the choice set formation stage is shown in Figure 6.7. The MNL-WOR model is the best in this regard, showing a detection rate of 60%, followed by the MNL-WR model (detection rate of 54%). If the original choice set is much larger (at least double or triple) than the sampled choice set size, the predicting capability of the models is decreased. However, the rate of decrease is higher in the SRS model and lower in the MNL-WOR model. Figure 6.8 shows that the detection rate increases with an increase in the choice set size. Furthermore, Figure 6.8 shows that the MNL-WOR model has superior performance, especially when the choice set size is small. The MNL-WR model exceeds the performance of the MNL-WOR model when the choice set size is larger than 40. The FL-WOR and SRS models show similar performances.

Table 6.4 presents a comparison of the prediction capability for the different models in the choice set formation stage. The comparisons are based on the scenarios where the ratio between the choice set size and the number of alternatives increases. If we consider all the cases where the original choice set is at least twice the sampled choice set size, the SRS and FL-WR models seem worse as these models can detect the chosen alternative only 27.1% and 26% of the time respectively compared to the MNL-WOR (53.7%) and MNL-WR (37.4%) models. The FL-WOR models seem to be better than the SRS models. If we increase the ratio between the choice set size and the number of alternatives by considering cases where the alternatives are at least three times the desired choice set size, the performance of the models seems to be the same with the MNL-WOR models being the best followed by the MNL-WR, FL-WOR and SRS models. Similar results are evident when we further increase the ratio between the choice set size and the number of alternatives by considering the cases where the desired choice set size is between 5 and 20. However, the prediction percentages seem to decrease with the increase of the ratio between the choice set size and the number of alternatives. From Table 6.4, the WOR protocol seems to be better than the WR protocol, the importance sampling method appears to be better than the uniform selection method, and that the use of a priori information seems to give a better prediction potential.

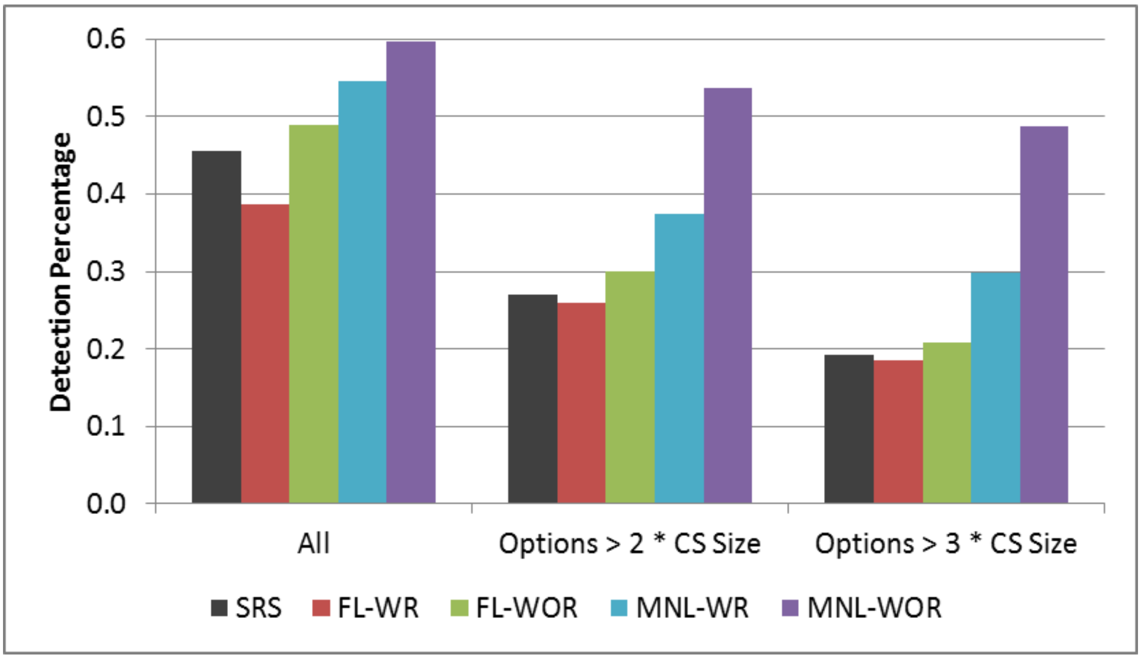

Figure 6.7 Predicting the Chosen Alternative in the Choice Set Formation Stage: According to Option Size

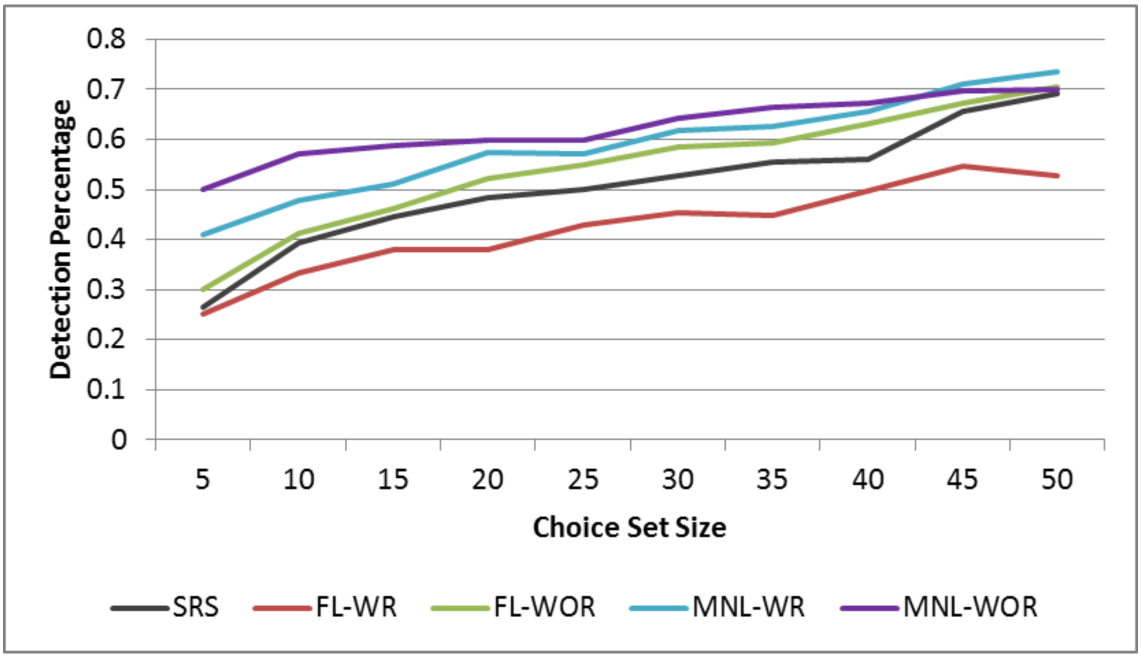

Figure 6.8 Predicting the Chosen Alternative in the Choice Set Formation Stage: According to Choice Set Size

Table 6.4 Prediction Capability in the Choice set Formation Stage

Models	**Detection of the Chosen Alternative (%)**		
	All Cases		**Choice set size 5 to 20**
	Alternatives > 2 x Choice Set Size (Total Cases 900)	**Alternatives > 3 x Choice Set Size (Total Cases 572)**	**Alternatives > 3 x Choice Set Size (Total Cases 480)**
SRS	27.1%	19.2%	18.0%
FL-WR	26.0%	18.5%	17.6%
FL-WOR	30.0%	20.9%	19.6%
MNL-WR	37.4%	29.9%	29.1%
MNL-WOR	53.7%	48.7%	45.6%

6.3.2.2 Prediction in the Second Stage (MNL Choice Probability)

When the chosen alternative is predicted successfully in the choice set formation stage it is further assessed in the second stage. In this stage, the choice probability of the alternatives picked in the choice set formation stage is calculated and ranked. The rank of the chosen alternative is considered for evaluation in three cases: 1) whether the chosen alternative is top-ranked, 2) ranked within the top 5, and 3) ranked within the top 10. This assessment also considers two scenarios: 1) cases where alternatives are more than double of the choice set size, and 2) cases where alternatives are more than triple the choice set size.

Figure 6.9 shows the detailed prediction results of the models according to the choice set size. The SRS models seem to be best according to the first criterion by better predicting the chosen alternative as the top-ranked alternative. The FL-WR model seems to be the second-best. However, the performance of the other models in this aspect seems to be very poor, especially when the choice set size increases. The MNL-WR and MNL-WOR models perform similarly to SRS when the choice set size is 5. However, the performance drops dramatically with the increase in choice set size. Although the SRS and FL-WR models show the worst performance in the first stage of prediction, the choice models developed from them show an excellent prediction potential of detecting the chosen alternative correctly in the second stage of prediction. This might be attributed to the weak composition of their choice sets as discussed in subsection 6.3.1. In these cases, their choice sets cannot include many competitive alternatives which makes it easier to detect the chosen alternative as the utility of chosen alternative will be comparatively higher. However, the situation is opposite for the MNL-WOR or MNL-WR models as they can better detect the competitive alternatives in the choice set.

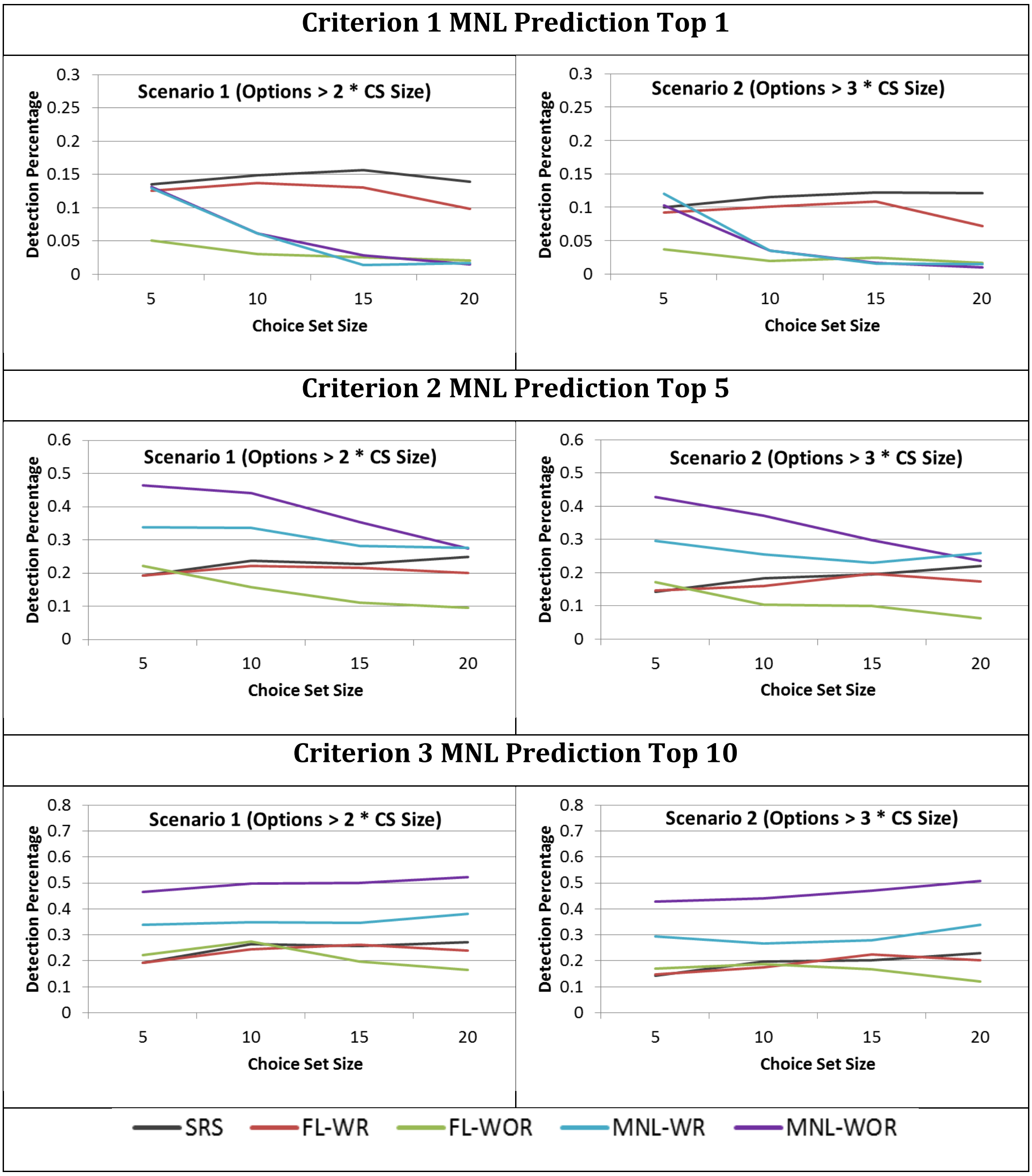

Figure 6.9 Prediction Capabilities of Different Models in Different Choice Set Size

Interestingly, in the second criteria, the MNL-WOR and MNL-WR models show better results for smaller choice sets in all the scenarios. However, their performance decreases when the choice set size increases. On the contrary, the SRS and FL-WR models show a steady performance for all choice set settings. Finally, in the third criteria, the MNL-WOR model seems to be the best in all the scenarios followed by the MNL-WR model. The effect of the choice set size is not that prominent in this criterion.

6.3.2.3 Summary of the Prediction

The prediction results are summarised in Figure 6.10 to Figure 6.12 by showing the overall performance of the models in three different scenarios and criteria. Considering all the cases, the MNL-WOR and MNL-WR models show superior performance as they can include the chosen option more often in the choice set formation stage and thus have an advantage over the other models. However, they perform poorly at predicting the chosen option in the MNL stage. Conversely, the SRS and FL-WR models perform relatively poorly in the choice set generation stage while their performance in the MNL stage is far better than the other models as they can correctly predict the chosen option as the best option. Almost the same pattern is seen in the other two scenarios. However, the MNL-WOR model is a much better predictor in the last two criteria.

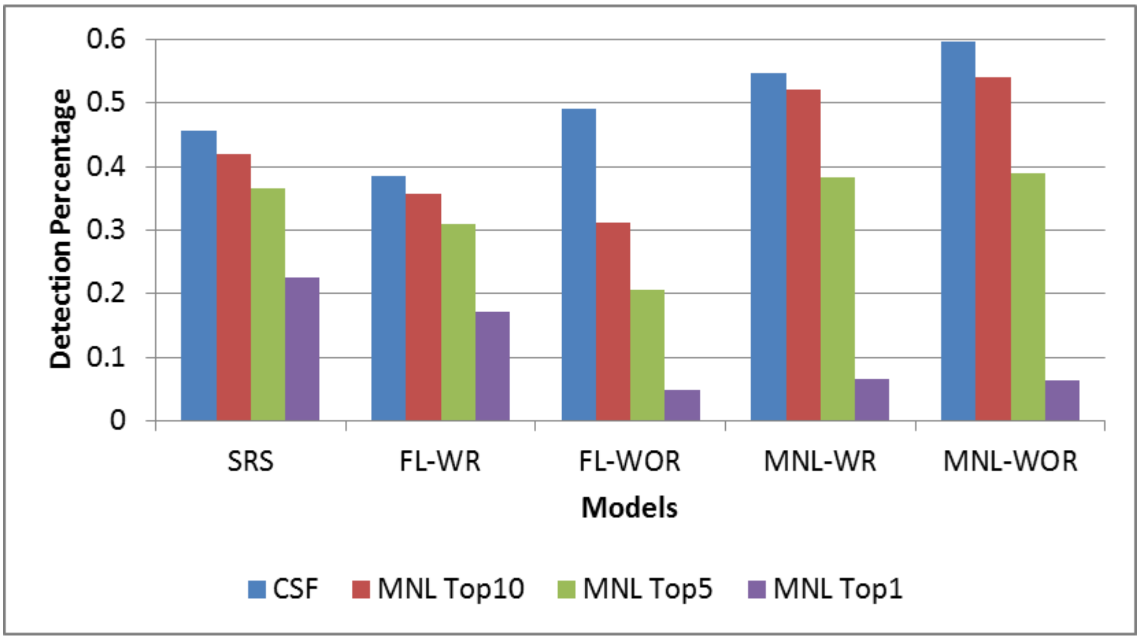

Figure 6.10 **Prediction Capabilities of the Models: Scenario 1 (All Cases)**

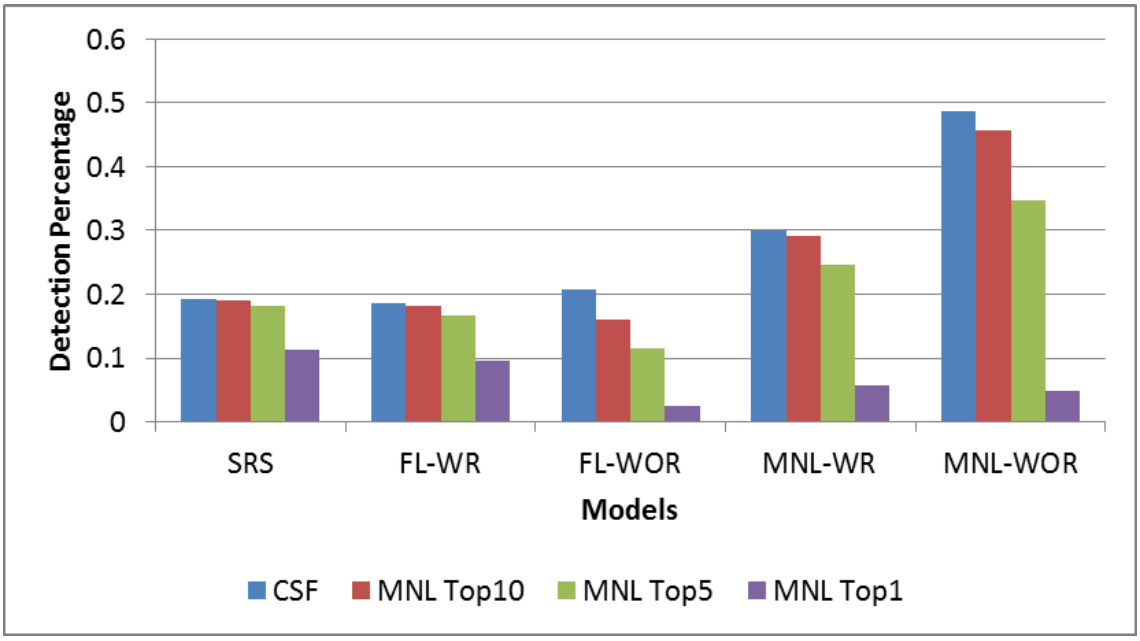

Figure 6.11 Prediction Capabilities of the Models: Scenario 2 (Options > 2 * CS Size)

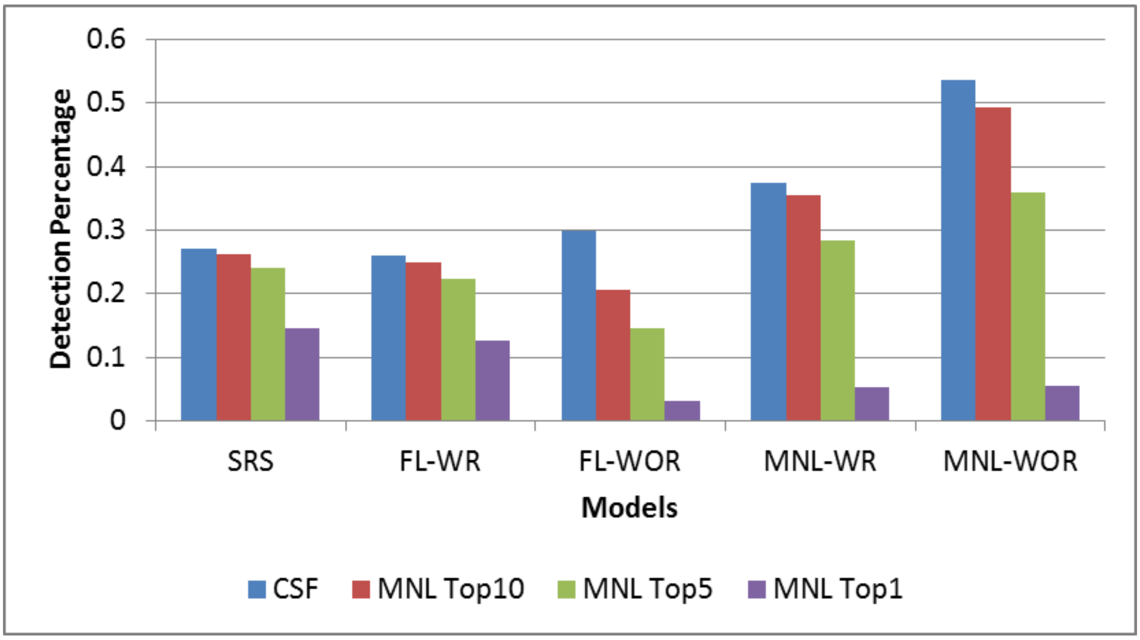

Figure 6.12 Prediction Capabilities of the Models: Scenario 3 (Options > 3 * CS Size)

Figure 6.13 to Figure 6.15 summarises the prediction capabilities according to the methods. In all the scenarios, the MNL models are better than the other methods for the last two criteria. Furthermore, if the options are increased (scenario 2 and 3), MNL outperforms the other methods. However, in all scenarios, SRS is dominant when the first criterion is considered.

