THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Strategic Approach In Analysing The Demand For Park-and-Ride Facilities

Bibhuti Sharma

B.E. (Civil), M.Eng. (Transport)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2019

School of Civil Engineering

Abstract

This study aimed to understand demand of park-and-ride (PNR) during different times of day. Surprisingly, there is no significant research in the area of dynamic demand of PNR. To fill this gap, this study of PNR demand was done in three stages. First, to understand why PNR users choose one PNR lot versus another, PNR lot choice models were developed. PNR lot choice behaviour was studied using two decision constructs, random utility maximization (RUM) and random regret minimization (RRM). Second, in order to understand the nature of utilization of PNR lots, a discrete time hazard model was developed based on the car arrival data in the morning period, at PNR lots. Finally, mode choice models (including PNR as one of the modes) were prepared to understand the choice of PNR as a mode.

From the developed lot choice models, it is understood that the PNR users' choice of PNR lot could also be explained by the RRM concept. In absence of any applications of RRM in PNR modelling, these new models serve as an important contribution. Further, the lot choice models suggested that the utilization of PNR lots is endogenous in nature. The identification of utilization as an endogenous variable is performed for the first time. The correction of endogeneity is completed using a two-stage control function method. Since the correction of endogeneity in the case of discrete choice transport models is a relatively new area, this work serves as additional evidence of the value of correcting for endogeneity using the control function method.

This research modelled the utilisation of parking spaces using a discrete-time logistic regression model and calculated the probability that each parking space is occupied at the end of one of 60 time-intervals between 4:00am and 9:00am on a weekday. The findings from the model suggest that the probability of a parking space to be occupied increases with a larger capacity of the PNR lot, a larger number of public transport services, and a lower walking time to the platforms. Moreover, the results suggest that a parking space is more likely to be occupied in PNR lots farther from the CBD until 8.00am, but it is more probable to be occupied in PNR lots closer to the CBD from 8.00am onwards.

Further, to understand the choice of PNR as a mode, mode choice models were prepared. With the aim of capturing a household's long-term decisions (like owing car, motorbike, bicycle etc.) on everyday short-term decision like mode choice, a portfolio-based multinomial logit framework was used to model the mode choice behaviour; where

portfolios are simply the set of modes enabled by the resources. Apart from conforming to some established results such as travellers are likely to choose modes which minimize their travel time), results suggested that long-term decisions do have an effect on the mode choice decisions. Further, a generalised nested logit (GNL) model was prepared as an alternative to the portfolio framework. In this model the portfolios (defined in the former model) act as nests. The GNL model was also able to capture the unobserved 'perceived activity set' of travellers as was the portfolio based model.

To connect the lot choice and mode choice model, the composite utility form of the lot choice model was used as a variable in the mode choice model. However, the variable is not significant, indicating that the results do not necessarily suggest that travellers will change their mode when they do not find parking at the PNR lots.

In overall, by answering questions such as why travellers choose one mode versus another and why PNR users choose one PNR lot versus another for different times of the day, and how PNR lot's utilization changes for different times of the day, this research explored the PNR demand and established that PNR demand is dynamic in nature.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications during candidature

Peer-reviewed journal paper:

Sharma, B., Hickman, M., & Nassir, N. (2017). Park-and-ride lot choice model using random utility maximization and random regret minimization