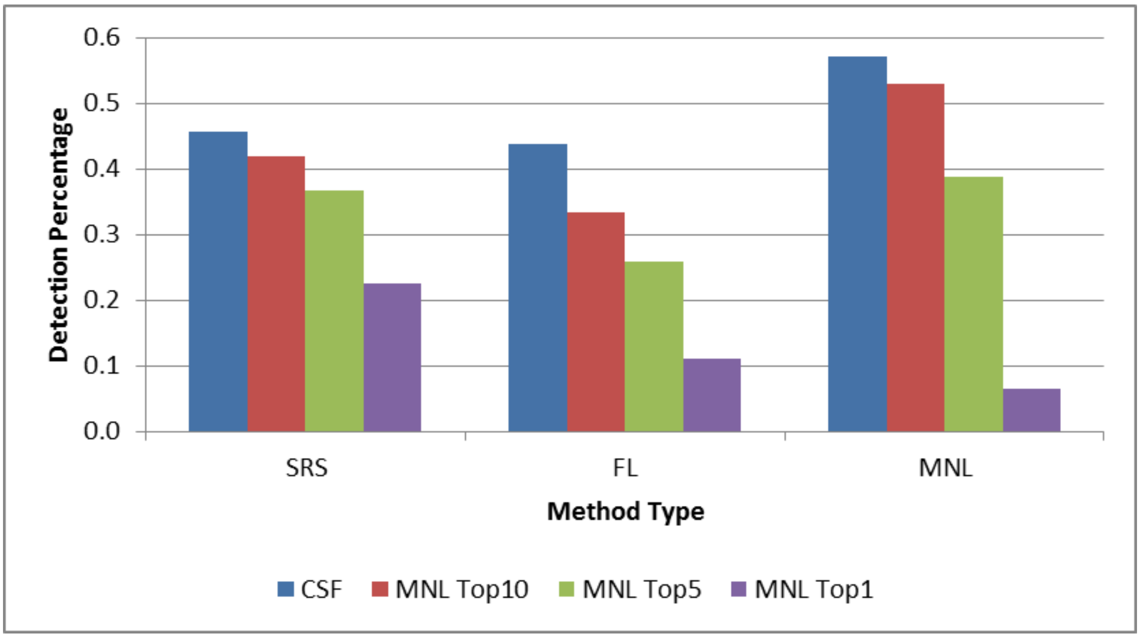

Figure 6.13 Prediction Capabilities of the Methods: Scenario 1 (All Cases)

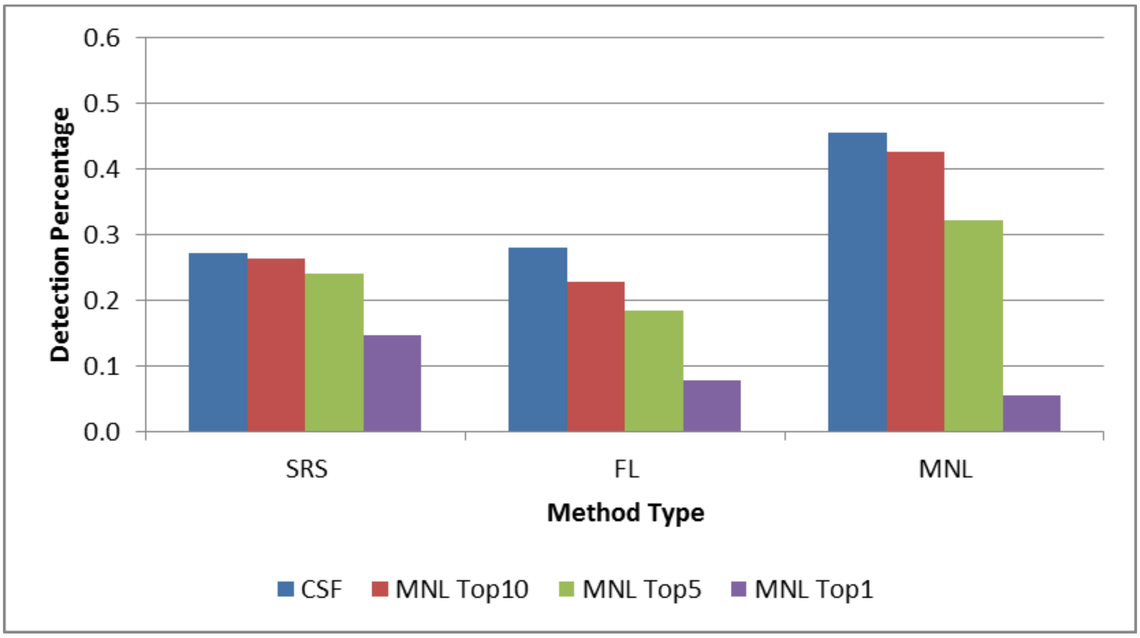

Figure 6.14 Prediction Capabilities of the Methods: Scenario 2 (Options > 2 * CS Size)

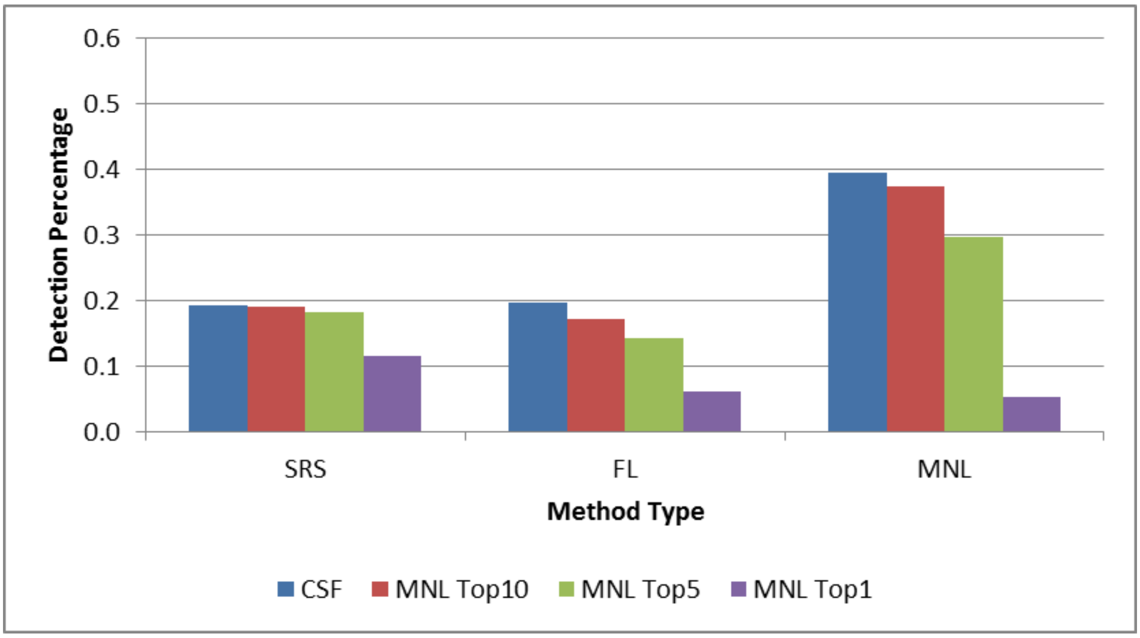

Figure 6.15 Prediction Capabilities of the Methods: Scenario 3 (Options > 3 * CS Size)

6.4 Conclusions

This study re-answers an important question, "Does the choice set composition affect model results?" which was previously answered by Bliemer and Bovy (2008) (in a simple network context), Prato and Bekhor (2007) (in a car route choice context) and Lemp and Kockelman (2012) (in a simulation study). However, this study answered this in a transit route choice context. Lemp and Kockelman (2012) showed that the *true choice sets*[5] could efficiently and accurately help in estimating model parameters and the current study confirms that the parameters estimated in such a way can also significantly improve the prediction capacity of the model. This study also showed that using the uniform selection probability (simple random sampling method) can be an alternative approach only when the prediction accuracy (correctly predicting the chosen alternative) is essential. However, when the ratio of choice set size and the available alternatives is large, the importance sampling methods (unequal selection probability) provide more reliable model performance. Consequently, the importance sampling methods can

[5] In reality, no one knows the true choice set. However, a true choice set refers to a choice set where all the competitive alternatives are included.

be useful when a high prediction accuracy (finding the exact choice that is selected) is not necessarily the goal. For example, route choice, destination choice, housing search and so forth.

This study shows that sampling protocols have a considerable impact on the formation of the choice set. The commonly used SRS sampling protocol cannot represent the *true choice set* as it shows inferior performance compared to the other protocols. The WOR sampling protocol appears to be better than the other two as it can accommodate more competitive alternatives in the choice set. Furthermore, the WOR protocol seems to be indirectly influencing the final choice probability of the models as models calibrated based on the WOR protocol can predict the final choice probability much better than the models calibrated using the WR or SRS protocols.

The other question that this study answer is whether a choice set formation method hampers the prediction capability. The study shows that in the first stage of the prediction where the choice set is formed, the sampling methods can make significant differences by accurately identifying the chosen alternative in the choice set. In this regard, prior information can be highly useful as these models (MNL models) show superior performance in the prediction stage. However, the other importance sampling models developed using Fuzzy Logic seems to be similar to SRS, although the FL-WOR models show a slightly better performance.

Modelling route choice, destination choice or housing location choice is crucial in transport demand modelling. This study suggests that using prior information in the sampling process can help to develop better demand models. If prior information is not available, the FL-WOR model structure can be used. However, a better way can be to develop a simple model using SRS and then use the estimated parameters of the SRS model as prior information to calculate the sampling probability. Next, use this sampling probability and the proposed importance sampling method to estimate the final model. As seen from this study, the proposed sampling protocol improves the quality of the choice set by including essential alternatives if prior information is used.

The limitations of this chapter are as follows:

- This chapter considers testing the performance of the proposed sampling method and the simple random sampling method with different modelling scenarios and sampling protocols. The data used in this analysis was obtained from a trip based shortest path algorithm discussed in subsection 2.2.2. The paths are generated only once, and the entire analysis is performed on that data. Thus, testing the proposed protocol on different path generation algorithms and protocols are out of scope for the study. Since the proposed sampling protocol provides promising results, future studies can investigate the applicability and performance of this protocol on the different path generation algorithms. Furthermore, future studies can also investigate the performance of this sampling protocol with other path generation protocols.
- In this study, it is understood that the choice set size plays a vital role in sampling, which is aligned with the literature (Daly et al., 2014; Nerella and Bhat, 2004). The results of this study suggest choosing unequal selection probabilities, especially when the options are very high compared to the choice set size. A future study can be undertaken to further understand the appropriate choice set size for sampling with unequal probabilities as this is suggested for uniform selection probabilities (Nerella and Bhat, 2004).

7 CONCLUSIONS AND FUTURE RESEARCH POTENTIAL

7.1 Summary of Contribution

This Thesis presents different modelling specifications to study transit stop choice and route choice behaviour with some interesting findings. The contribution of this study can be summarised as follows.

An important contribution of this study is including different travel strategies in the discrete choice modelling framework. The definition for the strategy used in this study capture the uniqueness of the strategy attributes (except the number of transfer variable as in the observed choice there was not enough variability) by a simple transformation process. The modelling exercise shows that this extra piece of derived information can explain users' transit stop and route choice behaviour better. Again, including these variables can significantly improve the overall goodness-of-fit of the models. Moreover, incorporating the strategy attributes in the modelling structure significantly improves the model's prediction capability, resulting in a better transit demand estimation.

The presented stop choice study can be useful for transport planners and policy-makers. The study revealed that SEQ passengers had a higher disutility for transfers and preferred more direct routes. They also showed a preference for stops where facilities like lighting are available. The preference for trains appeared to be very high as they might prefer to walk an extra 12 minutes to access a train station. The stop choice models also depict the influence of passengers' socio-demographic attributes. For instance, family size, accommodation type and birthplace affect their choice of the stop. Again, the day of the week and peak-off peak hour also influences the choice of access stop. All this information can help transport authorities to design and provide better facilities for transit passengers.

The route choice model also showed a very high preference for exclusive train routes, and passengers were willing to spend around 12 extra minutes of walking time to reach these routes. This might encourage policy-makers to introduce more train services. Again, passengers seem to have a higher disutility if the entire trip is travelled by bus. This suggests that the transit agency should improve the bus service. Moreover, passengers seem to prefer overlapping routes rather than unique routes.

The study compared two discrete choice modelling approaches, namely, RUM and RRM with different model specifications. To the authors' knowledge, this was the first attempt to analyse the RRM approach to investigate public transit stop choice and route choice behaviour. The study reveals the importance of using the RRM approach, particularly in a RUM-RRM hybrid specification as the SEQ passengers appeared to maximise utility for some of the attributes and minimise the regret from another group of attributes. Hybrid RUM-RRM models predict better when the choice set size is relatively small; in this particular study, it was 30. RUM based models seem to predict better when the choice set size is more than 30. Again, the study found that the prediction capability of the models decreases from a particular choice set size. Therefore, this study suggests investigating the optimum choice set size for modelling.

The most significant contribution of this study is to develop a probability-based importance sampling method which uses the sampling formulation of Guevara and Ben-Akiva (2013b) and Guevara et al. (2014). However, the alternatives are randomly selected from a probability profile of the universal set of alternatives. Therefore, the probability of including attractive alternatives in the choice set is higher than the simple random sampling protocol used in other studies (Guevara and Ben-Akiva, 2013b; Guevara et al., 2014; Nerella and Bhat, 2004). Consequently, this method solves the problem of unequal selection probability of sampling in a simple and straightforward way compare to the other solutions found in the literature (Frejinger et al., 2009; Zimmermann et al., 2018, Vacca et al., 2015).

Again, this method can be applied to other fields including housing search, and destination choice.

A comprehensive study was also undertaken to understand the effect of different sampling methods in the choice set formation, model estimation and prediction. In all these aspects, the proposed sampling protocol provides significantly better results. The proposed sampling method of this thesis has already been used in a destination choice context (Hassan et al., 2019) and promising results were found. Again, the route choice model can easily be implemented in transit assignment models because of its simple structure and ability to form reasonable choice sets with preferable alternatives with the help of the proposed sampling method.

7.2 Future Research Potentials

This study developed a transit access stop choice model which is a relatively new idea. The stop choice model provides a stepping stone to a futuristic behaviour-based transit choice model that uses smart card data and high-resolution origin-destination location data. In this modelling approach, an access stop choice model (proposed in this thesis) is developed from the trip origin to the departure stop. Again, from the departure stop to the destination stop, boarding strategy-based models (using smart card data) can be developed (Nassir et al., 2018). Eventually, the combination of these two models, along with an egress stop choice model, can effectively estimate and evaluate future transit demand from any given origin to destination.

One of the future directions of the transit choice modelling is to further investigate the choice structures jointly by adding the logsum of the route choice and strategy choice probability in a stop choice model. Alternatively, an inverse structure can be investigated, where the route choice model considers the logsum of stop choice and strategy choice probabilities. Again, further behavioural realism can be added to these models by introducing preference heterogeneity and taste heterogeneity.

REFERENCES

Abdel-Aty, M., Huang, Y. (2004) Exploratory spatial analysis of expressway ramps and its effect on route choice. *Journal of Transportation Engineering* 130, 104-112.

Abdelghany, K.F., Mahmassani, H.S. (1999) Multi-Objective Shortest Path Algorithm for Large Scale Intermodal Networks. *INFORMS*.

Abdelghany, K.F., Mahmassani, H.S. (2001) Dynamic trip assignment-simulation model for intermodal transportation networks. *Transportation Research Record* 1771, 52-60.

Ahuja, R., Magnanti, T., Orlin, J. (1993) *Network flows: theory, algorithms, and applications*. Prentice Hall, NJ.

Akamatsu, T. (1996) Cyclic flows, Markov process and stochastic traffic assignment. *Transportation Research Part B: Methodological* 30, 369-386.

Anderson, M.K., Nielsen, O.A., Prato, C.G. (2017) Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO Journal on Transportation and Logistics* 6, 221-245.

Antonini, G., Bierlaire, M., Weber, M. (2004) Simulation of pedestrian behaviour using a discrete choice model calibrated on actual motion data. *Proceedings of Swiss Transport Research Conference*.

Arslan, T., Khisty, J. (2006) A rational approach to handling fuzzy perceptions in route choice. *European Journal of Operational Research* 168, 571-583.

Auld, J., Mohammadian, A. (2011) Planning-constrained destination choice in activity-based model: agent-based dynamic activity planning and travel scheduling. *Transportation Research Record* 2254, 170-179.

Azevedo, J., Costa, M.E.O.S., Madeira, J.J.E.S., Martins, E.Q.V. (1993) An algorithm for the ranking of shortest paths. *European Journal of Operational Research* 69, 97-106.

Bajwa, S., Bekhor, S., Kuwahara, M., Chung, E. (2008) Discrete Choice Modeling of Combined Mode and Departure Time. *Transportmetrica* 4, 155-177.

Beimborn, E., Greenwald, M., Jin, X. (2003) Accessibility, Connectivity, and Captivity: Impacts on Transit Choice. *World Transit Research*.

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S. (2006) Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research* 144, 235-247.

Bekhor, S., Chorus, C., Toledo, T. (2012) Stochastic user equilibrium for route choice model based on random regret minimization. *Transportation Research Record* 2284, 100-108.

Bekhor, S., Prato, C.G. (2009) Methodological transferability in route choice modeling. *Transportation Research Part B: Methodological* 43, 422-437.

Bekhor, S., Toledo, T., Prashker, J.N. (2008) Effects of choice set size and route choice models on path-based traffic assignment. *Transportmetrica* 4, 117-133.

Bell, D.E. (1982) Regret in decision making under uncertainty. *Operations research* 30, 961-981.

Bell, M.G. (1995) Alternatives to Dial's logit assignment algorithm. *Transportation Research Part B: Methodological* 29, 287-295.

Ben-Akiva, M., Bergman, M., Daly, A.J., Ramaswamy, R. (1984) Modeling inter-urban route choice behaviour. *Proceedings of Proceedings of the 9th international symposium on transportation and traffic theory*, pp. 299-330.

Ben-Akiva, M., Bierlaire, M. (1999) Discrete choice methods and their applications to short term travel decisions. *Handbook of transportation science*. Springer, pp. 5-33.

Ben-Akiva, M., Bowman, J.L. (1998) Integration of an activity-based model system and a residential location model. *Urban Studies* 35, 1131-1153.

Ben-Akiva, M.E., Lerman, S.R. (1985) *Discrete choice analysis: theory and application to travel demand*. MIT press.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R. (1985) *Discrete choice analysis: theory and application to travel demand*. MIT press.

Benjamins, M., Lindveld, C., Van Nes, R. (2001) Multimodal travel choice modelling: a supernetwork approach. *Proceedings of 28th Colloquium Transport Planological Exhibition: Who Does What ?: About The Resistant Reality Of Market Operation And Decentralization: Bundling Of Contributions To The Colloquium* Amsterdam.

Bezdek, J. (1993) Fuzzy models - What are they, and why? *IEEE Transactions on Fuzzy Systems* 1, 1-6.

Bhat, C., Govindarajan, A., Pulugurta, V. (1998) Disaggregate attraction-end choice modeling: formulation and empirical analysis. *Transportation research record* 1645, 60-68.

Bhatta, B.P., Larsen, O.I. (2011) Errors in variables in multinomial choice modeling: A simulation study applied to a multinomial logit model of travel mode choice. *Transport policy* 18, 326-335.

Bielli, M., Boulmakoul, A., Mouncif, H. (2006) Object modeling and path computation for multimodal travel systems. *European Journal of Operational Research* 175, 1705-1730.

Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models. *Proceedings of Swiss Transport Research Conference*.

Bierlaire, M., Bolduc, D., McFadden, D. (2008) The estimation of generalized extreme value models from choice-based samples. *Transportation Research Part B: Methodological* 42, 381-394.

Bierlaire, M., Frejinger, E. (2005) Route choice models with subpath components. *Proceedings of Swiss Transportation Research Conference*.

Bliemer, M.C., Bovy, P.H. (2008) Impact of route choice set on route choice probabilities. *Transportation Research Record* 2076, 10-19.

Bliemer, M.C., Taale, H. (2006) Route generation and dynamic traffic assignment for large networks. *Proceedings of Proceedings of the First International Symposium on Dynamic Traffic Assignment, Leeds, UK*, pp. 90-99.

Boeri, M., Longo, A., Doherty, E., Hynes, S. (2012) Site choices in recreational demand: a matter of utility maximization or regret minimization? *Journal of Environmental Economics and Policy* 1, 32-47.

Boeri, M., Masiero, L. (2014) Regret minimisation and utility maximisation in a freight transport context. *Transportmetrica A: Transport Science* 10, 548-560.

Bovy, P.H., Bekhor, S., Prato, C.G. (2008) The factor of revised path size: an alternative derivation. *Transportation Research Record* 2076, 132-140.

Bovy, P.H., Fiorenzo-Catalano, S. (2007) Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica* 3, 173-189.

Bovy, P.H., Hoogendoorn-Lanser, S. (2005) Modelling route choice behaviour in multi-modal transport networks. *Transportation* 32, 341-368.

Carrasco, N. (2008) Deciding where to shop: disaggregate random utility destination choice modeling of grocery shopping in canton zurich. IVT, ETH Zürich.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A. (1996) A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. *Proceedings of Transportation and Traffic Theory. Proceedings of The 13th International Symposium On Transportation And Traffic Theory, Lyon, France, 24-26 July 1996*.

Cascetta, E., Papola, A. (2001) Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies* 9, 249-263.

Chakour, V., Eluru, N. (2013) Analyzing commuter train user behavior: a decision framework for access mode and station choice. *Transportation* 41, 211-228.

Chang, L.-Y., Mannering, F. (1999) Analysis of injury severity and vehicle occupancy in truck-and non-truck-involved accidents. *Accident Analysis & Prevention* 31, 579-592.

Chorus, C. (2016) Random Regret Minimization: A short History of Nearly Everything. *Random Regret Minimisation for travel behaviour research: State-of-the-art and the road ahead*, Delft.

Chorus, C.G. (2010) A new model of random regret minimization. *European Journal of Transport and Infrastructure Research* 10.

Chorus, C.G. (2012) *Random Regret-based discrete choice modeling: A tutorial*. Springer Science & Business Media.

Chorus, C.G. (2014) A generalized random regret minimization model. *Transportation research part B: Methodological* 68, 224-238.

Chorus, C.G., Arentze, T.A., Timmermans, H.J. (2008) A random regret-minimization model of travel choice. *Transportation Research Part B: Methodological* 42, 1-18.

Chorus, C.G., Bierlaire, M. (2012) An empirical comparison of travel choice models that capture preferences for compromise alternatives. *Transportation* 40, 549-562.

Chorus, C.G., De Jong, G.C. (2011) Modeling experienced accessibility for utility-maximizers and regret-minimizers. *Journal of Transport Geography* 19, 1155-1162.

Chorus, C.G., Rose, J.M., Hensher, D.A. (2013) Regret minimization or utility maximization: it depends on the attribute. *Environment and Planning B: Planning and Design* 40, 154-169.

Daly, A., Hess, S., Dekker, T. (2014) Practical solutions for sampling alternatives in large-scale models. *Transportation Research Record* 2429, 148-156.

De La Barra, T., Perez, B., Anez, J. (1993) Multidimensional path search and assignment. *Proceedings of PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom*.

Debrezion, G., Pels, E., Rietveld, P. (2009) Modelling the joint access mode and railway station choice. *Transportation Research Part E: Logistics and Transportation Review* 45, 270-283.

Dial, R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation research* 5, 83-111.

Ding, J., Gao, S., Jenelius, E., Rahmani, M., Huang, H., Ma, L., Pereira, F., Ben-Akiva, M. (2014) Routing policy choice set generation in stochastic time-dependent networks: Case studies for Stockholm, Sweden, and Singapore. *Transportation Research Record* 2466, 76-86.

DTMR (2010a) Survey Procedures and Documentation-Brisbane Survey 2009. ed Roads, D.o.T.a.M. Department of Transport and Main Roads.

DTMR (2010b) Survey Procedures and Documentation-Gold Coast Survey 2009. Department of Transport and Main Roads.

DTMR (2010c) Survey Procedures and Documentation-Sunshine Coast Survey 2009. ed Roads, D.o.T.a.M. Department of Transport and Main Roads.

Eluru, N., Chakour, V., El-Geneidy, A.M. (2012) Travel mode choice and transit route choice behavior in Montreal: insights from McGill University members commute patterns. *Public Transport* 4, 129-149.

Eppstein, D. (1998) Finding the k shortest paths. *SIAM Journal on computing* 28, 652-673.

Fatmi, M.R., Chowdhury, S., Habib, M.A. (2017) Life history-oriented residential location choice model: A stress-based two-tier panel modeling approach. *Transportation Research Part A: Policy and Practice* 104, 293-307.

Fiorenzo-Catalano, S., Van Nes, R., Bovy, P.H. (2004) Choice set generation for multi-modal travel analysis. *European journal of transport and infrastructure research EJTIR, 4 (2)*.

Fishburn, P.C. (1982) Nontransitive measurable utility. *Journal of Mathematical Psychology* 26, 31-67.

Florian, M. (2004) Finding shortest time-dependent paths in schedule-based transit networks: a label setting algorithm. *Schedule-based Dynamic Transit Modeling: Theory and Applications*. Springer, pp. 43-52.

Fonzone, A., Bell, M.G.H. (2010) Bounded rationality in hyperpath assignment: the locally rational traveller model., *89th Annual TRB Meeting*, Washington, D.C., USA.

Fonzone, A., Schmöcker, J.-D., Bell, M.G.H., Gentile, G., Kurauchi, F., Nökel, K., Wilson, N.H.M. (2010) Do "Hyper-Travellers" Exist? – Initial Results of an International Survey on Public Transport User Behaviour. *Proceedings of 12th WCTR*.

Fonzone, A., Schocker, J.-D., Kurauchi, F., Hassan, S.M. (2013) Strategy Choice in Transit Networks.

Fosgerau, M., Frejinger, E., Karlstrom, A. (2013) A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological* 56, 70-80.

Fosgerau, M., Hjorth, K., Lyk-Jensen, S.V. (2007a) The Danish value of time study: results for experiment 1. Note 5.

Fosgerau, M., Hjorth, K., Lyk-Jensen, S.V. (2007b) The Danish value of time study: results for experiment 1. Note 6.

Frejinger, E., Bierlaire, M. (2007) Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological* 41, 363-378.

Frejinger, E., Bierlaire, M., Ben-Akiva, M. (2009) Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological* 43, 984-994.

Friedrich, M. (1994) Rechnergestütztes Entwurfsverfahren für den ÖPNV im ländlichen Raum. Techn. Univ., Lehrstuhl für Verkehrs-und Stadtplanung.

Friedrich, M., Hofsaess, I., Wekeck, S. (2001) Timetable-based transit assignment using branch and bound techniques. *Transportation Research Record* 1752, 100-107.

Gao, S., Chabini, I. (2006) Optimal routing policy problems in stochastic time-dependent networks. *Transportation Research Part B: Methodological* 40, 93-122.

Garcia-Martinez, A., Cascajo, R., Jara-Diaz, S.R., Chowdhury, S., Monzon, A. (2018) Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice* 114, 52-66.

Garrow, L.A., Koppelman, F.S., Nelson, B.L. (2005) Efficient estimation of nested logit models using choice-based samples. *Proceedings of Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory*.

Guevara, C.A., Ben-Akiva, M.E. (2013a) Sampling of alternatives in logit mixture models. *Transportation Research Part B: Methodological* 58, 185-198.

Guevara, C.A., Ben-Akiva, M.E. (2013b) Sampling of alternatives in multivariate extreme value (MEV) models. *Transportation Research Part B: Methodological* 48, 31-52.

Guevara, C.A., Chorus, C.G., Ben-Akiva, M.E. (2014) Sampling of alternatives in random regret minimization models. *Transportation Science* 50, 306-321.

Habib, K.M.N. (2012) Modeling commuting mode choice jointly with work start time and work duration. *Transportation research part A: Policy and practice* 46, 33-47.

Habib, K.M.N., Sasic, A. (2014) A GEV model with scale heterogeneity for investigating the role of mobility tool ownership in peak period non-work travel mode choices. *Journal of choice modelling* 10, 46-59.

Hammadou, H., Thomas, I., Verhetsel, A., Witlox, F. (2008) How to incorporate the spatial dimension in destination choice models: The case of Antwerp. *Transportation Planning and Technology* 31, 153-181.

Hara, Y., Akamatsu, T. (2012) Stochastic user equilibrium traffic assignment with a network GEV based route choice model. *JSCE Proc. Infrastruct. Plann. Rev* 46.

Hassan, M.N., Najmi, A., Rashidi, T.H. (2019) A two-stage recreational destination choice study incorporating fuzzy logic in discrete choice modelling. *Transportation Research Part F: Traffic Psychology and Behaviour* 67, 123-141.

Hassan, M.N., Rashidi, T.H., Waller, S.T., Nassir, N., Hickman, M. (2016) Modeling Transit Users Stop Choice Behavior: Do Travelers Strategize? *Journal of Public Transportation* 19, 6.

Henn, V., Ottomanelli, M. (2006) Handling uncertainty in route choice models: From probabilistic to possibilistic approaches. *European Journal of Operational Research* 175, 1526-1538.

Hensher, D.A., Greene, W.H. (2001) The mixed logit model: The state of practice and warnings for the unwary. *Institute of Transport Studies*, 1-39.

Hensher, D.A., Greene, W.H., Ho, C.Q. (2016) Random regret minimization and random utility maximization in the presence of preference heterogeneity: an empirical contrast. *Journal of Transportation Engineering* 142, 04016009.

Hensher, D.A., Rose, J.M., Greene, W.H. (2005) *Applied choice analysis: a primer*. Cambridge University Press.

Hess, S., Bierlaire, M., Polak, J. (2005) Mixed GEV modelling of mode-choice. *Proceedings of Swiss Transportation Research Conference*.

Hess, S., Chorus, C.G. (2015) Utility maximisation and regret minimisation: a mixture of a generalisation. *Bounded rational choice behaviour: applications in transport*. Emerald Group Publishing Limited, pp. 31-47.

Hess, S., Daly, A., Rohr, C., Hyman, G. (2007) On the development of time period and mode choice models for use in large scale modelling forecasting systems. *Transportation Research Part A: Policy and Practice* 41, 802-826.

Hess, S., Ryley, T., Davison, L., Adler, T. (2013) Improving the quality of demand forecasts through cross nested logit: a stated choice case study of airport, airline and access mode choice. *Transportmetrica A: Transport Science* 9, 358-384.

Hess, S., Stathopoulos, A. (2013) A mixed random utility — Random regret model linking the choice of decision rule to latent character traits. *Journal of Choice Modelling* 9, 27-38.

Hoogendoorn-Lanser, S. (2005) *Modelling travel behaviour in multi-modal networks*.

Horn, M.E. (2003) An extended model and procedural framework for planning multi-modal passenger journeys. *Transportation Research Part B: Methodological* 37, 641-660.

Jánošíková, Ľ., Slavík, J., Koháni, M. (2014) Estimation of a route choice model for urban public transport using smart card data. *Transportation planning and technology* 37, 638-648.

Kedia, A.S., Saw, K.B., Katti, B.K. (2015) Fuzzy logic approach in mode choice modelling for education trips: a case study of Indian metropolitan city. *Transport* 30, 286-293.

Khani, A. (2013) Models and solution algorithms for transit and intermodal passenger assignment (development of FAST-TrIPs model). The University of Arizona.

Khani, A., Hickman, M., Noh, H. (2014) Trip-Based Path Algorithms Using the Transit Network Hierarchy. *Networks and Spatial Economics* 15, 635-653.

Khani, A., Lee, S., Hickman, M., Noh, H., Nassir, N. (2012) Intermodal path algorithm for time-dependent auto network and scheduled transit service. *Transportation Research Record* 2284, 40-46.

Kim, H., PARK, D., KIM, C., KIM, Y. (2011) A tour-based approach to destination choice modeling incorporating agglomeration and competition effects. *Proceedings of Proceedings of the Eastern Asia Society for Transportation Studies*, pp. 101-112.

Kulkarni, A.D. (2001) *Computer vision and fuzzy-neural systems*. Prentice Hall PTR, Upper Saddle River, NJ.

Kurauchi, F., Schmöcker, J.-D., Fonzone, A., Hemdan, S.M.H., Shimamoto, H., Bell, M.G.H. (2012) Estimating Weights of Times and Transfers for Hyperpath

Travelers. *Transportation Research Record: Journal of the Transportation Research Board* 2284, 89-99.

Kurauchi, F., Schmöcker, J.-D., Shimamoto, H., Hassan, S.M. (2014) Variability of commuters' bus line choice: an analysis of oyster card data. *Public Transport* 6, 21-34.

Lam, W.H.-K., Gao, Z., Chan, K., Yang, H. (1999) A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B: Methodological* 33, 351-368.

Lee, B.H., Waddell, P. (2010) Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation* 37, 587-601.

Lemp, J.D., Kockelman, K.M. (2012) Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice* 46, 602-613.

Leong, W., Hensher, D.A. (2015) Contrasts of Relative Advantage Maximisation with Random Utility Maximisation and Regret Minimisation. *Journal of Transport Economics and Policy (JTEP)* 49, 167-186.

Li, M.-T., Chow, L.-F., Zhao, F., Li, S.-C. (2005) Geographically stratified importance sampling for the calibration of aggregated destination choice models for trip distribution. *Transportation research record* 1935, 85-92.

Li, M., Huang, H.-J. (2017) A regret theory-based route choice model. *Transportmetrica A: Transport Science* 13, 250-272.

Loomes, G., Sugden, R. (1982) Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal* 92, 805-824.

Lotan, T. (1997) Effects of familiarity on route choice behavior in the presence of information. *Transportation Research Part C: Emerging Technologies* 5, 225-243.

Lotan, T., Koutsopoulos, H.N. (1993) Models for route choice behavior in the presence of information using concepts from fuzzy set theory and approximate reasoning. *Transportation* 20, 129-155.

Lozano, A., Storchi, G. (2001) Shortest viable path algorithm in multimodal networks. *Transportation Research Part A: Policy and Practice* 35, 225-241.

Mahmoud, M., Habib, K., Shalaby, A. (2014) Park-and-Ride Access Station Choice Model for Cross-Regional Commuting: Case Study of Greater Toronto and Hamilton Area, Canada. *Transportation Research Record: Journal of the Transportation Research Board* 2419, 92-100.

Malekzadeh, A. (2015) Measurement of transit network accessibility based on access stop choice behaviour. Queensland University of Technology.

Manski, C.F. (1977) The structure of random utility models. *Theory and decision* 8, 229-254.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior.

(1977) *Modelling the Choice of Residential Location*.

McFadden, D., Train, K., Tye, W. (1978) An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. *Transportation Research Record* 637, 39-46.

Meng, M., Rau, A., Mahardhika, H. (2018) Public transport travel time perception: Effects of socioeconomic characteristics, trip characteristics and facility usage. *Transportation Research Part A: Policy and Practice* 114, 24-37.

Nassir, N., Hickman, M., Ma, Z.-L. (2015a) Activity detection and transfer identification for public transit fare card data. *Transportation* 42, 683-705.

Nassir, N., Hickman, M., Ma, Z.-L. (2018) A strategy-based recursive path choice model for public transit smart card data. *Transportation Research Part B: Methodological*.

Nassir, N., Hickman, M., Malekzadeh, A., Irannezhad, E. (2015b) Modeling transit access stop choices. *Proceedings of Transportation Research Board 94th Annual Meeting*.

Nassir, N., Hickman, M., Malekzadeh, A., Irannezhad, E. (2016) A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice* 88, 26-39.

Nerella, S., Bhat, C. (2004) Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*, 11-19.

Nguyen, S., Pallottino, S. (1988) Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research* 37, 176-186.

Nguyen, S., Pallottino, S., Gendreau, M. (1998) Implicit enumeration of hyperpaths in a logit model for transit networks. *Transportation Science* 32, 54-64.

Nielsen, O.A. (2000) A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological* 34, 377-402.

Nielsen, O.A., Frederiksen, R.D. (2006) Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research* 144, 263-285.

Nielsen, O.A., Hansen, C.O., Daly, A. (2000) A Large-scale model system for the Copenhagen-Ringsted railway project. *Proceedings of International Conference On Travel Behaviour Research, 9th, 2000, Gold Coast, Queensland, Australia, Vol 12*.

Noh, H., Hickman, M., Khani, A. (2012) Hyperpaths in network based on transit schedules. *Transportation Research Record* 2284, 29-39.

Nuzzolo, A., Russo, F., Crisalli, U. (2001) A doubly dynamic schedule-based assignment model for transit networks. *Transportation Science* 35, 268-285.

Papola, A., Marzano, V. (2013) A network generalized extreme value model for route choice allowing implicit route enumeration. *Computer-Aided Civil and Infrastructure Engineering* 28, 560-580.

Park, H., Park, D., Kim, C., Kim, H., Park, M. (2012) A comparative study on sampling strategies for truck destination choice model: Case of Seoul Metropolitan Area. *Canadian Journal of Civil Engineering* 40, 19-26.

Pozsgay, M.A., Bhat, C.R. (2001) Destination choice modeling for home-based recreational trips: analysis and implications for land use, transportation, and air quality planning. *Transportation research record* 1777, 47-54.

Prato, C.G. (2009) Route choice modeling: past, present and future research directions. *Journal of choice modelling* 2, 65-100.

Prato, C.G. (2014) Expanding the applicability of random regret minimization for route choice analysis. *Transportation* 41, 351-375.

Prato, C.G., Bekhor, S. (2006) Applying branch-and-bound technique to route choice set generation. *Transportation Research Record* 1985, 19-28.

Prato, C.G., Bekhor, S. (2007) Modeling route choice behavior: How relevant is the composition of choice set? *Transportation Research Record* 2003, 64-73.

Prato, C.G., Bekhor, S., Pronello, C. (2012) Latent variables and route choice behavior. *Transportation* 39, 299-319.

Quattrone, A., Vitetta, A. (2011) Random and fuzzy utility models for road route choice. *Transportation Research Part E: Logistics and Transportation Review* 47, 1126-1139.

Ramming, M. (2002) Network Knowledge and Route Choice [PhD thesis]. *Cambridge, USA: Massachusetts Institute of Technology*.

Rasmussen, T.K., Nielsen, O.A., Watling, D.P., Prato, C.G. (2017) The Restricted Stochastic User Equilibrium with Threshold model: Large-scale application and parameter testing. *European Journal of Transport and Infrastructure Research* 17, 1-24.

Raveau, S., Muñoz, J.C., De Grange, L. (2011) A topological route choice model for metro. *Transportation Research Part A: Policy and Practice* 45, 138-147.

Ridwan, M. (2004) Fuzzy preference based traffic assignment problem. *Transportation Research Part C: Emerging Technologies* 12, 209-233.

Ross, T.J. (2004) *Fuzzy logic with engineering applications*, 2nd ed ed. John Wiley, Hoboken, NJ.

Schakenbos, R., La Paix, L., Nijenstein, S., Geurs, K.T. (2016) Valuation of a transfer in a multimodal public transport trip. *Transport Policy* 46, 72-81.

Schmöcker, J.-D., Shimamoto, H., Kurauchi, F. (2013) Generation and calibration of transit hyperpaths. *Procedia-Social and Behavioral Sciences* 80, 211-230.

Schüssler, N., Axhausen, K.W. (2009) Accounting for route overlap in urban and suburban route choice decisions derived from GPS observations. *Arbeitsberichte Verkehrs-und Raumplanung* 590.

Scott, K., Pabón-Jiménez, G., Bernstein, D. (1997) Finding alternatives to the best path.

Sermons, M.W., Koppelman, F.S. (2001) Representing the differences between female and male commute behavior in residential location choice models. *Journal of transport geography* 9, 101-110.

Shabanpour, R., Golshani, N., Auld, J., Mohammadian, A. (2017) Dynamics of activity time-of-day choice. *Transportation Research Record* 2665, 51-59.

Shakeel, K., Rashidi, T.H., Waller, T.S. (2016) Choice set formation behavior: joint mode and route choice selection model. *Transportation Research Record* 2563, 96-104.

Sheffi, Y., Powell, W.B. (1982) An algorithm for the equilibrium assignment problem with random link times. *Networks* 12, 191-207.

Simma, A., Schlich, R., Axhausen, K. (2001) Destination choice modelling of leisure trips: The case of Switzerland. *Arbeitsberichte Verkehrs-und Raumplanung* 99.

Spiess, H., Florian, M. (1989) Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B: Methodological* 23, 83-102.

Tan, R. (2016) Modeling Route Choice Behaviour in Public Transport Network. Unpublished PhD thesis, National University of Singapore (Singapore).

Tan, R., Raveau, S., Lee, D. H., & Ben-Akiva, M. (2016) *A Public Transport Route Choice Model with Multimodal Access and Egress: Application to the Large-Scale Network of Singapore*.

Ulfarsson, G.F., Mannering, F.L. (2004) Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accident Analysis & Prevention* 36, 135-147.

Vacca, A., Prato, C.G., Meloni, I. (2015) Estimating route choice models from stochastically generated choice sets on large-scale networks: Correcting for unequal sampling probability. *Transportation Research Record* 2493, 11-18.

Van Cranenburgh, S., Guevara, C.A., Chorus, C.G. (2015a) New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice* 74, 91-109.

Van Cranenburgh, S., Prato, C.G. (2016) On the robustness of random regret minimization modelling outcomes towards omitted attributes. *Journal of choice modelling* 18, 51-70.

Van Cranenburgh, S., Prato, C.G., Chorus, C. (2015b) Accounting for variation in choice set size in Random Regret Minimization models. working paper.

Van der Waard, J. (1988) The relative importance of public transport trip time attributes in route choice. *Proceedings of PTRC Summer Annual Meeting, 16th, 1988, Bath, United Kingdom*.

Van der Zijpp, N., Catalano, S.F. (2005) Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B: Methodological* 39, 545-563.

Vij, A., Carrel, A., Walker, J.L. (2013) Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice* 54, 164-178.

Vrtic, M., Axhausen, K. (2003) The impact of tilting trains in Switzerland: A route choice model of regional-and long distance public transport trips. *Proceedings of 83rd Annual Meeting of the Transportation Research Board, Washington, DC*.

Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.

Wahba, M., Shalaby, A. (2009) MILATRAS: A new modeling framework for the transit assignment problem. *Schedule-Based Modeling of Transportation Networks*. Springer, pp. 1-24.

Washington, S.P., Karlaftis, M.G., Mannering, F. (2010) *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.

Wen, C.-H., Wang, W.-C., Fu, C. (2012) Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review* 48, 545-554.

Yang, L., Lam, W. (2006) Probit-type reliability-based transit network assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 154-163.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control* 8, 338-353.

Zhou, M., Lu, J. (2011) Research on Prediction of Traffic Mode Choice of Urban Residents. *ICCTP 2011: Towards Sustainable Transportation Systems*, pp. 449-460.

Ziliaskopoulos, A., Wardell, W. (2000) An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research* 125, 486-502.

Zimmermann, M., Axhausen, K., Frejinger, E. (2018) Multi-modal route choice modeling in a dynamic schedule-based transit network. *Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation*.

Zolfaghari, A., Sivakumar, A., Polak, J.W. (2012) Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. *Transportation Planning and Technology* 35, 87-106.

APPENDICES

APPENDIX A

DESCRIPTION OF THE VARIABLES

Table A1 Descriptions of the Impedance Variables for the Stop Choice Model

Variable	Mean	SD	Description
FastestTT	46.95	19.85	Travel time (min) of fastest path to destination from the stop (excluding AccessWalk)
MinTransfer	0.83	0.84	Minimum number of transfer among paths from the stop to destination
MinWalk	19.05	8.88	Minimum walk time (min) among paths from the stop to destination (excluding AccessWalk)
MinFare	1.15	1.10	Minimum fare among paths from the stop to destination
MinWait	10.56	14.34	Minimum wait time (min) among paths from the stop to destination
NumRoutes	1.74	1.96	Number of available paths from the stop to destination
TotalFreq	4.55	7.46	Summation of frequency for all the paths from the stop to destination
AveTT	48.50	20.33	Average travel time of all paths from the stop to destination (excluding AccessWalk)
AveTransfer	0.95	0.83	Average number of required transfers for all paths from the stop to destination
AveWalk	19.70	8.83	Average walking time (min) for all paths from the stop to destination (excluding AccessWalk)
AveFare	1.17	1.10	Average fare for all paths from the stop to destination
AveWait	12.37	15.17	Average waiting time (min) for all paths from the stop to destination

Table A2 Descriptions of the Impedance Variables for the Route Choice Model

Variable	Mean	SD	Description
TravelTime	60.74	40.78	Travel time (min) of path to destination from the stop (excluding AccessWalk)
NumberOfTransfer	1.2	1.23	Number of transfer within the path
WalkTime	21.54	10.97	Walk time within paths from the stop to destination (excluding AccessWalk)
WaitTime	21.25	29.08	Wait time (min) within paths from the stop to destination
TravelTimeTrain	29.36	23.19	In-vehicle travel time (min) in train
TravelTimeBus	34.25	32.07	In-vehicle travel time (min) in bus
TravelTimeFerry	17.2	14.13	In-vehicle travel time (min) in bus

Table A3 Descriptions of the Access Stop Variables for the Stop Choice Model

Variable	Mean	SD	Description
Train	0.05	0.22	Binary variable indicating the trip access stop is a train station
AccessWalk	11.33	7.11	Walk time from origin location to stop (min)
Shelter	0.41	0.49	Binary variable indicating a sheltered stop
StopLight	0.34	0.47	Binary variable indicating an illuminated stop
StreetLight	0.31	0.46	Binary variable indicating an illuminated street
BoardingSlab	0.88	0.32	Binary variable indicating existence of a boarding slab
FootPath	0.87	0.34	Binary variable indicating existence of foot path
Map	1.65	2.74	Total number of printed map/schedule at the stop

Table A4 Descriptions of the Access Stop Variables for the Route Choice Model

Variable	Mean	SD	Description
AccessWalk	11.85	7.2	Walk time from an origin location to stop (min)
Shelter	0.35	0.48	Binary variable indicating a sheltered stop
StopLight	0.36	0.48	Binary variable indicating an illuminated stop
StreetLight	0.26	0.44	Binary variable indicating an illuminated street
BoardingSlab	0.87	0.33	Binary variable indicating existence of a boarding slab
FootPath	0.86	0.34	Binary variable indicating existence of foot path
Map A4	0.86	1.48	Number of printed A4 map/schedule at the stop
Map A3	0.16	0.66	Number of printed A3 map/schedule at the stop

Table A5 Descriptions of the Mode Variables used in Route Choice Model

Variable	Mean	SD	Description
MainTrain	0.12	0.00	Binary variable indicating the maximum in vehicle travel time in the route is in train
MainBus	0.88	0.00	Binary variable indicating the maximum in vehicle travel time in the route is in bus
MainFerry	0.01	0.00	Binary variable indicating the maximum in vehicle travel time in the route is in ferry
OnlyTrain	0.04	0.00	Binary variable indicating the route consist only train
OnlyBus	0.81	0.00	Binary variable indicating the route consist only bus
MixedMode	0.15	0.00	Binary variable indicating the route consist different modes

Table A6 Descriptions of the Strategy Variables for Access Stop Choice

Variable	Mean	SD	Description
MTT Strategy	0.23	0.42	Binary variable indicating the option offers minimum travel time
MTR Strategy	0.45	0.50	Binary variable indicating the option offer minimum number of transfer
MAT Strategy	0.17	0.37	Binary variable indicating the option offers min walking access time

Table A7 Descriptions of the Trip Variables

Variable	Mean	SD	Description
AMPeakDep	0.28	0.45	Binary variable indicating the trip starts in AM Peak Hour
PMPeakDep	0.17	0.38	Binary variable indicating the trip starts in PM Peak Hour
PeakHourDep	0.45	0.50	Binary variable indicating the trip starts in a Peak Hour
AMPeakArv	0.25	0.44	Binary variable indicating the trip ends in AM Peak Hour
PMPeakArv	0.22	0.42	Binary variable indicating the trip ends in PM Peak Hour
PeakHourArv	0.48	0.50	Binary variable indicating the trip ends in a Peak Hour
Weekday	0.90	0.30	Binary variable indicating the trip was in a weekday
PurposeWork	0.67	0.47	Binary variable indicating the trip was made for work purpose

Table A8 Descriptions of the User Variables

Variable	Mean	SD	Description
Age	35.55	19.20	Age of the user
Male	0.44	0.50	Binary variable indicating the user is male
HHSize	3.02	1.35	Total number of members in the HH (Household)
CoupleKids	0.36	0.48	Binary variable indicating the user H/H type is couple with kids
OneParent	0.08	0.27	Binary variable indicating the user H/H type is one parent with kids
Sole	0.13	0.33	Binary variable indicating the user H/H type is sole
Couple	0.20	0.40	Binary variable indicating the user H/H type is couple
OtherHHType	0.23	0.42	Binary variable indicating the user H/H type is other
House	0.81	0.40	Binary variable indicating the user lives in a house
Flat	0.15	0.35	Binary variable indicating the user lives in a flat
Townhouse	0.05	0.21	Binary variable indicating the user lives in a townhouse
Bedrooms	3.13	0.98	Number of bedrooms in the accommodation
OwnedProp	0.57	0.50	Binary variable indicating the user lives in an owned property
LivedInTheProp	99.38	127.87	Total number of months lived on the accommodation
HHIncome	1850.28	1340.09	Weekly income of the H/H
HighPerIncome	0.11	0.31	Binary variable indicating the user falls in the high income group
MedPerIncome	0.38	0.49	Binary variable indicating the user falls in the medium income group

LowPerIncome	0.51	0.50	Binary variable indicating the user falls in the low income group
FullTimeWork	0.37	0.48	Binary variable indicating the user is a full time worker
AnyWork	0.58	0.49	Binary variable indicating the user works
Student	0.00	0.06	Binary variable indicating the user is a student
AustralianBorn	0.72	0.45	Binary variable indicating the user born in Australia
CarLicence	0.51	0.50	Binary variable indicating the user has a car license
BikeLicence	0.02	0.15	Binary variable indicating the user has a motorbike license
NoLicence	0.39	0.49	Binary variable indicating the user has no license
TotalVehs	1.37	0.99	Total number of vehicles in the HH
PersonalVeh	0.75	0.43	Binary variable indicating the user has a personal vehicle
Bicycles	1.39	1.52	Total number of bicycles in the HH

APPENDIX B

CORRELATION ANALYSIS OF THE VARIABLES

Table B1: Correlation Analysis of the Stop Choice Model Variables

	Fastest Travel Time	Average Travel Time	Minimun Number of Transfers	Average Number of Transfer	Minimum Total Walk Time	Average Total Walk Time	Minimum Wait Time	Average Wait Time	Minimum Fare	Average Fare	Access Walk Time	Minimum Other Walk Time	Number of Routes
Fastest Travel Time	1.00												
Average Travel Time	0.97	1.00											
Minimun Number of Transfers	0.29	0.25	1.00										
Average Number of Transfer	0.27	0.26	0.95	1.00									
MinimumTotal Walk Time	0.39	0.36	-0.05	-0.09	1.00								
Average Walk Time	0.39	0.38	-0.08	-0.09	0.98	1.00							
Minimum Wait Time	0.44	0.43	0.29	0.27	0.06	0.04	1.00						
Average Wait Time	0.41	0.45	0.24	0.26	0.02	0.04	0.95	1.00					
Minimum Fare	0.54	0.53	0.35	0.36	-0.02	-0.03	0.16	0.15	1.00				
Average Fare	0.54	0.54	0.35	0.37	-0.03	-0.03	0.16	0.16	0.99	1.00			
Access Walk Time	0.18	0.17	-0.19	-0.20	0.74	0.74	0.02	0.01	-0.14	-0.15	1.00		
Minimum Other Walk Time	0.36	0.33	0.14	0.10	0.60	0.58	0.06	0.02	0.13	0.13	-0.09	1.00	
Number of Routes	-0.13	-0.04	-0.19	-0.02	-0.11	-0.03	-0.14	-0.01	-0.05	-0.02	-0.02	-0.14	1.00
Street Light	-0.01	-0.01	-0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.03	-0.02	-0.01
Map A4	-0.04	-0.02	-0.05	-0.02	0.08	0.09	0.00	0.03	-0.03	-0.03	0.17	-0.08	0.04
Map A3	0.05	0.08	-0.10	-0.07	0.05	0.06	0.03	0.07	0.01	0.02	0.09	-0.02	0.07
Shelter	0.01	0.01	-0.03	-0.02	0.03	0.03	0.03	0.04	-0.01	-0.01	0.08	-0.05	-0.03
Stop Light	-0.07	-0.06	-0.03	0.00	0.02	0.03	-0.11	-0.09	0.03	0.03	0.04	-0.02	0.16
Footpath Access	-0.10	-0.10	-0.01	0.00	-0.04	-0.04	-0.10	-0.10	0.02	0.02	-0.03	-0.02	0.05
Boarding Slab	-0.10	-0.10	0.00	0.01	-0.04	-0.04	-0.09	-0.09	0.02	0.02	-0.03	-0.02	0.04
Bus Stop	0.01	-0.01	0.04	0.01	-0.05	-0.05	0.04	0.01	-0.03	-0.04	-0.01	-0.06	-0.05
Train Station	-0.01	0.01	-0.04	-0.01	0.03	0.03	-0.03	0.00	0.04	0.05	0.04	-0.02	0.07
Ferry Terminal	-0.01	-0.01	-0.01	-0.01	0.04	0.04	-0.01	-0.01	-0.01	-0.01	-0.10	0.18	0.00
Total Frequency	-0.13	-0.08	-0.11	0.02	-0.07	-0.02	-0.17	-0.09	-0.01	0.00	-0.01	-0.10	0.86
MTT Strategy	-0.03	0.02	-0.06	-0.02	-0.21	-0.19	0.01	0.06	0.08	0.09	-0.20	-0.07	0.13
MTR Strategy	-0.18	-0.14	-0.59	-0.73	0.11	0.13	-0.14	-0.09	-0.27	-0.27	0.25	-0.13	0.18
MAT Strategy	-0.05	-0.03	0.11	0.12	-0.39	-0.39	0.07	0.09	0.05	0.05	-0.49	0.00	0.02

Table B1: Correlation Analysis of the Stop Choice Model Variables (cont.)

	Street Light	Map A4	Map A3	Shelter	Stop Light	Footpath Access	Boarding Slab	Bus Stop	Train Station	Ferry Terminal	Total Frequency	MTT Strategy	MTR Strategy	MAT Strategy
Fastest Travel Time														
Average Travel Time														
Minimun Number of Transfers														
Average Number of Transfer														
MinimumTotal Walk Time														
Average Walk Time														
Minimum Wait Time														
Average Wait Time														
Minimum Fare														
Average Fare														
Access Walk Time														
Minimum Other Walk Time														
Number of Routes														
Street Light	1.00													
Map A4	0.35	1.00												
Map A3	0.33	0.53	1.00											
Shelter	0.24	0.37	0.31	1.00										
Stop Light	-0.24	0.23	0.19	0.24	1.00									
Footpath Access	0.20	0.20	0.10	0.24	0.24	1.00								
Boarding Slab	0.21	0.20	0.10	0.30	0.26	0.80	1.00							
Bus Stop	-0.24	-0.55	-0.64	-0.19	-0.28	-0.06	-0.05	1.00						
Train Station	0.28	0.64	0.73	0.24	0.28	0.08	0.08	-0.83	1.00					
Ferry Terminal	0.01	0.01	-0.01	0.00	0.00	0.01	0.00	-0.31	-0.02	1.00				
Total Frequency	-0.09	-0.05	-0.05	-0.10	0.16	0.01	-0.01	0.01	-0.02	0.01	1.00			
MTT Strategy	0.01	0.03	0.12	0.02	-0.01	-0.04	-0.05	-0.03	0.06	-0.04	0.05	1.00		
MTR Strategy	0.03	0.12	0.12	0.07	0.02	-0.01	-0.01	-0.05	0.06	0.01	0.09	0.08	1.00	
MAT Strategy	-0.02	-0.03	0.00	0.00	-0.03	-0.01	0.01	0.01	0.00	-0.04	-0.01	0.24	-0.10	1.00

Table B2: Correlation Analysis of the Route Choice Model Variables

	Travel Time	Walk Time	Wait Time	Access Time	Number Of Transfer	Only Train	Only Bus	Street Lighting	Map A4	Map A3	Shelter	Stop Lighting	Footpath	Boarding Slab	Number Of Routes	MTT	MAT	MTR
Travel Time	1.00																	
Walk Time	0.56	1.00																
Wait Time	0.60	0.22	1.00															
Access Time	0.26	0.72	0.08	1.00														
Number Of Transfer	0.27	0.08	0.47	-0.07	1.00													
Only Train	0.01	0.00	0.00	0.01	-0.10	1.00												
Only Bus	-0.21	-0.06	-0.10	0.06	-0.21	-0.38	1.00											
Street Lighting	0.05	0.02	0.03	0.02	0.01	-0.06	0.04	1.00										
Map A4	-0.01	0.06	0.04	0.14	0.04	-0.06	0.08	0.26	1.00									
Map A3	-0.01	-0.01	0.03	0.03	-0.08	-0.04	0.11	0.11	-0.13	1.00								
Shelter	-0.01	-0.03	0.02	-0.01	-0.01	0.01	0.02	0.21	0.34	0.15	1.00							
Stop Lighting	-0.03	0.02	-0.06	0.05	-0.01	0.01	0.01	-0.41	-0.09	-0.15	0.11	1.00						
Footpath	0.04	0.02	0.00	-0.02	0.01	-0.12	0.10	0.22	0.22	0.07	0.28	0.33	1.00					
Boarding Slab	0.04	0.02	-0.01	-0.01	0.01	-0.12	0.10	0.23	0.22	0.07	0.31	0.34	0.90	1.00				
Number Of Routes	-0.01	0.08	-0.06	0.10	-0.02	0.03	0.00	-0.19	-0.20	-0.10	-0.24	0.44	0.10	0.09	1.00			
MTT	-0.09	-0.19	-0.10	-0.20	-0.03	0.03	-0.04	-0.01	-0.03	0.09	0.02	-0.05	-0.03	-0.04	-0.09	1.00		
MAT	-0.08	-0.28	0.00	-0.36	0.03	0.03	-0.05	0.00	-0.02	0.03	0.00	-0.03	-0.04	-0.03	-0.05	0.19	1.00	
MTR	-0.23	0.01	-0.20	0.16	-0.65	0.10	0.18	0.01	0.05	0.11	0.04	-0.05	-0.04	-0.04	-0.06	0.04	-0.02	1.00

Table B3: Correlation Analysis of the Socio-Demographic and Trip Variables

	Age	Gender	Male	AMPeak Departure	PMPeak Departure	PeakHour Departure	Purpose Work	AMPeak Arrival	PMPeak Arrival	PeakHour Arrival	Household Size	Couple with Kids	One Parent	Sole	Couple
Age	1.00														
Gender	0.09	1.00													
Male	-0.09	-1.00	1.00												
AMPeak Departure	0.00	0.00	0.00	1.00											
PMPeak Departure	0.02	-0.04	0.04	-0.29	1.00										
PeakHour Departure	0.02	-0.03	0.03	0.69	0.50	1.00									
Purpose Work	-0.16	-0.08	0.08	0.25	0.05	0.26	1.00								
AMPeak Arrival	-0.04	-0.02	0.02	0.80	-0.27	0.53	0.30	1.00							
PMPeak Arrival	0.01	0.00	0.00	-0.33	0.70	0.23	0.08	-0.31	1.00						
PeakHour Arrival	-0.03	-0.01	0.01	0.42	0.35	0.65	0.33	0.61	0.56	1.00					
Household Size	-0.50	-0.03	0.03	-0.05	-0.05	-0.08	0.08	0.00	-0.02	-0.02	1.00				
Couple with Kids	-0.32	-0.06	0.06	-0.03	-0.04	-0.05	0.07	0.00	-0.02	-0.02	0.57	1.00			
One Parent	-0.20	0.06	-0.06	0.03	-0.10	-0.05	-0.05	0.00	-0.09	-0.08	0.03	-0.22	1.00		
Sole	0.43	0.07	-0.07	-0.02	0.01	-0.01	-0.09	-0.04	-0.01	-0.05	-0.57	-0.28	-0.11	1.00	
Couple	0.20	0.00	0.00	0.05	0.07	0.10	0.00	0.06	0.04	0.09	-0.32	-0.37	-0.15	-0.19	1.00
Other HH Type	-0.04	-0.02	0.02	-0.02	0.03	0.01	0.03	-0.03	0.05	0.02	0.09	-0.41	-0.16	-0.21	-0.28
House	-0.13	-0.02	0.02	-0.01	0.00	-0.01	-0.01	0.02	0.01	0.02	0.39	0.32	-0.05	-0.34	-0.09
Flat	0.12	0.00	0.00	0.00	-0.02	-0.01	0.03	-0.02	-0.01	-0.02	-0.35	-0.30	0.04	0.30	0.11
Townhouse	0.04	0.04	-0.04	0.02	0.04	0.04	-0.03	0.00	0.01	0.00	-0.15	-0.11	0.04	0.13	-0.01
Number of Bedrooms	-0.19	-0.01	0.01	-0.01	-0.02	-0.02	0.00	0.02	0.01	0.02	0.58	0.32	-0.04	-0.34	-0.18
Owned Property	0.16	-0.03	0.03	0.04	0.04	0.06	0.05	0.06	0.10	0.14	0.01	0.21	-0.17	-0.07	0.13
Lived In The Property	0.42	-0.01	0.01	-0.02	-0.05	-0.06	-0.12	-0.01	-0.02	-0.02	-0.15	-0.08	-0.07	0.19	-0.02
Number of Bicycle	-0.31	-0.07	0.07	0.02	-0.05	-0.01	0.08	0.07	-0.05	0.02	0.49	0.41	0.02	-0.31	-0.13
Total Vehicles	-0.25	0.01	-0.01	0.02	0.04	0.04	0.13	0.08	0.05	0.11	0.43	0.33	-0.17	-0.38	0.00
Weekday	0.01	-0.02	0.02	0.13	0.05	0.16	0.36	0.11	0.03	0.12	0.04	0.04	-0.07	-0.02	0.00
Australian Born	-0.04	0.00	0.00	0.00	-0.01	-0.01	-0.06	0.01	0.00	0.01	-0.08	-0.03	0.11	0.04	-0.05
Car Licence	0.30	-0.10	0.10	0.09	0.21	0.24	0.12	0.11	0.14	0.21	-0.22	-0.15	-0.23	0.00	0.28
Bike Licence	0.10	-0.10	0.10	0.01	-0.03	-0.01	-0.09	0.00	-0.02	-0.02	-0.09	-0.08	-0.01	-0.03	0.14
No Licence	-0.19	0.04	-0.04	-0.07	-0.19	-0.21	-0.13	-0.09	-0.16	-0.21	0.12	0.09	0.21	0.05	-0.22
Full Time Work	0.12	-0.10	0.10	0.16	0.27	0.35	0.26	0.25	0.24	0.42	-0.18	-0.12	-0.19	0.01	0.22
Any Work	0.02	-0.03	0.03	0.11	0.22	0.26	0.33	0.19	0.19	0.32	-0.10	-0.07	-0.17	-0.06	0.16
Student	-0.10	-0.02	0.02	-0.01	-0.03	-0.03	-0.01	-0.01	-0.03	-0.04	0.05	0.09	-0.02	-0.02	-0.03
Household Income	-0.15	-0.02	0.02	0.10	0.12	0.19	0.17	0.15	0.10	0.21	0.36	0.27	-0.22	-0.30	0.04
High Personal Income	0.13	-0.10	0.10	0.09	0.16	0.20	0.11	0.12	0.10	0.18	-0.07	0.06	-0.10	0.00	0.13
Medium Personal Income	0.20	0.04	-0.04	0.08	0.17	0.20	0.14	0.13	0.16	0.25	-0.20	-0.21	-0.09	0.08	0.13
Low Personal Income	-0.27	0.02	-0.02	-0.13	-0.26	-0.32	-0.20	-0.20	-0.22	-0.35	0.24	0.16	0.15	-0.08	-0.21
Vehicle Ownership	-0.20	-0.06	0.06	0.06	0.05	0.09	0.15	0.10	0.03	0.11	0.27	0.20	-0.13	-0.35	0.12

Table B3: Correlation Analysis of the Socio-Demographic and Trip Variables (cont.)

	Other HH Type	House	Flat	Townhouse	Number of Bedrooms	Owned Property	Lived In The Property	Number of Bicycle	Total Vehicles	Weekday	Australian Born	Car Licence	Bike Licence	No Licence
Age														
Gender														
Male														
AMPeak Departure														
PMPeak Departure														
PeakHour Departure														
Purpose Work														
AMPeak Arrival														
PMPeak Arrival														
PeakHour Arrival														
Household Size														
Couple with Kids														
One Parent														
Sole														
Couple														
Other HH Type	1.00													
House	0.03	1.00												
Flat	-0.03	-0.84	1.00											
Townhouse	0.00	-0.46	-0.09	1.00										
Number of Bedrooms	0.10	0.50	-0.50	-0.09	1.00									
Owned Property	-0.20	0.23	-0.25	-0.02	0.25	1.00								
Lived In The Property	0.01	0.16	-0.13	-0.07	0.10	0.35	1.00							
Number of Bicycle	-0.10	0.25	-0.21	-0.12	0.32	0.10	-0.10	1.00						
Total Vehicles	0.04	0.34	-0.30	-0.13	0.45	0.31	0.03	0.29	1.00					
Weekday	0.01	0.06	-0.04	-0.04	0.02	0.05	-0.02	0.02	0.00	1.00				
Australian Born	-0.02	0.14	-0.10	-0.10	0.05	0.14	0.20	0.10	0.09	0.02	1.00			
Car Licence	0.06	-0.02	0.03	-0.01	-0.05	0.16	-0.01	-0.11	0.12	0.04	-0.08	1.00		
Bike Licence	-0.01	0.05	-0.03	-0.03	0.02	0.05	0.06	0.00	0.06	-0.04	0.01	0.13	1.00	
No Licence	-0.06	-0.03	0.02	0.02	0.00	-0.16	0.03	0.07	-0.21	-0.05	0.09	-0.82	-0.12	1.00
Full Time Work	0.04	-0.01	0.00	0.02	-0.08	0.13	-0.03	-0.07	0.07	0.11	0.00	0.43	0.00	-0.38
Any Work	0.09	0.01	0.00	-0.01	-0.05	0.09	-0.08	-0.01	0.14	0.09	-0.02	0.41	0.04	-0.45
Student	-0.04	0.03	-0.03	-0.01	0.02	0.00	-0.02	0.04	-0.02	0.02	0.04	-0.07	-0.01	0.08
Household Income	0.02	0.23	-0.21	-0.07	0.33	0.27	-0.06	0.29	0.46	0.10	-0.04	0.21	-0.02	-0.22
High Personal Income	-0.13	0.05	-0.05	-0.01	-0.01	0.16	-0.03	0.02	0.03	0.08	-0.02	0.26	0.05	-0.23
Medium Personal Income	0.11	-0.04	0.03	0.02	-0.08	0.02	0.02	-0.12	0.02	0.03	-0.01	0.34	0.02	-0.29
Low Personal Income	-0.03	0.01	0.00	-0.02	0.09	-0.12	0.00	0.10	-0.04	-0.08	0.02	-0.50	-0.05	0.42
Vehicle Ownership	0.02	0.26	-0.21	-0.13	0.26	0.20	-0.06	0.21	0.57	0.06	0.05	0.22	0.03	-0.26

Table B3: Correlation Analysis of the Socio-Demographic and Trip Variables (cont.)

	Full Time Work	Any Work	Student	Household Income	High Personal Income	Medium Personal Income	Low Personal Income	Vehicle Ownership
Age								
Gender								
Male								
AMPeak Departure								
PMPeak Departure								
PeakHour Departure								
Purpose Work								
AMPeak Arrival								
PMPeak Arrival								
PeakHour Arrival								
Household Size								
Couple with Kids								
One Parent								
Sole								
Couple								
Other HH Type								
House								
Flat								
Townhouse								
Number of Bedrooms								
Owned Property								
Lived In The Property								
Number of Bicycle								
Total Vehicles								
Weekday								
Australian Born								
Car Licence								
Bike Licence								
No Licence								
Full Time Work	1.00							
Any Work	0.65	1.00						
Student	-0.05	-0.08	1.00					
Household Income	0.35	0.32	-0.03	1.00				
High Personal Income	0.35	0.27	-0.02	0.39	1.00			
Medium Personal Income	0.54	0.48	-0.05	0.10	-0.28	1.00		
Low Personal Income	-0.75	-0.63	0.06	-0.34	-0.36	-0.80	1.00	
Vehicle Ownership	0.16	0.20	-0.05	0.31	0.09	0.07	-0.13	1.00

APPENDIX C

STOP CHOICE MODEL ESTIMATION RESULTS

Model:	**MNL**				
Number of estimated parameters:	7				
Number of observations:	1238				
Number of individuals:	1238				
Null log-likelihood:	-2937.928				
Init log-likelihood:	-2937.928				
Final log-likelihood:	-1977.056				
Likelihood ratio test:	1921.744				
Rho-square:	0.327				
Adjusted rho-square:	0.325				

Utility parameters

Name	Value	Robust Std err	Robust t-test	p-value	
B_AccessTime	-0.196	0.00939	-20.91	0	
B_MinWalkTime	-0.0337	0.0105	-3.21	0	
B_NumofRoutes	0.059	0.0138	4.27	0	
B_Str_TT	0.656	0.0887	7.39	0	
B_Str_Tr	1.4	0.0982	14.23	0	
B_Train	2.32	0.123	18.87	0	
B_stopLight	0.393	0.095	4.14	0	

Model:	**TT**
f estimated parameters:	9
umber of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-1984.771
Likelihood ratio test:	1906.315
Rho-square:	0.324
Adjusted rho-square:	0.321

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_AccessTime	-0.134	0.014	-9.58	0		0.0132	-10.19	0	
B_AustralianBorn_TT	0.695	0.11	6.34	0		0.11	6.29	0	
B_MinTransfers	-0.867	0.0778	-11.15	0		0.0819	-10.59	0	
B_MinWalkTime	-0.0304	0.0092	-3.3	0		0.0091	-3.35	0	
B_NumofRoutes	0.0528	0.013	4.07	0		0.0129	4.11	0	
B_Train	1.96	0.149	13.16	0		0.144	13.6	0	
B_stopLight	0.307	0.0845	3.63	0		0.0831	3.69	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std e	bust t-tes	p-value	bust t-tes	p-value
MTT	1.24	0.108	11.48	0	2.21	0.03		0.111	11.13	0	2.15	0.03
NoMTT	1.21	0.0831	14.58	0	2.55	0.01		0.0802	15.12	0	2.65	0.01

Model:	**AT**
Number of estimated parameters:	10
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-2054.072
Likelihood ratio test:	1767.713
Rho-square:	0.301
Adjusted rho-square:	0.297
BIC	4179

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_FastestTT	-0.0085	0.0038	-2.25	0.02		0.0041	-2.09	0.04	
B_HHSize_AT	0.126	0.028	4.49	0		0.0284	4.44	0	
B_MinTransfers	-0.728	0.0687	-10.6	0		0.0761	-9.58	0	
B_MinWalkTime	-0.104	0.0091	-11.43	0		0.0095	-10.94	0	
B_NumofRoutes	0.0542	0.0109	4.97	0		0.0105	5.16	0	
B_PMPeak_AT	-0.54	0.194	-2.79	0.01		0.193	-2.79	0.01	
B_Train	1.72	0.144	11.9	0		0.144	11.88	0	
B_stopLight	0.284	0.0768	3.69	0		0.0751	3.78	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinAT	1.21	0.119	10.15	0	1.76	0.08	*	0.128	9.43	0	1.63	0.1
NoMinAT	1.4	0.103	13.6	0	3.86	0		0.104	13.43	0	3.81	0

Model:	**Tr**
Number of estimated parameters:	13
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-1971.252
Likelihood ratio test:	1933.353
Rho-square:	0.329
Adjusted rho-square:	0.325
BIC	4035

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_AccessTime	-0.166	0.0165	-10.04	0		0.0174	-9.56	0	
B_AustralianBorn_Tr	0.543	0.161	3.38	0		0.159	3.43	0	
B_BikeLicence_Tr	-1.28	0.511	-2.5	0.01		0.523	-2.44	0.01	
B_FastestTT	-0.0254	0.0048	-5.33	0		0.0057	-4.44	0	
B_MediumPerIncome_Tr	0.392	0.167	2.34	0.02		0.168	2.33	0.02	
B_MinWalkTime	-0.0333	0.0097	-3.44	0		0.0098	-3.41	0	
B_NumofRoutes	0.0543	0.0126	4.31	0		0.013	4.19	0	
B_PMPeak_Dep_Tr	0.709	0.219	3.24	0		0.219	3.24	0	
B_Train	2.09	0.168	12.43	0		0.16	13.06	0	
B_Weekday_Tr	0.953	0.167	5.72	0		0.16	5.97	0	
B_stopLight	0.33	0.0864	3.82	0		0.0881	3.74	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinTr	1.28	0.107	11.94	0	2.63	0.01		0.114	11.22	0	2.47	0.01
NoMinTr	1.05	0.0796	13.15	0	0.59	0.56	*	0.0755	13.87	0	0.62	0.53

Model:	**TT-AT**
Number of estimated parameters:	13
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-1997.259
Likelihood ratio test:	1881.338
Rho-square:	0.32
Adjusted rho-square:	0.316
BIC	4087

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value
B_AccessTime	-0.156	0.0123	-12.64	0		0.0122	-12.73	0
B_AustralianBorn_TT	0.796	0.116	6.85	0		0.116	6.85	0
B_Flat_TT_AT	0.553	0.202	2.74	0.01		0.188	2.95	0
B_HHSize_TT_AT	0.228	0.0294	7.74	0		0.0288	7.91	0
B_MinWalkTime	-0.0266	0.0089	-2.97	0		0.009	-2.96	0
B_Mintransfers	-0.859	0.0686	-12.53	0		0.0758	-11.33	0
B_NumofRoutes	0.041	0.0124	3.32	0		0.0124	3.32	0
B_Train	1.8	0.133	13.52	0		0.129	13.94	0
B_stopLight	0.453	0.0799	5.67	0		0.0794	5.7	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinAccessTime	1.25	0.174	7.21	0	1.46	0.14	*	0.177	7.08	0	1.43	0.15
MinTravelTime	1.43	0.147	9.77	0	2.95	0		0.149	9.62	0	2.9	0
NO_Strategy	1.28	0.0805	15.89	0	3.47	0		0.0778	16.44	0	3.59	0
TT_and_AT	1.7	0.268	6.34	0	2.6	0.01		0.254	6.69	0	2.75	0.01

Model:	**TT-Tr**
Number of estimated parameters:	19
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2776.403
Final log-likelihood:	-1942.09
Likelihood ratio test:	1991.676
Rho-square:	0.339
Adjusted rho-square:	0.332
BIC	4019

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_AccessTime	-0.128	0.013	-9.82	0		0.0126	-10.09	0	
B_AustralianBorn_TT	0.788	0.228	3.46	0		0.235	3.35	0	
B_AustralianBorn_TT_Tr	1.02	0.205	4.99	0		0.206	4.96	0	
B_AustralianBorn_Tr	0.774	0.212	3.65	0		0.216	3.58	0	
B_MediumPerIncome_TT	-0.463	0.216	-2.15	0.03		0.208	-2.23	0.03	
B_MinWalkTime	-0.0366	0.0092	-3.96	0		0.009	-4.08	0	
B_NoLicence_Tr	0.407	0.173	2.36	0.02		0.175	2.33	0.02	
B_NumofRoutes	0.042	0.013	3.23	0		0.0129	3.25	0	
B_PMPeak_Dep_TT_Tr	0.652	0.254	2.57	0.01		0.248	2.63	0.01	
B_PMPeak_Dep_Tr	1.02	0.245	4.15	0		0.252	4.03	0	
B_TOTALVEHS_TT	0.244	0.0929	2.63	0.01		0.0887	2.75	0.01	
B_Train	2.03	0.137	14.81	0		0.136	14.89	0	
B_Weekday_TT_Tr	1.38	0.183	7.55	0		0.178	7.77	0	
B_Weekday_Tr	0.787	0.194	4.07	0		0.197	3.99	0	
B_stopLight	0.291	0.0824	3.53	0		0.0833	3.49	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinTransfer	1.52	0.135	11.26	0	3.85	0		0.146	10.42	0	3.57	0
MinTravelTime	1.19	0.136	8.77	0	1.41	0.16	*	0.137	8.68	0	1.4	0.16
NO_Strategy	1.07	0.0769	13.86	0	0.85	0.39	*	0.0784	13.59	0	0.84	0.4
TT_and_Transfer	1.38	0.151	9.14	0	2.51	0.01		0.157	8.77	0	2.41	0.02

Model:	**AT-Tr**
Number of estimated parameters:	17
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-2013.01
Likelihood ratio test:	1849.836
Rho-square:	0.315
Adjusted rho-square:	0.309
BIC	4147

Utility parameters

Name	Value	Std err	t-test	p-value
B_AustralianBorn_Tr	0.422	0.143	2.95	0
B_BikeLicence_AT	1.11	0.519	2.14	0.03
B_FastestTT	-0.0104	0.004	-2.61	0.01
B_HHSize_AT_Tr	0.254	0.0608	4.17	0
B_MinWalkTime	-0.118	0.00801	-14.74	0
B_NumofRoutes	0.0451	0.0106	4.26	0
B_PMPeak_Arv_Tr	0.574	0.169	3.39	0
B_PeakHour_Arv_NoStr	-0.556	0.172	-3.22	0
B_PersonalVeh_AT_Tr	-0.483	0.211	-2.29	0.02
B_Train	1.89	0.127	14.86	0
B_Weekday_AT_Tr	1.06	0.21	5.05	0
B_Weekday_Tr	0.801	0.154	5.2	0
B_stopLight	0.319	0.0787	4.05	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value
AT_and_Transfer	1	6.57E-09	152320351	0	0	1
MinAccessTime	1	1.80e+30	0	1	0	1
MinTransfer	1.65	0.127	12.96	0	5.1	0
NO_Strategy	1.1	0.0686	16.09	0	1.52	0.13

Model:	**TT-AT-Tr**
Number of estimated parameters	24
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.826
Init log-likelihood:	-2937.826
Final log-likelihood:	-1966.695
Likelihood ratio test:	1942.263
Rho-square:	0.331
Adjusted rho-square:	0.322
BIC	4104

Utility parameters

Name	Value	Std err	t-test	p-value
B_AustralianBorn_NoStr	-0.532	0.191	-2.79	0.01
B_AustralianBorn_TT_Tr	0.68	0.184	3.71	0
B_BEDROOMS_AT_Tr	0.363	0.0624	5.81	0
B_BEDROOMS_TT_AT	0.237	0.0497	4.77	0
B_MinWalkTime	-0.0981	0.00775	-12.65	0
B_NumofRoutes	0.0329	0.0116	2.84	0
B_OtherHHType_AT_Tr	0.91	0.337	2.7	0.01
B_PMPeak_Dep_TT_Tr	0.707	0.25	2.82	0
B_PMPeak_Dep_Tr	0.913	0.221	4.13	0
B_PeakHour_Arv_TT	-0.739	0.306	-2.41	0.02
B_Train	1.86	0.127	14.59	0
B_Weekday_TT	0.796	0.254	3.14	0
B_Weekday_TT_AT_Tr	1.93	0.159	12.17	0
B_Weekday_TT_Tr	1.21	0.198	6.08	0
B_Weekday_Tr	0.948	0.167	5.68	0
B_stopLight	0.271	0.0754	3.59	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value
ATandTr	1.42	0.451	3.14	0	0.92	0.36
MinAccessTime	1.03	0.152	6.77	0	0.21	0.84
MinTransfer	1.79	0.179	9.97	0	4.39	0
MinTravelTime	1.38	0.277	5	0	1.39	0.16
NO_Strategy	1.12	0.0852	13.08	0	1.35	0.18
TTandAT	1.18	0.247	4.77	0	0.72	0.47
TTandATandTr	1	9.81E-09	101954190	0	0	1
TTandTr	2.46	0.39	6.31	0	3.74	0

Model:	**Mixed MNL**
Number of draws:	100
Number of estimated parameters:	9
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-5930.679
Final log-likelihood:	-2002.611
Likelihood ratio test:	1870.635
Rho-square:	0.318
Adjusted rho-square:	0.315
BIC	4069.3133

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value
B_AccessTime	-1.63	0.0514	-31.62	0		0.0527	-30.85	0
B_AccessTime_S	0.391	0.112	3.89	0		0.108	4.03	0
B_MinWalkTime	-4.26	0.884	-4.92	0		0.744	-5.85	0
B_MinWalkTime_S	1.74	0.703	2.62	0.01		0.521	3.54	0
B_NumofRoutes	0.0582	0.0143	4.09	0		0.015	3.9	0
B_Str_TT	0.713	0.0893	7.98	0		0.091	7.84	0
B_Str_Tr	1.43	0.0962	14.89	0		0.104	13.84	0
B_Train	1.36	0.0839	16.26	0		0.0879	15.53	0
B_stopLight	0.537	0.0958	5.63	0		0.093	5.79	0

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_AccessTime_B_AccessTime_S	0.592	0.0972	6.1		
B_MinWalkTime_B_MinWalkTime_S	7.11	2.59	2.74		

Model:	**TT[M]**
Number of estimated parameters:	13
Number of observations:	1237
Number of individuals:	1237
Null log-likelihood:	-2934.8
Init log-likelihood:	-5623.5
Final log-likelihood:	-2002.2
Likelihood ratio test:	1905.09
Rho-square:	0.32
Adjusted rho-square:	0.315
BIC	4097.06

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value
B_AccessTime	-0.15	0.0146	-10.31	0		0.0141	-10.67	0
B_AustralianBorn_TT	0.419	0.165	2.53	0.01		0.172	2.44	0.01
B_Map	0.171	0.0144	11.89	0		0.0142	12.05	0
B_MinTransfers	0.326	0.207	1.57	0.12	*	0.0949	3.43	0
B_MinTransfers_S	-0.749	0.164	-4.56	0		0.112	-6.7	0
B_MinWalkTime	-4.23	0.788	-5.36	0		0.739	-5.72	0
B_MinWalkTime_S	-1.79	0.604	-2.96	0		0.562	-3.18	0
B_NumofRoutes	0.0417	0.0149	2.8	0.01		0.0147	2.83	0
B_PMPeak_Arv_TT	-0.435	0.192	-2.26	0.02		0.199	-2.19	0.03
B_Weekday_TT	0.506	0.164	3.08	0		0.174	2.91	0
B_stopLight	0.454	0.0972	4.67	0		0.0962	4.71	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value	
MTT	1.12	0.0953	11.72	0	1.23	0.22	*	0.0972	11.49	0	1.2	0.23	*
NoMTT	1.1	0.0753	14.58	0	1.31	0.19	*	0.0796	13.8	0	1.24	0.21	*

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinTransfers_B_MinTransfers_S	0.56	0.246	2.28		
B_MinWalkTime_B_MinWalkTime_S	3.19	2.16	1.48		

Model:	**AT[M]**
Number of estimated parameters:	12
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.9
Init log-likelihood:	-2802.9
Final log-likelihood:	-1994.5
Likelihood ratio test:	1926.79
Rho-square:	0.323
Adjusted rho-square:	0.318
BIC	4074.52

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value
B_FastestTT	-0.0258	0.0048	-5.43	0		0.0055	-4.71	0
B_HHSize_AT	0.128	0.0419	3.05	0		0.0407	3.15	0
B_MinTransfers	-0.32	0.13	-2.46	0.01		0.132	-2.43	0.02
B_MinTransfers_S	1.03	0.208	4.98	0		0.192	5.4	0
B_MinWalkTime	-0.0241	0.0093	-2.6	0.01		0.0097	-2.48	0.01
B_NumofRoutes	0.0556	0.0121	4.58	0		0.012	4.64	0
B_PMPeak_AT	-0.484	0.205	-2.37	0.02		0.212	-2.28	0.02
B_PersonalVeh_AT	-0.369	0.159	-2.32	0.02		0.159	-2.31	0.02
B_Train	2.53	0.337	7.51	0		0.346	7.32	0
B_stopLight	0.313	0.0821	3.82	0		0.081	3.87	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinAT	1.29	0.144	8.99	0	2.04	0.04		0.166	7.82	0	1.78	0.08
NoMinAT	1.31	0.105	12.53	0	2.99	0		0.105	12.53	0	2.99	0

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinTransfers_B_MinTransfers_S	1.07	0.43	2.49		

Model:	**TR[M]**
Number of estimated parameters:	13
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.93
Init log-likelihood:	-2802.92
Final log-likelihood:	-1990.11
Likelihood ratio test:	1955.64
Rho-square:	0.326
Adjusted rho-square:	0.32
BIC	4072.79

Utility parameters

Name	Value	Std err	t-test	p-value
B_AccessTime	-0.173	0.0153	-11.32	0
B_AustralianBorn_Tr	0.417	0.172	2.43	0.02
B_BikeLicence_Tr	-1.38	0.53	-2.59	0.01
B_FastestTT	-0.0257	0.0048	-5.33	0
B_MediumPerIncome_Tr	0.346	0.173	2	0.05
B_MinWalkTime	-0.0345	0.0098	-3.52	0
B_MinWalkTime_S	-0.0607	0.0163	-3.72	0
B_NumofRoutes	0.0507	0.0128	3.97	0
B_PMPeak_Dep_Tr	0.725	0.219	3.3	0
B_Train	2.77	0.312	8.88	0
B_Weekday_Tr	0.585	0.203	2.88	0
B_stopLight	0.318	0.0874	3.64	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value
MinTr	1.23	0.0775	15.89	0	2.98	0
NoMinTr	1	1.80e+30	0	1	0	1

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinWalkTime_B_MinWalkTime_S	0.00345	0.0019	1.1		

Model:	**TT-AT[M]**
Number of estimated parameters:	16
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-2006.316
Likelihood ratio test:	1923.224
Rho-square:	0.32
Adjusted rho-square:	0.314
BIC	4126.57

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_AustralianBorn_TT	0.845	0.125	6.77	0		0.126	6.69	0	
B_Flat_TT_AT	0.587	0.213	2.75	0.01		0.195	3.01	0	
B_HHSize_TT_AT	0.263	0.0357	7.36	0		0.0351	7.49	0	
B_Male_TT_AT	-0.331	0.162	-2.04	0.04		0.164	-2.02	0.04	
B_MinTransfers	-0.878	0.0739	-11.87	0		0.0805	-10.9	0	
B_MinWalkTime	-0.0397	0.0119	-3.34	0		0.0127	-3.14	0	
B_MinWalkTime_S	-0.0607	0.0163	-3.72	0		0.0171	-3.55	0	
B_NumofRoutes	0.0448	0.0128	3.5	0		0.0126	3.56	0	
B_PMPeak_TT_AT	-0.406	0.235	-1.72	0.08	*	0.234	-1.74	0.08	*
B_Student_TT_AT	20.1	1.65E+04	0	1	*	0.75	26.76	0	
B_Train	1.09	0.195	5.61	0		0.2	5.48	0	
B_stopLight	0.403	0.0841	4.79	0		0.0837	4.81	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinAccessTime	1.27	0.18	7.06	0	1.51	0.13	*	0.188	6.76	0	1.44	0.15
MinTravelTime	1.35	0.14	9.63	0	2.5	0.01		0.144	9.4	0	2.44	0.01
NO_Strategy	1.23	0.0819	14.97	0	2.75	0.01		0.0818	14.97	0	2.75	0.01
TT_and_AT	1.65	0.277	5.97	0	2.36	0.02		0.258	6.41	0	2.53	0.01

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinWalkTime_B_MinWalkTime_S	0.00369	0.00198	1.86		

Model:	**TT-TR[N]**
Number of estimated parameters:	20
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.93
Init log-likelihood:	-2776.4
Final log-likelihood:	-1968.77
Likelihood ratio test:	2018.315
Rho-square:	0.334
Adjusted rho-square:	0.327
BIC	4079.97

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
B_AccessTime	-0.133	0.014	-9.52	0		0.0136	-9.84	0	
B_AustralianBorn_TT	0.796	0.237	3.36	0		0.246	3.24	0	
B_AustralianBorn_TT_Tr	0.918	0.213	4.31	0		0.214	4.3	0	
B_AustralianBorn_Tr	0.69	0.219	3.15	0		0.221	3.12	0	
B_MediumPerIncome_TT	-0.482	0.224	-2.16	0.03		0.217	-2.22	0.03	
B_MinWalkTime	-0.0469	0.0117	-4.02	0		0.0119	-3.94	0	
B_MinWalkTime_S	-0.0597	0.0157	-3.81	0		0.017	-3.52	0	
B_NoLicence_Tr	0.361	0.179	2.02	0.04		0.179	2.02	0.04	
B_NumofRoutes	0.0393	0.0133	2.95	0		0.0131	2.99	0	
B_PMPeak_Dep_TT_Tr	0.617	0.257	2.39	0.02		0.249	2.48	0.01	
B_PMPeak_Dep_Tr	1.04	0.251	4.16	0		0.257	4.05	0	
B_TOTALVEHS_TT	0.275	0.0966	2.85	0		0.0922	2.98	0	
B_Train	2.71	0.309	8.77	0		0.318	8.52	0	
B_Weekday_TT_Tr	1.02	0.211	4.84	0		0.207	4.93	0	
B_Weekday_Tr	0.417	0.219	1.91	0.06	*	0.22	1.9	0.06	*
B_stopLight	0.292	0.0853	3.43	0		0.0859	3.4	0	

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
MinTransfer	1.54	0.147	10.5	0	3.7	0		0.158	9.8	0	3.46	0
MinTravelTime	1.18	0.144	8.21	0	1.26	0.21	*	0.148	8	0	1.23	0.22
NO_Strategy	1	0.0772	13.02	0	0.06	0.95	*	0.0772	13.02	0	0.06	0.95
TT_and_Transfer	1.34	0.154	8.66	0	2.19	0.03		0.157	8.5	0	2.15	0.03

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinWalkTime_B_MinWalkTime_S	0.00357	0.0019	1.9		

Model:	**AT-TR[M]**
Number of estimated parameters:	20
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-1978.566
Likelihood ratio test:	1998.724
Rho-square:	0.333
Adjusted rho-square:	0.325
BIC	4099.56

Utility parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value
B_AustralianBorn_Tr	0.507	0.154	3.28	0		0.158	3.21	0
B_BikeLicence_AT	1.17	0.553	2.12	0.03		0.515	2.28	0.02
B_FastestTT	-0.0288	0.005	-5.79	0		0.0059	-4.89	0
B_HHSize_AT_Tr	0.247	0.0649	3.81	0		0.0655	3.78	0
B_MinWalkTime	-0.0509	0.0126	-4.05	0		0.0128	-3.97	0
B_MinWalkTime_S	-0.0789	0.0142	-5.54	0		0.0145	-5.44	0
B_NumofRoutes	0.0478	0.012	3.98	0		0.0119	4.01	0
B_PMPeak_Arv_Tr	0.767	0.176	4.36	0		0.172	4.46	0
B_PersonalVeh_AT_Tr	-0.601	0.225	-2.67	0.01		0.225	-2.67	0.01
B_Student_AT_Tr	7.04	65.3	0.11	0.91	*	0.738	9.54	0
B_Student_NoStr	9.99	65.3	0.15	0.88	*	1.23	8.13	0
B_Student_Tr	-6.03	618	-0.01	0.99	*	0.432	-13.96	0
B_Train	2.66	0.334	7.97	0		0.337	7.89	0
B_Weekday_AT_Tr	0.785	0.252	3.12	0		0.256	3.06	0
B_Weekday_Tr	0.72	0.197	3.66	0		0.205	3.51	0
B_stopLight	0.345	0.0864	3.99	0		0.0864	3.99	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value		Robust Std err	Robust t-test 0	p-value	Robust t-test 1	p-value
AT_and_Transfer	1.08	0.212	5.09	0	0.37	0.71	*	0.24	4.49	0	0.33	0.74
MinAccessTime	1.04	0.12	8.68	0	0.31	0.75	*	0.124	8.38	0	0.3	0.76
MinTransfer	1.51	0.133	11.29	0	3.8	0		0.138	10.89	0	3.67	0
NO_Strategy	1.05	0.0792	13.21	0	0.58	0.57	*	0.078	13.41	0	0.58	0.56

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinWalkTime_B_MinWalkTime_S	0.00622	0.0023	2.77		

Model:	**TT-AT-TR[M]**
Number of estimated parameters:	24
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.826
Init log-likelihood:	-2937.826
Final log-likelihood:	-1956.693
Likelihood ratio test:	2042.266
Rho-square:	0.338
Adjusted rho-square:	0.328
BIC	4084.30

Utility parameters

Name	Value	Std err	t-test	p-value
B_AustralianBorn_TT_Tr	0.771	0.189	4.07	0
B_BEDROOMS_AT_Tr	0.255	0.0676	3.77	0
B_BEDROOMS_TT_AT	0.249	0.0491	5.06	0
B_MinWalkTime	-0.0457	0.0119	-3.85	0
B_MinWalkTime_S	-0.0595	0.0155	-3.83	0
B_NumofRoutes	0.0412	0.0129	3.2	0
B_OtherHHType_AT_Tr	0.821	0.348	2.36	0.02
B_PMPeak_Dep_TT_Tr	0.71	0.259	2.74	0.01
B_PMPeak_Dep_Tr	0.956	0.23	4.16	0
B_PeakHour_Arv_TT	-0.714	0.317	-2.26	0.02
B_Train	2.56	0.304	8.42	0
B_Weekday_TT	1.16	0.256	4.54	0
B_Weekday_TT_AT_Tr	1.51	0.187	8.09	0
B_Weekday_TT_Tr	1.01	0.222	4.54	0
B_Weekday_Tr	0.812	0.194	4.19	0
B_stopLight	0.275	0.0916	3	0

Model parameters

Name	Value	Std err	t-test 0	p-value	t-test 1	p-value
ATandTr	3.31	2.38	1.39	0.16	0.97	0.33
MinAccessTime	1	1.80e+308	0	1	0	1
MinTransfer	1.6	0.163	9.83	0	3.7	0
MinTravelTime	1.37	0.288	4.75	0	1.27	0.2
NO_Strategy	1.06	0.0748	14.17	0	0.8	0.42
TTandAT	1.36	0.371	3.66	0	0.97	0.33
TTandATandTr	1	2.20E-08	45529634	0	0	1
TTandTr	1.86	0.295	6.32	0	2.92	0

Variance of random coefficients

Name	Value	Std err	t-test	Robust Std err	Robust t-test
B_MinWalkTime_B_MinWalkTime_S	0.00355	0.00185	1.91		

Model:	**P-RRM**						
Number of estimated parameters:	4						
Number of observations:	1238						
Number of individuals:	1238						
Null log-likelihood:	-2937.9						
Init log-likelihood:	-2937.9						
Final log-likelihood:	-2629.4						
Likelihood ratio test:	616.98						
Rho-square:	0.105						
Adjusted rho-square:	0.104						
BIC	5287						
Utility parameters							
Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
B_AccessTime	-0.0004	3E-05	-15.45	0	3E-05	-12.72	0
B_Str_TT	0.0064	0.0014	4.47	0	0.0016	4.11	0
B_Train	0.0778	0.0064	12.11	0	0.0104	7.49	0
B_stopLight	0.0063	0.0015	4.15	0	0.0016	3.98	0

Model:	**muRRM**
Number of estimated parameters:	7
Number of observations:	1238
Number of individuals:	1238
Null log-likelihood:	-2937.928
Init log-likelihood:	-2937.928
Final log-likelihood:	-2802.468
Likelihood ratio test:	270.921
Rho-square:	0.046
Adjusted rho-square:	0.044
BIC	5655

Utility parameters

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
B_AccessTime	-0.0336	0.0028	-12.07	0	0.00324	-10.35	0
B_MinOtherWalkTime	-	-	-	-	-	-	-
B_NumofRoutes	0.00396	0.0036	1.09	0.28	0.00186	2.12	0.03
B_Str_TT	0.00528	0.0138	0.38	0.7	0.00373	1.42	0.16
B_Str_Tr	0.0143	0.0148	0.97	0.33	0.00615	2.32	0.02
B_Train	0.00526	0.0144	0.37	0.71	0.00271	1.94	0.05
B_stopLight	0.0104	0.0148	0.7	0.48	0.00512	2.04	0.04
mu	0.0658	0.0277	2.37	0.02	0.02	3.29	0

Model:	**RRM-RUM Hybrid**						
Number of estimated parameters:	7						
Number of observations:	1238						
Number of individuals:	1238						
Null log-likelihood:	-2937.928						
Init log-likelihood:	-2937.928						
Final log-likelihood:	-1982.286						
Likelihood ratio test:	1911.284						
Rho-square:	0.325						
Adjusted rho-square:	0.323						
BIC	4014						
Utility parameters							
Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
B_AccessTime	-0.198	0.0088	-22.53	0	0.00941	-21	0
B_MinOtherWalkTime	-0.038	0.0106	-3.62	0	0.0107	-3.59	0
B_NumofRoutes	0.0005	0.0002	2.56	0.01	0.000172	2.68	0.01
B_Str_TT	0.736	0.0841	8.75	0	0.0869	8.47	0
B_Str_Tr	1.46	0.0897	16.23	0	0.0965	15.08	0
B_Train	2.3	0.123	18.65	0	0.123	18.71	0
B_stopLight	0.444	0.0947	4.69	0	0.093	4.77	0

APPENDIX D

ROUTE CHOICE MODEL ESTIMATION RESULTS

Model Name: MNL-FL Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.202	-0.207	-0.208	-0.212	-0.211	-0.213	-0.213	-0.213	-0.215	-0.216
TravelTime	-0.046	-0.052	-0.051	-0.052	-0.051	-0.053	-0.053	-0.054	-0.055	-0.054
Number of Transfer	-1.384	-1.427	-1.452	-1.470	-1.485	-1.482	-1.496	-1.495	-1.507	-1.510
Train	2.579	2.559	2.559	2.565	2.507	2.510	2.512	2.502	2.521	2.524
Bus	-0.573	-0.567	-0.569	-0.553	-0.562	-0.567	-0.582	-0.570	-0.585	-0.585
LightingStop	0.290	0.299	0.305	0.297	0.286	0.283	0.280	0.268	0.270	0.270
Travel Time Strategy	0.434	0.404	0.452	0.464	0.478	0.453	0.459	0.456	0.446	0.468
t-statistics of the Estimated Parmeters										
Access Time	-8.860	-8.987	-8.945	-8.851	-8.864	-8.988	-8.891	-8.869	-8.900	-8.897
TravelTime	-4.439	-4.506	-4.336	-4.434	-4.556	-4.334	-4.572	-4.668	-4.320	-4.585
Number of Transfer	-2.197	-2.670	-2.973	-1.608	-3.142	-2.653	-4.103	-1.711	-2.650	-3.339
Train	8.316	9.023	8.952	7.453	10.075	8.291	9.594	8.146	9.093	8.241
Bus	-2.348	-2.118	-2.062	-2.110	-2.055	-2.293	-2.106	-2.316	-2.280	-2.119
LightingStop	2.453	2.770	2.737	2.461	2.679	2.572	2.676	2.487	2.503	2.398
Travel Time Strategy	1.851	1.956	2.144	1.859	2.079	1.904	2.123	1.877	1.985	2.001
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1242	-1690	-1914	-2049	-2137	-2205	-2250	-2292	-2319	-2347
Final Log Likelihood	-710	-1027	-1202	-1302	-1375	-1430	-1465	-1500	-1518	-1541
Rho_Square	0.428	0.392	0.372	0.364	0.356	0.351	0.349	0.345	0.345	0.345
BIC	1467	2101	2452	2652	2797	2908	2978	3047	3083	3129

Model Name: MNL-SRS Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.190	-0.195	-0.195	-0.195	-0.194	-0.195	-0.195	-0.195	-0.195	-0.194
TravelTime	-0.053	-0.054	-0.055	-0.055	-0.054	-0.054	-0.054	-0.054	-0.054	-0.052
Number of Transfer	-1.177	-1.206	-1.215	-1.221	-1.225	-1.230	-1.236	-1.236	-1.236	-1.510
Train	2.600	2.526	2.522	2.513	2.484	2.468	2.488	2.474	2.479	2.540
Bus	-0.615	-0.612	-0.613	-0.611	-0.621	-0.631	-0.642	-0.643	-0.645	-0.633
LightingStop	0.331	0.320	0.288	0.285	0.278	0.272	0.268	0.264	0.263	0.275
Travel Time Strategy	0.354	0.375	0.388	0.405	0.411	0.426	0.424	0.437	0.437	0.467
t-statistics of the Estimated Parmeters										
Access Time	-10.183	-11.277	-11.579	-11.180	-11.047	-11.437	-11.306	-11.518	-11.510	-11.663
TravelTime	-4.579	-5.716	-5.483	-5.597	-5.789	-5.870	-6.153	-6.136	-5.926	-5.970
Number of Transfer	-8.498	-9.650	-10.140	-10.168	-10.259	-10.254	-10.373	-10.301	-10.234	-10.265
Train	7.223	8.668	9.223	9.593	9.753	9.765	9.741	9.892	9.939	10.294
Bus	-2.788	-3.048	-3.171	-3.217	-3.401	-3.458	-3.360	-3.420	-3.519	-3.525
LightingStop	2.178	2.380	2.254	2.271	2.096	2.150	2.070	2.261	2.159	2.260
Travel Time Strategy	1.836	2.142	2.365	2.442	2.649	2.603	2.840	2.737	2.795	2.783
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1334	-1796	-2013	-2141	-2223	-2281	-2322	-2354	-2378	-2397
Final Log Likelihood	-734	-1070	-1243	-1346	-1416	-1466	-1500	-1528	-1550	-1565
Rho_Square	0.450	0.404	0.383	0.371	0.363	0.357	0.354	0.351	0.348	0.345
BIC	1516	2187	2533	2739	2880	2979	3048	3104	3147	3178

Model Name: TT Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.118	-0.093	-0.086	-0.080	-0.078	-0.079	-0.078	-0.081	-0.082	-0.087
TravelTime	-0.063	-0.050	-0.046	-0.043	-0.041	-0.040	-0.039	-0.039	-0.039	-0.041
Number of Transfer	-0.744	-0.584	-0.554	-0.516	-0.505	-0.518	-0.519	-0.536	-0.544	-0.575
Train	1.549	1.228	1.140	1.065	1.027	1.028	1.031	1.056	1.068	1.135
Bus	-0.436	-0.299	-0.254	-0.227	-0.217	-0.218	-0.216	-0.226	-0.223	-0.238
LightingStop	0.138	0.108	0.098	0.091	0.084	0.087	0.088	0.088	0.089	0.097
Weekday X Walk Time	-0.013	-0.010	-0.010	-0.010	-0.010	-0.009	-0.010	-0.010	-0.010	-0.011
Estimated Nest Co-efficients (μ)										
MTT	0.594	0.474	0.438	0.400	0.389	0.403	0.402	0.410	0.418	0.441
MoMTT	0.560	0.418	0.378	0.352	0.341	0.340	0.339	0.351	0.356	0.379
t-statistics of the Estimated Parmeters										
Access Time	-5.489	-4.712	-4.754	-4.658	-4.719	-4.970	-4.996	-4.452	-4.487	-4.820
TravelTime	-2.082	-1.739	-1.609	-2.086	-2.341	-1.895	-2.035	-1.690	-1.612	-1.940
Number of Transfer	-5.243	-4.847	-4.805	-4.732	-4.611	-4.638	-4.819	-4.159	-4.390	-4.602
Train	4.896	4.286	4.631	4.407	4.399	4.892	4.792	4.248	4.567	4.158
Bus	-2.308	-2.197	-2.068	-1.829	-1.978	-1.948	-1.774	-1.774	-1.856	-1.747
LightingStop	1.359	1.326	1.367	1.306	1.207	1.315	1.290	1.279	1.225	1.264
Weekday X Walk Time	-1.133	-1.055	-1.281	-1.242	-1.293	-1.244	-1.296	-1.272	-1.281	-1.379
t-statistics of the Nest Co-efficients										
MTT	5.493	4.398	4.409	4.102	4.112	3.816	3.686	3.256	3.430	3.616
MoMTT	10.453	8.720	9.055	7.950	7.765	8.411	8.187	7.951	7.907	6.593
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	10	10	10	10	10	10	10	10	10	10
Initial Log Likelihood	-1966	-2379	-2554	-2622	-2648	-2665	-2655	-2633	-2627	-2594
Final Log Likelihood	-1057	-1300	-1413	-1484	-1523	-1553	-1575	-1587	-1599	-1602
Rho_Square	0.462	0.454	0.447	0.434	0.425	0.417	0.407	0.397	0.391	0.383
BIC	2182	2668	2894	3036	3114	3174	3217	3242	3266	3271

Model Name: AT Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.151	-0.119	-0.108	-0.107	-0.107	-0.107	-0.107	-0.111	-0.118	-0.120
TravelTime	-0.036	-0.032	-0.031	-0.032	-0.033	-0.033	-0.034	-0.034	-0.036	-0.037
Number of Transfer	-0.842	-0.720	-0.701	-0.702	-0.716	-0.728	-0.731	-0.765	-0.813	-0.828
Train	1.891	1.500	1.371	1.348	1.341	1.370	1.389	1.433	1.506	1.545
Bus	-0.372	-0.320	-0.320	-0.320	-0.324	-0.335	-0.342	-0.359	-0.391	-0.399
LightingStop	0.208	0.172	0.148	0.147	0.147	0.139	0.138	0.143	0.153	0.153
Weekday X Walk Time	-0.019	-0.016	-0.016	-0.016	-0.017	-0.017	-0.018	-0.018	-0.019	-0.019
Estimated Nest Co-efficients (μ)										
MAT	0.674	0.592	0.589	0.572	0.589	0.586	0.577	0.588	0.633	0.637
MoMAT	0.640	0.513	0.484	0.488	0.501	0.509	0.514	0.536	0.569	0.585
t-statistics of the Estimated Parmeters										
Access Time	-7.997	-8.107	-8.379	-8.447	-8.471	-8.357	-8.657	-8.921	-8.481	-8.574
TravelTime	-4.414	-4.991	-5.321	-5.421	-5.271	-5.209	-5.688	-5.398	-5.697	-5.620
Number of Transfer	-7.874	-8.079	-8.335	-8.487	-8.596	-8.577	-8.786	-9.014	-8.874	-8.809
Train	7.016	7.455	7.913	7.803	7.926	7.864	8.294	8.225	8.363	8.044
Bus	-2.479	-2.764	-3.047	-3.002	-3.023	-2.970	-3.098	-3.186	-3.335	-3.230
LightingStop	2.034	2.251	2.176	2.160	2.131	1.943	1.982	1.989	2.018	1.913
Weekday X Walk Time	-1.784	-1.932	-2.163	-2.174	-2.261	-2.182	-2.325	-2.336	-2.314	-2.230
t-statistics of the Nest Co-efficients										
MAT	9.626	8.316	8.115	7.984	8.148	7.796	8.522	8.441	8.426	8.312
MoMAT	14.556	14.922	15.001	14.535	14.389	13.684	14.414	14.109	13.917	13.406
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	9	9	9	9	9	9	9	9	9	9
Initial Log Likelihood	-1792	-2129	-2264	-2309	-2343	-2375	-2399	-2414	-2418	-2431
Final Log Likelihood	-972	-1194	-1307	-1381	-1433	-1471	-1501	-1526	-1537	-1557
Rho_Square	0.458	0.439	0.423	0.402	0.388	0.381	0.374	0.368	0.364	0.360
BIC	2004	2448	2675	2823	2928	3003	3062	3112	3134	3174

Model Name: TR Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.142	-0.102	-0.100	-0.094	-0.095	-0.093	-0.098	-0.101	-0.105	-0.107
TravelTime	-0.043	-0.031	-0.030	-0.028	-0.028	-0.027	-0.028	-0.029	-0.030	-0.031
Train	1.993	1.319	1.265	1.168	1.187	1.175	1.234	1.269	1.323	1.349
Bus	-0.444	-0.327	-0.333	-0.307	-0.318	-0.312	-0.322	-0.325	-0.339	-0.346
LightingStop	0.224	0.124	0.114	0.100	0.109	0.106	0.108	0.108	0.112	0.111
Weekday X Walk Time	-0.023	-0.014	-0.014	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013	-0.014
Estimated Nest Co-efficients (μ)										
MTR	0.691	0.484	0.475	0.444	0.451	0.441	0.462	0.476	0.496	0.504
MoMTR	0.774	0.500	0.467	0.423	0.424	0.416	0.429	0.432	0.450	0.462
t-statistics of the Estimated Parmeters										
Access Time	-8.696	-8.655	-8.686	-8.789	-8.442	-8.700	-8.624	-8.480	-8.462	-8.803
TravelTime	-5.250	-5.349	-5.229	-5.627	-5.361	-5.565	-5.496	-5.316	-5.490	-5.367
Train	7.404	7.241	7.356	7.329	7.345	7.390	7.494	7.594	7.333	7.796
Bus	-2.816	-3.134	-3.357	-3.520	-3.573	-3.598	-3.603	-3.493	-3.496	-3.579
LightingStop	2.196	1.892	1.816	1.831	1.893	1.923	1.903	1.785	1.848	1.803
Weekday X Walk Time	-2.136	-2.037	-1.996	-2.025	-2.075	-2.111	-2.015	-2.016	-2.014	-2.045
t-statistics of the Nest Co-efficients										
MTR	13.800	12.043	11.196	10.994	10.556	10.642	10.441	10.221	9.749	10.440
MoMTR	13.284	12.309	12.237	12.525	12.263	12.508	12.010	11.501	11.047	12.002
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	8	8	8	8	8	8	8	8	8	8
Initial Log Likelihood	-1861	-2307	-2445	-2547	-2578	-2582	-2586	-2589	-2576	-2566
Final Log Likelihood	-1035	-1308	-1409	-1486	-1528	-1561	-1585	-1602	-1607	-1614
Rho_Square	0.444	0.433	0.424	0.416	0.407	0.395	0.387	0.381	0.376	0.371
BIC	2123	2670	2873	3026	3110	3176	3224	3258	3268	3281

Model Name: TT-AT Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
TravelTime	-0.071	-0.062	-0.057	-0.055	-0.055	-0.053	-0.053	-0.053	-0.053	-0.054
Number of Transfer	-0.852	-0.723	-0.682	-0.678	-0.706	-0.704	-0.714	-0.731	-0.750	-0.789
Train	1.789	1.442	1.304	1.277	1.296	1.272	1.290	1.311	1.335	1.417
Bus	-0.491	-0.347	-0.307	-0.288	-0.279	-0.270	-0.272	-0.281	-0.283	-0.305
LightingStop	0.175	0.131	0.114	0.102	0.108	0.103	0.107	0.107	0.107	0.114
Weekday X Walk Time	-0.015	-0.013	-0.013	-0.013	-0.012	-0.013	-0.014	-0.014	-0.015	-0.015
Estimated Nest Co-efficients (μ)										
MTT	0.580	0.479	0.440	0.439	0.465	0.477	0.484	0.487	0.504	0.527
MAT	0.644	0.591	0.542	0.551	0.566	0.548	0.550	0.551	0.581	0.601
MTT-MAT	0.611	0.481	0.435	0.415	0.427	0.419	0.425	0.432	0.446	0.471
NoStrategy	0.600	0.467	0.416	0.400	0.401	0.395	0.402	0.409	0.420	0.453
t-statistics of the Estimated Parmeters										
TravelTime	-2.192	-2.115	-2.421	-1.878	-2.009	-2.018	-1.897	-1.837	-1.995	-1.840
Number of Transfer	-5.858	-5.705	-5.456	-5.468	-5.608	-6.371	-5.511	-5.680	-5.657	-6.121
Train	5.788	5.680	5.447	5.509	5.183	5.247	5.168	5.574	5.670	5.810
Bus	-2.388	-2.277	-2.080	-2.037	-1.926	-1.886	-1.811	-1.812	-1.768	-1.919
LightingStop	2.441	2.340	2.260	2.208	2.163	2.263	2.273	2.200	2.164	2.255
Weekday X Walk Time	-2.114	-2.238	-2.364	-2.393	-2.280	-2.403	-2.461	-2.414	-2.480	-2.455
t-statistics of the Nest Co-efficients										
MTT	6.972	6.282	5.772	5.279	4.574	4.587	4.329	4.303	4.366	4.734
MAT	5.984	4.779	4.358	4.597	4.216	4.215	3.879	4.065	4.198	4.316
MTT-MAT	3.628	3.035	2.729	2.949	2.779	2.919	3.062	2.918	2.944	2.920
NoStrategy	11.560	11.928	10.848	11.246	10.225	10.632	9.826	11.117	10.555	10.245
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	10	10	10	10	10	10	10	10	10	10
Initial Log Likelihood	-2254	-2547	-2696	-2722	-2722	-2709	-2690	-2676	-2651	-2597
Final Log Likelihood	-1214	-1383	-1480	-1530	-1551	-1580	-1593	-1605	-1613	-1609
Rho_Square	0.462	0.457	0.451	0.438	0.430	0.417	0.408	0.400	0.392	0.380
BIC	2495	2833	3027	3127	3169	3227	3253	3278	3293	3286

Model Name: TT-TR Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.120	-0.100	-0.097	-0.083	-0.081	-0.089	-0.101	-0.106	-0.108	-0.107
TravelTime	-0.068	-0.057	-0.053	-0.050	-0.045	-0.050	-0.049	-0.049	-0.050	-0.049
Train	1.611	1.277	1.231	1.213	1.253	1.310	1.350	1.361	1.384	1.382
Bus	-0.198	-0.208	-0.189	-0.194	-0.192	-0.211	-0.201	-0.199	-0.197	-0.193
Weekday X Walk Time	-0.017	-0.013	-0.012	-0.012	-0.011	-0.013	-0.013	-0.013	-0.013	-0.013
Estimated Nest Co-efficients (μ)										
MTT	0.589	0.452	0.445	0.422	0.432	0.508	0.519	0.527	0.527	0.518
MTR	0.531	0.416	0.392	0.377	0.381	0.409	0.419	0.429	0.432	0.432
MTT-MTR	0.602	0.495	0.481	0.461	0.455	0.482	0.495	0.500	0.503	0.498
NoStrategy	0.612	0.453	0.414	0.393	0.393	0.416	0.423	0.429	0.440	0.439
t-statistics of the Estimated Parmeters										
Access Time	-5.298	-5.084	-4.853	-4.785	-4.907	-4.973	-4.948	-5.072	-5.037	-4.970
TravelTime	-2.106	-2.266	-2.069	-1.893	-2.080	-1.726	-1.877	-1.705	-1.947	-1.579
Train	5.155	4.556	4.973	4.750	4.882	5.068	5.243	5.105	5.002	4.900
Bus	-2.419	-2.192	-2.173	-2.504	-2.550	-2.456	-2.240	-2.392	-2.300	-2.349
Weekday X Walk Time	-1.439	-1.325	-1.316	-1.335	-1.327	-1.430	-1.330	-1.335	-1.287	-1.305
t-statistics of the Nest Co-efficients										
MTT	3.585	2.890	2.567	2.344	2.162	1.575	1.779	1.909	2.042	1.874
MTR	6.782	5.232	5.071	4.638	4.300	4.413	4.421	3.562	3.695	3.967
MTT-MTR	2.878	2.145	2.084	1.631	1.770	2.009	2.099	2.055	2.096	2.319
NoStrategy	5.377	4.939	4.589	4.794	4.730	4.816	4.296	4.263	4.216	4.193
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	9	9	9	9	9	9	9	9	9	9
Initial Log Likelihood	-2387	-2708	-2776	-2791	-2782	-2711	-2686	-2662	-2644	-2638
Final Log Likelihood	-1312	-1483	-1541	-1584	-1607	-1614	-1623	-1633	-1635	-1644
Rho_Square	0.454	0.455	0.448	0.435	0.423	0.408	0.398	0.389	0.384	0.379
BIC	2684	3026	3142	3229	3275	3288	3307	3327	3331	3349

Model Name: AT-TR Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.156	-0.119	-0.112	-0.114	-0.109	-0.108	-0.116	-0.117	-0.120	-0.121
TravelTime	-0.042	-0.036	-0.035	-0.037	-0.036	-0.035	-0.037	-0.037	-0.038	-0.039
Train	1.854	1.426	1.295	1.347	1.289	1.304	1.399	1.412	1.447	1.480
Bus	-0.412	-0.358	-0.374	-0.396	-0.382	-0.381	-0.409	-0.410	-0.420	-0.418
LightingStop	0.198	0.145	0.129	0.124	0.124	0.123	0.130	0.132	0.136	0.137
Weekday X Walk Time	-0.018	-0.013	-0.013	-0.014	-0.014	-0.014	-0.015	-0.014	-0.015	-0.015
Estimated Nest Co-efficients (μ)										
MAT	0.671	0.611	0.600	0.643	0.638	0.636	0.652	0.641	0.646	0.645
MTR	0.607	0.470	0.452	0.470	0.457	0.457	0.493	0.505	0.521	0.527
MAT-MTR	0.497	0.414	0.393	0.395	0.372	0.383	0.401	0.405	0.407	0.413
NoStrategy	0.674	0.509	0.468	0.470	0.450	0.450	0.472	0.473	0.479	0.495
t-statistics of the Estimated Parmeters										
Access Time	-9.380	-9.054	-9.244	-8.996	-9.218	-9.022	-9.139	-9.313	-9.448	-9.157
TravelTime	-5.735	-5.846	-5.974	-5.888	-6.083	-6.018	-5.794	-6.177	-6.303	-6.116
Train	8.097	7.933	8.174	8.027	8.001	7.967	8.088	8.107	8.476	8.413
Bus	-2.949	-3.163	-3.598	-3.757	-3.655	-3.715	-3.814	-3.901	-3.863	-3.690
LightingStop	2.147	2.052	1.965	1.879	1.920	1.877	1.941	1.922	1.975	1.957
Weekday X Walk Time	-1.892	-1.761	-1.852	-1.851	-1.925	-2.019	-1.913	-1.888	-1.887	-1.874
t-statistics of the Nest Co-efficients										
MAT	8.925	7.693	7.411	6.907	6.883	6.918	6.110	6.653	6.437	6.586
MTR	15.594	14.292	13.970	13.056	12.555	12.300	11.917	11.831	11.723	11.457
MAT-MTR	5.478	5.392	5.465	5.478	5.497	5.416	5.438	5.383	5.582	5.659
NoStrategy	15.672	14.697	15.280	15.059	14.838	14.280	13.896	14.124	14.563	13.760
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	10	10	10	10	10	10	10	10	10	10
Initial Log Likelihood	-2191	-2508	-2603	-2618	-2657	-2645	-2621	-2612	-2613	-2589
Final Log Likelihood	-1204	-1389	-1465	-1505	-1552	-1579	-1592	-1602	-1609	-1614
Rho_Square	0.450	0.446	0.437	0.425	0.416	0.403	0.393	0.387	0.384	0.376
BIC	2476	2845	2998	3077	3171	3225	3252	3272	3286	3296

Model Name: NL Model - Mode

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.083	-0.072	-0.068	-0.066	-0.067	-0.068	-0.069	-0.072	-0.073	-0.075
TravelTime	-0.022	-0.020	-0.019	-0.018	-0.019	-0.019	-0.020	-0.020	-0.021	-0.021
Number of Transfer	-0.534	-0.450	-0.416	-0.405	-0.412	-0.421	-0.427	-0.443	-0.450	-0.463
LightingStop	0.130	0.111	0.090	0.088	0.088	0.089	0.088	0.090	0.089	0.092
Travel Time Strategy	0.224	0.161	0.150	0.142	0.143	0.144	0.146	0.151	0.152	0.159
Weekday X Walk Time	-0.007	-0.006	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
Estimated Nest Co-efficients (μ)										
Only Bus	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Only Train	0.413	0.336	0.309	0.296	0.300	0.305	0.308	0.319	0.323	0.333
Mixed	0.395	0.308	0.287	0.278	0.288	0.294	0.300	0.307	0.309	0.315
t-statistics of the Estimated Parmeters										
Access Time	-2.575	-1.577	-1.441	-1.446	-1.566	-1.404	-1.472	-1.374	-1.482	-1.620
TravelTime	-1.670	-1.189	-0.957	-0.965	-1.109	-0.952	-0.995	-0.987	-1.073	-1.049
Number of Transfer	-2.414	-1.582	-1.364	-1.326	-1.358	-1.303	-1.281	-1.352	-1.301	-1.505
LightingStop	1.147	0.836	0.599	0.612	0.599	0.604	0.565	0.614	0.581	0.614
Travel Time Strategy	1.111	0.700	0.607	0.592	0.583	0.562	0.524	0.512	0.521	0.680
Weekday X Walk Time	-0.709	-0.423	-0.379	-0.365	-0.413	-0.323	-0.370	-0.377	-0.369	-0.388
t-statistics of the Nest Co-efficients										
Only Bus	0.981	0.438	0.359	0.344	0.336	0.332	0.310	0.312	0.321	0.366
Only Train	2.767	1.611	1.444	1.580	1.621	1.463	1.501	1.453	1.650	1.663
Mixed	3.309	1.461	1.310	1.225	1.172	1.156	1.154	1.185	1.134	1.296
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	9	9	9	9	9	9	9	9	9	9
Initial Log Likelihood	-1714	-2152	-2341	-2453	-2491	-2512	-2530	-2529	-2535	-2536
Final Log Likelihood	-975	-1253	-1392	-1474	-1527	-1562	-1585	-1613	-1629	-1641
Rho_Square	0.431	0.417	0.405	0.399	0.387	0.378	0.373	0.362	0.357	0.353
BIC	2010	2568	2845	3009	3114	3184	3232	3287	3320	3343

Model Name: RRM Model With Sampling (Guevara Sampling Protocol)

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.013	-0.012	-0.012	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
TravelTime	-0.005	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
Number of Transfer	-0.081	-0.076	-0.075	-0.073	-0.072	-0.071	-0.071	-0.070	-0.070	-0.070
Train	0.151	0.133	0.133	0.118	0.120	0.116	0.118	0.111	0.107	0.112
Bus	-0.022	-0.024	-0.024	-0.026	-0.023	-0.025	-0.022	-0.024	-0.026	-0.023
LightingStop	0.012	0.011	0.011	0.007	0.008	0.007	0.007	0.007	0.006	0.005
Travel Time Strategy	0.014	0.020	0.017	0.017	0.021	0.019	0.022	0.020	0.021	0.021
t-statistics of the Estimated Parmeters										
Access Time	-16.904	-13.024	-11.296	-11.763	-14.058	-15.020	-12.131	-12.620	-13.578	-14.254
TravelTime	-10.466	-15.730	-15.869	-8.776	-5.937	-14.979	-5.544	-5.335	-5.761	-7.214
Number of Transfer	-72.334	-11.728	-10.869	-11.192	-13.568	-15.885	-11.082	-12.932	-12.840	-13.273
Train	6.058	7.583	6.779	7.609	9.436	8.938	7.951	7.656	7.992	8.481
Bus	-1.504	-1.998	-1.715	-2.463	-2.409	-2.546	-2.081	-2.260	-2.526	-2.656
LightingStop	4.987	1.367	1.671	1.254	1.340	1.332	3.571	1.082	1.011	1.046
Travel Time Strategy	3.213	1.797	1.429	1.664	2.257	2.426	1.980	1.974	2.337	2.177
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1322	-1777	-1992	-2117	-2198	-2254	-2294	-2325	-2348	-2366
Final Log Likelihood	-916	-1290	-1472	-1579	-1640	-1693	-1727	-1754	-1772	-1788
Rho_Square	0.307	0.274	0.261	0.254	0.254	0.249	0.247	0.245	0.245	0.244
BIC	1879	2627	2991	3205	3328	3432	3502	3556	3592	3624

Model Name: RRM Model With Sampling (Proposed Sampling Protocol)

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.013	-0.012	-0.013	-0.012	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
TravelTime	-0.005	-0.005	-0.005	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
Number of Transfer	-0.083	-0.080	-0.077	-0.074	-0.073	-0.072	-0.072	-0.071	-0.071	-0.070
Train	0.151	0.133	0.136	0.126	0.122	0.116	0.114	0.112	0.115	0.110
Bus	-0.025	-0.025	-0.024	-0.023	-0.024	-0.029	-0.025	-0.027	-0.023	-0.026
LightingStop	0.008	0.013	0.008	0.008	0.007	0.009	0.008	0.006	0.008	0.007
Travel Time Strategy	0.018	0.021	0.022	0.017	0.023	0.020	0.020	0.023	0.025	0.024
t-statistics of the Estimated Parmeters										
Access Time	-12.995	-16.127	-12.901	-17.084	-12.011	-17.376	-22.255	-19.288	-15.729	-16.885
TravelTime	-10.508	-16.801	-30.032	-38.621	-12.130	-14.635	-42.476	-13.089	-28.227	-44.937
Number of Transfer	-12.437	-70.112	-16.851	-11.033	-11.758	-8.913	-13.577	-10.691	-14.389	-12.209
Train	9.043	7.516	8.811	17.882	6.383	6.764	17.388	13.018	9.173	8.880
Bus	-2.167	-1.931	-4.174	-2.359	-1.732	-1.921	-2.820	-2.130	-2.396	-2.522
LightingStop	1.236	1.662	1.785	1.350	2.301	2.694	1.734	1.197	2.422	1.325
Travel Time Strategy	1.160	2.160	9.466	3.153	3.756	17.515	2.544	2.271	12.785	13.376
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1262	-1694	-1907	-2039	-2118	-2181	-2223	-2265	-2290	-2315
Final Log Likelihood	-837	-1184	-1378	-1489	-1558	-1615	-1652	-1689	-1712	-1735
Rho_Square	0.337	0.301	0.277	0.270	0.264	0.260	0.257	0.254	0.252	0.251
BIC	1722	2415	2804	3025	3164	3277	3351	3426	3472	3517

Model Name: RRM-RUM Hybrid Model Guevara Sampling Protocol)

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.012	-0.011	-0.011	-0.011	-0.010	-0.010	-0.010	-0.010	-0.010	-0.010
TravelTime	-0.003	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
Number of Transfer	-0.075	-0.070	-0.068	-0.067	-0.065	-0.065	-0.063	-0.064	-0.063	-0.063
Train	0.142	0.123	0.119	0.117	0.115	0.109	0.107	0.105	0.104	0.103
Bus	-0.423	-0.449	-0.451	-0.461	-0.445	-0.466	-0.489	-0.489	-0.482	-0.496
LightingStop	0.281	0.305	0.289	0.291	0.264	0.276	0.271	0.260	0.259	0.252
Travel Time Strategy	0.934	0.962	0.952	0.981	0.985	0.993	0.995	1.005	1.004	1.010
t-statistics of the Estimated Parmeters										
Access Time	-10.387	-12.033	-14.022	-11.686	-13.277	-13.680	-12.379	-13.852	-13.335	-12.622
TravelTime	-4.067	-3.742	-13.114	-4.258	-8.257	-5.770	-6.673	-7.688	-10.637	-7.969
Number of Transfer	-10.122	-11.422	-16.456	-12.483	-11.639	-17.195	-13.318	-13.068	-12.488	-17.269
Train	6.716	8.015	8.523	9.324	8.773	9.714	10.719	10.759	9.064	9.150
Bus	-2.247	-2.448	-2.601	-2.868	-4.019	-3.162	-2.986	-3.236	-2.944	-3.685
LightingStop	1.899	2.194	2.240	2.467	2.488	2.560	2.385	2.471	2.341	2.295
Travel Time Strategy	6.936	7.993	8.334	7.543	8.077	9.254	7.837	9.223	9.232	11.269
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1328	-1786	-2003	-2129	-2210	-2267	-2308	-2340	-2363	-2381
Final Log Likelihood	-890	-1261	-1436	-1539	-1612	-1659	-1697	-1723	-1741	-1756
Rho_Square	0.330	0.294	0.283	0.277	0.271	0.268	0.265	0.264	0.263	0.263
BIC	1827	2569	2920	3126	3270	3365	3442	3493	3530	3560

Model Name: RRM-RUM Hybrid Model Proposed Sampling Protocol)

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.012	-0.011	-0.011	-0.011	-0.011	-0.011	-0.010	-0.010	-0.010	-0.011
TravelTime	-0.003	-0.002	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
Number of Transfer	-0.075	-0.071	-0.068	-0.068	-0.067	-0.067	-0.064	-0.065	-0.065	-0.068
Train	0.152	0.129	0.121	0.119	0.119	0.114	0.110	0.111	0.111	0.116
Bus	-0.443	-0.470	-0.490	-0.468	-0.463	-0.484	-0.481	-0.505	-0.498	-0.528
LightingStop	0.315	0.301	0.299	0.310	0.278	0.281	0.264	0.265	0.267	0.272
Travel Time Strategy	0.944	0.993	1.005	0.994	0.990	0.986	1.002	0.997	1.001	1.036
t-statistics of the Estimated Parmeters										
Access Time	-10.783	-12.323	-11.879	-13.327	-14.216	-14.521	-14.997	-11.412	-13.815	-12.638
TravelTime	-3.797	-5.182	-4.082	-7.851	-11.764	-7.375	-10.192	-6.380	-9.931	-9.077
Number of Transfer	-9.998	-12.220	-11.388	-12.906	-12.506	-16.639	-16.789	-11.179	-12.285	-23.408
Train	6.799	8.264	8.464	8.529	9.180	10.012	9.189	9.493	10.528	11.685
Bus	-2.233	-2.730	-2.818	-10.016	-2.533	-3.144	-5.605	-3.371	-4.880	-12.787
LightingStop	2.299	2.329	2.427	2.188	2.369	2.419	2.159	2.497	2.147	2.129
Travel Time Strategy	7.656	8.078	7.973	7.788	9.333	9.472	9.149	8.949	9.002	14.735
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	7	7	7	7	7	7	7	7	7	7
Initial Log Likelihood	-1272	-1704	-1918	-2045	-2133	-2195	-2237	-2276	-2306	-2331
Final Log Likelihood	-812	-1155	-1336	-1439	-1517	-1574	-1613	-1644	-1670	-1701
Rho_Square	0.362	0.323	0.303	0.296	0.289	0.283	0.279	0.278	0.276	0.270
BIC	1670	2357	2720	2926	3082	3196	3274	3336	3388	3450

Model Name: PSC-RRM Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.012	-0.012	-0.011	-0.011	-0.011	-0.011	-0.011	-0.010	-0.010	-0.010
TravelTime	-0.003	-0.003	-0.003	-0.002	-0.002	-0.003	-0.003	-0.002	-0.002	-0.002
Number of Transfer	-0.072	-0.069	-0.067	-0.065	-0.065	-0.064	-0.064	-0.062	-0.063	-0.063
Train	0.142	0.129	0.124	0.118	0.115	0.114	0.113	0.112	0.112	0.108
Bus	-0.504	-0.520	-0.547	-0.526	-0.521	-0.524	-0.539	-0.534	-0.524	-0.542
LightingStop	0.309	0.316	0.309	0.285	0.295	0.277	0.287	0.269	0.267	0.258
Travel Time Strategy	0.947	1.002	1.008	1.010	1.008	0.998	0.999	1.012	1.018	1.016
PathSizeCorrection	-0.026	-0.024	-0.021	-0.020	-0.017	-0.020	-0.019	-0.018	-0.017	-0.017
t-statistics of the Estimated Parmeters										
Access Time	-11.257	-12.357	-12.315	-13.043	-12.380	-15.107	-14.029	-12.817	-13.200	-12.372
TravelTime	-4.145	-4.149	-4.298	-5.253	-4.809	-7.574	-5.390	-6.208	-4.764	-5.715
Number of Transfer	-9.949	-11.255	-10.869	-12.290	-11.175	-12.089	-12.371	-11.436	-11.755	-10.486
Train	7.047	8.178	8.727	8.839	8.763	10.258	9.731	9.843	9.482	8.266
Bus	-2.465	-2.859	-3.038	-3.213	-3.205	-3.790	-2.912	-2.822	-3.165	-3.279
LightingStop	2.090	2.436	2.330	2.611	2.351	2.360	2.468	2.162	2.296	2.156
Travel Time Strategy	7.471	8.685	8.029	9.755	7.659	8.439	8.417	12.849	9.336	9.610
PathSizeCorrection	-1.592	-1.858	-1.557	-1.617	-1.285	-1.719	-1.562	-1.641	-1.375	-1.395
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	8	8	8	8	8	8	8	8	8	8
Initial Log Likelihood	-1271	-1704	-1915	-2044	-2127	-2189	-2237	-2276	-2299	-2326
Final Log Likelihood	-812	-1152	-1330	-1448	-1516	-1571	-1613	-1650	-1668	-1691
Rho_Square	0.362	0.324	0.305	0.292	0.287	0.282	0.279	0.275	0.274	0.273
BIC	1677	2357	2715	2949	3087	3196	3279	3354	3391	3436

Model Name: PSC-RRM-RUM Model

	Choice Set Size									
	5	10	15	20	25	30	35	40	45	50
Estimated Parmeters (β)										
Access Time	-0.012	-0.012	-0.011	-0.011	-0.011	-0.011	-0.011	-0.010	-0.011	-0.010
TravelTime	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
Number of Transfer	-0.072	-0.068	-0.065	-0.066	-0.064	-0.063	-0.063	-0.063	-0.063	-0.062
Train	0.147	0.125	0.125	0.118	0.116	0.112	0.111	0.113	0.108	0.109
Bus	-0.493	-0.569	-0.516	-0.565	-0.565	-0.580	-0.598	-0.583	-0.594	-0.606
LightingStop	0.316	0.299	0.297	0.300	0.292	0.287	0.280	0.276	0.272	0.265
Travel Time Strategy	0.981	1.016	1.030	1.018	1.029	1.044	1.028	1.021	1.027	1.036
PathSizeCorrection	-0.374	-0.498	-0.471	-0.513	-0.532	-0.545	-0.560	-0.552	-0.545	-0.560
t-statistics of the Estimated Parmeters										
Access Time	-11.126	-12.263	-12.710	-11.570	-12.694	-12.565	-12.360	-11.021	-16.016	-13.424
TravelTime	-3.950	-4.300	-4.453	-6.493	-4.217	-5.112	-8.775	-6.689	-20.893	-9.892
Number of Transfer	-9.780	-11.061	-11.495	-13.415	-11.644	-13.030	-12.165	-44.813	-42.960	-20.349
Train	6.679	7.491	8.660	8.594	8.832	9.363	10.658	8.089	13.125	10.837
Bus	-2.376	-3.110	-2.963	-2.715	-2.820	-6.443	-3.859	-3.422	-6.074	-9.706
LightingStop	2.334	2.371	2.739	2.164	2.427	2.492	3.262	2.034	2.398	2.517
Travel Time Strategy	7.639	7.579	8.289	8.496	7.571	8.314	8.424	6.717	8.470	16.734
PathSizeCorrection	-2.019	-2.281	-2.260	-2.490	-2.394	-3.002	-2.403	-2.193	-2.596	-4.242
Model Statistics										
Number of Observation	867	867	867	867	867	867	867	867	867	867
Number of Parameter	8	8	8	8	8	8	8	8	8	8
Initial Log Likelihood	-1275	-1704	-1920	-2043	-2130	-2197	-2239	-2279	-2303	-2330
Final Log Likelihood	-810	-1153	-1337	-1434	-1514	-1577	-1608	-1643	-1666	-1688
Rho_Square	0.365	0.323	0.304	0.298	0.289	0.282	0.281	0.279	0.277	0.275
BIC	1674	2360	2728	2923	3081	3207	3271	3341	3386	3431

APPENDIX E

SAMPLE CODES

APPENDIX E-1 MNL MODEL FOR STOP CHOICE (BIOGEME CODE)

```
[ModelDescription]
MNL Model for Stop Choice

[Choice]
CHOICE

[Beta]

// Name Value  LowerBound UpperBound  status (0=variable, 1=fixed)

B_AccessTime        0   -10   10   0
B_stopLight     0   -10   10   0
B_NumofRoutes       0   -10   10   0
B_MinWalkTime       0   -10   10   0
B_Train             0   -10   10   0
B_Str_Tr        0   -10   10   1
B_Str_TT        0   -10   10   0

[Utilities]

// Id Name    Avail     linear-in-parameter expression (beta1*x1 + beta2*x2 + ... )
1 Choice1 AV1 B_AccessTime * AccTime1 + B_stopLight * stopLight1 + B_NumofRoutes * NumofRoutes1 + B_Str_Tr * MinTr1 +B_Str_TT * MinTT1 + B_MinWalkTime * MiniOtherWalkTime1
2 Choice2 AV2 B_AccessTime * AccTime2 + B_stopLight * stopLight2 + B_NumofRoutes * NumofRoutes2 + B_Str_Tr * MinTr2 +B_Str_TT * MinTT2 + B_MinWalkTime * MiniOtherWalkTime2
3 Choice3 AV3 B_AccessTime * AccTime3 + B_stopLight * stopLight3 + B_NumofRoutes * NumofRoutes3 + B_Str_Tr * MinTr3 +B_Str_TT * MinTT3 + B_MinWalkTime * MiniOtherWalkTime3
4 Choice4 AV4 B_AccessTime * AccTime4 + B_stopLight * stopLight4 + B_NumofRoutes * NumofRoutes4 + B_Str_Tr * MinTr4 +B_Str_TT * MinTT4 + B_MinWalkTime * MiniOtherWalkTime4
5 Choice5 AV5 B_AccessTime * AccTime5 + B_stopLight * stopLight5 + B_NumofRoutes * NumofRoutes5 + B_Str_Tr * MinTr5 +B_Str_TT * MinTT5 + B_MinWalkTime * MiniOtherWalkTime5
6 Choice6 AV6 B_AccessTime * AccTime6 + B_stopLight * stopLight6 + B_NumofRoutes * NumofRoutes6 + B_Str_Tr * MinTr6 +B_Str_TT * MinTT6 + B_MinWalkTime * MiniOtherWalkTime6
7 Choice7 AV7 B_AccessTime * AccTime7 + B_stopLight * stopLight7 + B_NumofRoutes * NumofRoutes7 + B_Str_Tr * MinTr7 +B_Str_TT * MinTT7 + B_MinWalkTime * MiniOtherWalkTime7
8 Choice8 AV8 B_AccessTime * AccTime8 + B_stopLight * stopLight8 + B_NumofRoutes * NumofRoutes8 + B_Str_Tr * MinTr8 +B_Str_TT * MinTT8 + B_MinWalkTime * MiniOtherWalkTime8
```

9 Choice9 AV9 B_AccessTime * AccTime9 + B_stopLight * stopLight9 + B_NumofRoutes * NumofRoutes9 + B_Str_Tr * MinTr9 +B_Str_TT * MinTT9 + B_MinWalkTime * MiniOtherWalkTime9
10 Choice10 AV10 B_AccessTime * AccTime10 + B_stopLight * stopLight10 + B_NumofRoutes * NumofRoutes10 + B_Str_Tr * MinTr10 +B_Str_TT * MinTT10 + B_Train * Train10 + B_MinWalkTime * MiniOtherWalkTime10
11 Choice11 AV11 B_AccessTime * AccTime11 + B_stopLight * stopLight11 + B_NumofRoutes * NumofRoutes11 + B_Str_Tr * MinTr11 +B_Str_TT * MinTT11 + B_Train * Train11 + B_MinWalkTime * MiniOtherWalkTime11
12 Choice12 AV12 B_AccessTime * AccTime12 + B_stopLight * stopLight12 + B_NumofRoutes * NumofRoutes12 + B_Str_Tr * MinTr12 +B_Str_TT * MinTT12 + B_Train * Train12 + B_MinWalkTime * MiniOtherWalkTime12
13 Choice13 AV13 B_AccessTime * AccTime13 + B_stopLight * stopLight13 + B_NumofRoutes * NumofRoutes13 + B_Str_Tr * MinTr13 +B_Str_TT * MinTT13 + B_Train * Train13 + B_MinWalkTime * MiniOtherWalkTime13
14 Choice14 AV14 B_AccessTime * AccTime14 + B_stopLight * stopLight14 + B_NumofRoutes * NumofRoutes14 + B_Str_Tr * MinTr14 +B_Str_TT * MinTT14 + B_Train * Train14 + B_MinWalkTime * MiniOtherWalkTime14
15 Choice15 AV15 B_AccessTime * AccTime15 + B_stopLight * stopLight15 + B_NumofRoutes * NumofRoutes15 + B_Str_Tr * MinTr15 +B_Str_TT * MinTT15 + B_Train * Train15 + B_MinWalkTime * MiniOtherWalkTime15
16 Choice16 AV16 B_AccessTime * AccTime16 + B_stopLight * stopLight16 + B_NumofRoutes * NumofRoutes16 + B_Str_Tr * MinTr16 +B_Str_TT * MinTT16 + B_Train * Train16 + B_MinWalkTime * MiniOtherWalkTime16
17 Choice17 AV17 B_AccessTime * AccTime17 + B_stopLight * stopLight17 + B_NumofRoutes * NumofRoutes17 + B_Str_Tr * MinTr17 +B_Str_TT * MinTT17 + B_Train * Train17 + B_MinWalkTime * MiniOtherWalkTime17
18 Choice18 AV18 B_AccessTime * AccTime18 + B_stopLight * stopLight18 + B_NumofRoutes * NumofRoutes18 + B_Str_Tr * MinTr18 +B_Str_TT * MinTT18 + B_Train * Train18 + B_MinWalkTime * MiniOtherWalkTime18
19 Choice19 AV19 B_AccessTime * AccTime19 + B_stopLight * stopLight19 + B_NumofRoutes * NumofRoutes19 + B_Str_Tr * MinTr19 +B_Str_TT * MinTT19 + B_Train * Train19 + B_MinWalkTime * MiniOtherWalkTime19
20 Choice20 AV20 B_AccessTime * AccTime20 + B_stopLight * stopLight20 + B_NumofRoutes * NumofRoutes20 + B_Str_Tr * MinTr20 +B_Str_TT * MinTT20 + B_Train * Train20 + B_MinWalkTime * MiniOtherWalkTime20
21 Choice21 AV21 B_AccessTime * AccTime21 + B_stopLight * stopLight21 + B_NumofRoutes * NumofRoutes21 + B_Str_Tr * MinTr21 +B_Str_TT * MinTT21 + B_Train * Train21 + B_MinWalkTime * MiniOtherWalkTime21
22 Choice22 AV22 B_AccessTime * AccTime22 + B_stopLight * stopLight22 + B_NumofRoutes * NumofRoutes22 + B_Str_Tr * MinTr22 +B_Str_TT * MinTT22 + B_Train * Train22 + B_MinWalkTime * MiniOtherWalkTime22
23 Choice23 AV23 B_AccessTime * AccTime23 + B_stopLight * stopLight23 + B_NumofRoutes * NumofRoutes23 + B_Str_Tr * MinTr23 +B_Str_TT * MinTT23 + B_Train * Train23 + B_MinWalkTime * MiniOtherWalkTime23
24 Choice24 AV24 B_AccessTime * AccTime24 + B_stopLight * stopLight24 + B_NumofRoutes * NumofRoutes24 + B_Str_Tr * MinTr24 +B_Str_TT * MinTT24 + B_Train * Train24 + B_MinWalkTime * MiniOtherWalkTime24

25 Choice25 AV25 B_AccessTime * AccTime25 + B_stopLight * stopLight25 + B_NumofRoutes * NumofRoutes25 + B_Str_Tr * MinTr25 +B_Str_TT * MinTT25 + B_Train * Train25 + B_MinWalkTime * MiniOtherWalkTime25
26 Choice26 AV26 B_AccessTime * AccTime26 + B_stopLight * stopLight26 + B_NumofRoutes * NumofRoutes26 + B_Str_Tr * MinTr26 +B_Str_TT * MinTT26 + B_Train * Train26 + B_MinWalkTime * MiniOtherWalkTime26
27 Choice27 AV27 B_AccessTime * AccTime27 + B_stopLight * stopLight27 + B_NumofRoutes * NumofRoutes27 + B_Str_Tr * MinTr27 +B_Str_TT * MinTT27 + B_Train * Train27 + B_MinWalkTime * MiniOtherWalkTime27
28 Choice28 AV28 B_AccessTime * AccTime28 + B_stopLight * stopLight28 + B_NumofRoutes * NumofRoutes28 + B_Str_Tr * MinTr28 +B_Str_TT * MinTT28 + B_Train * Train28 + B_MinWalkTime * MiniOtherWalkTime28
29 Choice29 AV29 B_AccessTime * AccTime29 + B_stopLight * stopLight29 + B_NumofRoutes * NumofRoutes29 + B_Str_Tr * MinTr29 +B_Str_TT * MinTT29 + B_Train * Train29 + B_MinWalkTime * MiniOtherWalkTime29
30 Choice30 AV30 B_AccessTime * AccTime30 + B_stopLight * stopLight30 + B_NumofRoutes * NumofRoutes30 + B_Str_Tr * MinTr30 +B_Str_TT * MinTT30 + B_Train * Train30 + B_MinWalkTime * MiniOtherWalkTime30
31 Choice31 AV31 B_AccessTime * AccTime31 + B_stopLight * stopLight31 + B_NumofRoutes * NumofRoutes31 + B_Str_Tr * MinTr31 +B_Str_TT * MinTT31 + B_Train * Train31 + B_MinWalkTime * MiniOtherWalkTime31
32 Choice32 AV32 B_AccessTime * AccTime32 + B_stopLight * stopLight32 + B_NumofRoutes * NumofRoutes32 + B_Str_Tr * MinTr32 +B_Str_TT * MinTT32 + B_Train * Train32 + B_MinWalkTime * MiniOtherWalkTime32
33 Choice33 AV33 B_AccessTime * AccTime33 + B_stopLight * stopLight33 + B_NumofRoutes * NumofRoutes33 + B_Str_Tr * MinTr33 +B_Str_TT * MinTT33 + B_Train * Train33 + B_MinWalkTime * MiniOtherWalkTime33
34 Choice34 AV34 B_AccessTime * AccTime34 + B_stopLight * stopLight34 + B_NumofRoutes * NumofRoutes34 + B_Str_Tr * MinTr34 +B_Str_TT * MinTT34 + B_Train * Train34 + B_MinWalkTime * MiniOtherWalkTime34
35 Choice35 AV35 B_AccessTime * AccTime35 + B_stopLight * stopLight35 + B_NumofRoutes * NumofRoutes35 + B_Str_Tr * MinTr35 +B_Str_TT * MinTT35 + B_Train * Train35 + B_MinWalkTime * MiniOtherWalkTime35
36 Choice36 AV36 B_AccessTime * AccTime36 + B_stopLight * stopLight36 + B_NumofRoutes * NumofRoutes36 + B_Str_Tr * MinTr36 +B_Str_TT * MinTT36 + B_Train * Train36 + B_MinWalkTime * MiniOtherWalkTime36
37 Choice37 AV37 B_AccessTime * AccTime37 + B_stopLight * stopLight37 + B_NumofRoutes * NumofRoutes37 + B_Str_Tr * MinTr37 +B_Str_TT * MinTT37 + B_Train * Train37 + B_MinWalkTime * MiniOtherWalkTime37
38 Choice38 AV38 B_AccessTime * AccTime38 + B_stopLight * stopLight38 + B_NumofRoutes * NumofRoutes38 + B_Str_Tr * MinTr38 +B_Str_TT * MinTT38 + B_Train * Train38 + B_MinWalkTime * MiniOtherWalkTime38
39 Choice39 AV39 B_AccessTime * AccTime39 + B_stopLight * stopLight39 + B_NumofRoutes * NumofRoutes39 + B_Str_Tr * MinTr39 +B_Str_TT * MinTT39 + B_Train * Train39 + B_MinWalkTime * MiniOtherWalkTime39
40 Choice40 AV40 B_AccessTime * AccTime40 + B_stopLight * stopLight40 + B_NumofRoutes * NumofRoutes40 + B_Str_Tr * MinTr40 +B_Str_TT * MinTT40 + B_Train * Train40 + B_MinWalkTime * MiniOtherWalkTime40

41 Choice41 AV41 B_AccessTime * AccTime41 + B_stopLight * stopLight41 + B_NumofRoutes * NumofRoutes41 + B_Str_Tr * MinTr41 +B_Str_TT * MinTT41 + B_Train * Train41 + B_MinWalkTime * MiniOtherWalkTime41
42 Choice42 AV42 B_AccessTime * AccTime42 + B_stopLight * stopLight42 + B_NumofRoutes * NumofRoutes42 + B_Str_Tr * MinTr42 +B_Str_TT * MinTT42 + B_Train * Train42 + B_MinWalkTime * MiniOtherWalkTime42
43 Choice43 AV43 B_AccessTime * AccTime43 + B_stopLight * stopLight43 + B_NumofRoutes * NumofRoutes43 + B_Str_Tr * MinTr43 +B_Str_TT * MinTT43 + B_Train * Train43 + B_MinWalkTime * MiniOtherWalkTime43
44 Choice44 AV44 B_AccessTime * AccTime44 + B_stopLight * stopLight44 + B_NumofRoutes * NumofRoutes44 + B_Str_Tr * MinTr44 +B_Str_TT * MinTT44 + B_Train * Train44 + B_MinWalkTime * MiniOtherWalkTime44
45 Choice45 AV45 B_AccessTime * AccTime45 + B_stopLight * stopLight45 + B_NumofRoutes * NumofRoutes45 + B_Str_Tr * MinTr45 +B_Str_TT * MinTT45 + B_Train * Train45 + B_MinWalkTime * MiniOtherWalkTime45
46 Choice46 AV46 B_AccessTime * AccTime46 + B_stopLight * stopLight46 + B_NumofRoutes * NumofRoutes46 + B_Str_Tr * MinTr46 +B_Str_TT * MinTT46 + B_Train * Train46 + B_MinWalkTime * MiniOtherWalkTime46
47 Choice47 AV47 B_AccessTime * AccTime47 + B_stopLight * stopLight47 + B_NumofRoutes * NumofRoutes47 + B_Str_Tr * MinTr47 +B_Str_TT * MinTT47 + B_Train * Train47 + B_MinWalkTime * MiniOtherWalkTime47
48 Choice48 AV48 B_AccessTime * AccTime48 + B_stopLight * stopLight48 + B_NumofRoutes * NumofRoutes48 + B_Str_Tr * MinTr48 +B_Str_TT * MinTT48 + B_Train * Train48 + B_MinWalkTime * MiniOtherWalkTime48
49 Choice49 AV49 B_AccessTime * AccTime49 + B_stopLight * stopLight49 + B_NumofRoutes * NumofRoutes49 + B_Str_Tr * MinTr49 +B_Str_TT * MinTT49 + B_Train * Train49 + B_MinWalkTime * MiniOtherWalkTime49
50 Choice50 AV50 B_AccessTime * AccTime50 + B_stopLight * stopLight50 + B_NumofRoutes * NumofRoutes50 + B_Str_Tr * MinTr50 +B_Str_TT * MinTT50 + B_Train * Train50 + B_MinWalkTime * MiniOtherWalkTime50
51 Choice51 AV51 B_AccessTime * AccTime51 + B_stopLight * stopLight51 + B_NumofRoutes * NumofRoutes51 + B_Str_Tr * MinTr51 +B_Str_TT * MinTT51 + B_Train * Train51 + B_MinWalkTime * MiniOtherWalkTime51
52 Choice52 AV52 B_AccessTime * AccTime52 + B_stopLight * stopLight52 + B_NumofRoutes * NumofRoutes52 + B_Str_Tr * MinTr52 +B_Str_TT * MinTT52 + B_Train * Train52 + B_MinWalkTime * MiniOtherWalkTime52
53 Choice53 AV53 B_AccessTime * AccTime53 + B_stopLight * stopLight53 + B_NumofRoutes * NumofRoutes53 + B_Str_Tr * MinTr53 +B_Str_TT * MinTT53 + B_Train * Train53 + B_MinWalkTime * MiniOtherWalkTime53
54 Choice54 AV54 B_AccessTime * AccTime54 + B_stopLight * stopLight54 + B_NumofRoutes * NumofRoutes54 + B_Str_Tr * MinTr54 +B_Str_TT * MinTT54 + B_Train * Train54 + B_MinWalkTime * MiniOtherWalkTime54
55 Choice55 AV55 B_AccessTime * AccTime55 + B_stopLight * stopLight55 + B_NumofRoutes * NumofRoutes55 + B_Str_Tr * MinTr55 +B_Str_TT * MinTT55 + B_Train * Train55 + B_MinWalkTime * MiniOtherWalkTime55
56 Choice56 AV56 B_AccessTime * AccTime56 + B_stopLight * stopLight56 + B_NumofRoutes * NumofRoutes56 + B_Str_Tr * MinTr56 +B_Str_TT * MinTT56 + B_Train * Train56 + B_MinWalkTime * MiniOtherWalkTime56

57 Choice57 AV57 B_AccessTime * AccTime57 + B_stopLight * stopLight57 + B_NumofRoutes * NumofRoutes57 + B_Str_Tr * MinTr57 +B_Str_TT * MinTT57 + B_Train * Train57 + B_MinWalkTime * MiniOtherWalkTime57
58 Choice58 AV58 B_AccessTime * AccTime58 + B_stopLight * stopLight58 + B_NumofRoutes * NumofRoutes58 + B_Str_Tr * MinTr58 +B_Str_TT * MinTT58 + B_Train * Train58 + B_MinWalkTime * MiniOtherWalkTime58
59 Choice59 AV59 B_AccessTime * AccTime59 + B_stopLight * stopLight59 + B_NumofRoutes * NumofRoutes59 + B_Str_Tr * MinTr59 +B_Str_TT * MinTT59 + B_Train * Train59 + B_MinWalkTime * MiniOtherWalkTime59
60 Choice60 AV60 B_AccessTime * AccTime60 + B_stopLight * stopLight60 + B_NumofRoutes * NumofRoutes60 + B_Str_Tr * MinTr60 +B_Str_TT * MinTT60 + B_Train * Train60 + B_MinWalkTime * MiniOtherWalkTime60
61 Choice61 AV61 B_AccessTime * AccTime61 + B_stopLight * stopLight61 + B_NumofRoutes * NumofRoutes61 + B_Str_Tr * MinTr61 +B_Str_TT * MinTT61 + B_Train * Train61 + B_MinWalkTime * MiniOtherWalkTime61
62 Choice62 AV62 B_AccessTime * AccTime62 + B_stopLight * stopLight62 + B_NumofRoutes * NumofRoutes62 + B_Str_Tr * MinTr62 +B_Str_TT * MinTT62 + B_Train * Train62 + B_MinWalkTime * MiniOtherWalkTime62
63 Choice63 AV63 B_AccessTime * AccTime63 + B_stopLight * stopLight63 + B_NumofRoutes * NumofRoutes63 + B_Str_Tr * MinTr63 +B_Str_TT * MinTT63 + B_Train * Train63 + B_MinWalkTime * MiniOtherWalkTime63
64 Choice64 AV64 B_AccessTime * AccTime64 + B_stopLight * stopLight64 + B_NumofRoutes * NumofRoutes64 + B_Str_Tr * MinTr64 +B_Str_TT * MinTT64 + B_Train * Train64 + B_MinWalkTime * MiniOtherWalkTime64
65 Choice65 AV65 B_AccessTime * AccTime65 + B_stopLight * stopLight65 + B_NumofRoutes * NumofRoutes65 + B_Str_Tr * MinTr65 +B_Str_TT * MinTT65 + B_Train * Train65 + B_MinWalkTime * MiniOtherWalkTime65
66 Choice66 AV66 B_AccessTime * AccTime66 + B_stopLight * stopLight66 + B_NumofRoutes * NumofRoutes66 + B_Str_Tr * MinTr66 +B_Str_TT * MinTT66 + B_Train * Train66 + B_MinWalkTime * MiniOtherWalkTime66
67 Choice67 AV67 B_AccessTime * AccTime67 + B_stopLight * stopLight67 + B_NumofRoutes * NumofRoutes67 + B_Str_Tr * MinTr67 +B_Str_TT * MinTT67 + B_Train * Train67 + B_MinWalkTime * MiniOtherWalkTime67
68 Choice68 AV68 B_AccessTime * AccTime68 + B_stopLight * stopLight68 + B_NumofRoutes * NumofRoutes68 + B_Str_Tr * MinTr68 +B_Str_TT * MinTT68 + B_Train * Train68 + B_MinWalkTime * MiniOtherWalkTime68
69 Choice69 AV69 B_AccessTime * AccTime69 + B_stopLight * stopLight69 + B_NumofRoutes * NumofRoutes69 + B_Str_Tr * MinTr69 +B_Str_TT * MinTT69 + B_Train * Train69 + B_MinWalkTime * MiniOtherWalkTime69
70 Choice70 AV70 B_AccessTime * AccTime70 + B_stopLight * stopLight70 + B_NumofRoutes * NumofRoutes70 + B_Str_Tr * MinTr70 +B_Str_TT * MinTT70 + B_Train * Train70 + B_MinWalkTime * MiniOtherWalkTime70

[Expressions]

[Model]
// $MNL stands for "Multinomial logit model",
$MNL

APPENDIX E-2 CHOICE SET FORMATION CODE USED IN MATLAB 2013

```
current_ID = M(1,1);
list_of_rows = [1];
for i=1:length(M)
  row_ID = M(i,1);
  if (row_ID ~= current_ID);
    list_of_rows = [list_of_rows, i];
    current_ID = row_ID;
  end
end

DN1=[];
POSITIONS=[];
CombinedAllCases=[];

for i=1:length(list_of_rows) - 1          print=i;
  CombinedEachCase=zeros([1 (CHOICESETSIZE*16+6)]);                 %16
variables

  clear nest_subgroup nest_subgroup1 nest_subgroup2 nest_subgroup3
nest_subgroup4 V1 V2 Unique_CS_Index1 Unique_CS_Index2
  clear mu_Nest1 mu_Nest2 GIN EF CF TermB TermE CS_Index1 CS_Index2 count1
count2

  start_ID = list_of_rows(i);
  end_ID = list_of_rows(i+1)-1;
  subgroup_All = M(start_ID:end_ID,1:22);    %Total 22 columns
  subgroup_Path = M(start_ID:end_ID,2);     %path is in column 2

  subgroup_Prob = M(start_ID:end_ID, MUU);  %mu=1 > 20; mu=5 > 21; mu=10
>22

  Options = length(subgroup_Path);
  CaseID=subgroup_All(1,1);

  MTT_test=1.1*min(subgroup_All(:,3));
  MAT_test=1.1*min(subgroup_All(:,6));
  MTR_test=1.1*min(subgroup_All(:,7));

  Available=zeros([1 CHOICESETSIZE]);
  PathNumber=zeros([1 CHOICESETSIZE]);
  TravelTime=zeros([1 CHOICESETSIZE]);
  AccessWalkTime=zeros([1 CHOICESETSIZE]);
  NumberOfTransfer=zeros([1 CHOICESETSIZE]);
  WalkTime=zeros([1 CHOICESETSIZE]);
  Only_Train=zeros([1 CHOICESETSIZE]);
  Only_Bus=zeros([1 CHOICESETSIZE]);
```

```
Path_CF=zeros([1 CHOICESETSIZE]);
LightingStop=zeros([1 CHOICESETSIZE]);
Attractiveness=zeros([1 CHOICESETSIZE]);
Probability=zeros([1 CHOICESETSIZE]);

MTT=zeros([1 CHOICESETSIZE]);
MAT=zeros([1 CHOICESETSIZE]);
MTR=zeros([1 CHOICESETSIZE]);

CFactor=zeros([1 CHOICESETSIZE]);

ChoiceSet=[];
subgroup_All(:,23:26)=0;     %4 extra column created after 22 columns

for jjj=1:Options

  %Determining the basic strategy attributes

  if subgroup_All(jjj,3)<=MTT_test
    subgroup_All(jjj,23)=1;
  end
  if subgroup_All(jjj,6)<=MAT_test
    subgroup_All(jjj,24)=1;
  end
  if subgroup_All(jjj,7)<=MTR_test
    subgroup_All(jjj,25)=1;
  else
    subgroup_All(jjj,25)=0;
  end

end


if Options<= CHOICESETSIZE
  ChoiceSet=subgroup_All;

  PathNumber(1:Options)=subgroup_All(:,2)';
  TravelTime(1:Options)=subgroup_All(:,3)';
  AccessWalkTime(1:Options)=subgroup_All(:,6)';
  NumberOfTransfer(1:Options)=subgroup_All(:,7)';
  Only_Train(1:Options)=subgroup_All(:,8)';
  Only_Bus(1:Options)=subgroup_All(:,9)';
  WalkTime(1:Options)=subgroup_All(:,4)';
  Path_CF(1:Options)=subgroup_All(:,18)';
  LightingStop(1:Options)=subgroup_All(:,14)';
  Attractiveness(1:Options)=subgroup_All(:,19)';
  Probability(1:Options)=subgroup_All(:,MUU)';

  MTT(1:Options)=subgroup_All(:,23)';
```

```
    MAT(1:Options)=subgroup_All(:,24)';
    MTR(1:Options)=subgroup_All(:,25)';

    CFactor(1:Options)=subgroup_All(:,26)';
    Available(1,1:Options)=1;

  else
    %Checking for the chosen alternative in the nest and thus
    %defining the ModifiedChoiceSetSize
    for kk=1:Options
      if subgroup_All(kk,2) == 9999
        CS_Index = kk;
        ModifiedChoiceSetSize=CHOICESETSIZE-1;
        break;

      end
    end

    %Sampling for DN1
    for j=1:ModifiedChoiceSetSize
      Subgroup_cum_prob = cumsum(subgroup_All(:,MUU)); %calculating
cumulative probabilities
      rng(0,'twister');            %needed for Random number generator
      rng('shuffle');
      RandomNumber = (1-0).*rand(1,1);          %generates random number
between 0 and 1

      %Finding the option that has been picked by the random number
      %generator
      for k=1:length(Subgroup_cum_prob)
        if RandomNumber <= Subgroup_cum_prob(k,1);
          Path_Index = k;
          break;
        end
      end
      CS_Index = [CS_Index;Path_Index];

    end

    Unique_CS_Index=unique(CS_Index(:));
    count=(hist(CS_Index(:),Unique_CS_Index))';

    LengthCS=length(Unique_CS_Index);

    for j=1:LengthCS

      ChoiceSet(j,:)=subgroup_All(Unique_CS_Index(j,1),:);
      ChoiceSet(j,17)=log(count(j,1)/subgroup_All(j,MUU));
```

```
    end

    PathNumber(1:LengthCS)=ChoiceSet(:,2)';
    TravelTime(1:LengthCS)=ChoiceSet(:,3)';
    AccessWalkTime(1:LengthCS)=ChoiceSet(:,6)';
    NumberOfTransfer(1:LengthCS)=ChoiceSet(:,7)';
    Only_Train(1:LengthCS)=ChoiceSet(:,8)';
    Only_Bus(1:LengthCS)=ChoiceSet(:,9)';
    WalkTime(1:LengthCS)=ChoiceSet(:,4)';
    Path_CF(1:LengthCS)=ChoiceSet(:,18)';
    LightingStop(1:LengthCS)=ChoiceSet(:,14)';
    Attractiveness(1:LengthCS)=ChoiceSet(:,19)';
    Probability(1:LengthCS)=ChoiceSet(:,MUU)';

    MTT(1:LengthCS)=ChoiceSet(:,23)';
    MAT(1:LengthCS)=ChoiceSet(:,24)';
    MTR(1:LengthCS)=ChoiceSet(:,25)';

    CFactor(1:LengthCS)=ChoiceSet(:,26)';
    Available(1,1:length(Unique_CS_Index))=1;
  end
  ChosenOption=1;
  CS_Size=size(ChoiceSet,1);
  PurposeWork=CSA(i,7);
  Weekday=CSA(i,26);

  CombinedEachCase =[CaseID, ChosenOption, CS_Size, Options, Available,
PathNumber, TravelTime, AccessWalkTime, NumberOfTransfer, Only_Train,
Only_Bus, WalkTime, Path_CF, LightingStop, Attractiveness, Probability,
MTT, MAT, MTR, CFactor,PurposeWork,Weekday];
  CombinedAllCases=[CombinedAllCases;CombinedEachCase];

  % in POSITIONS we have the following information in different cells
  % 1. Case Number  2. start position of DN1 3. end position of DN1


  POSITIONS=[POSITIONS; [i, size(DN1,1)+1, size(DN1,1)+size(ChoiceSet,1)]];
  DN1=[DN1; ChoiceSet];

end
```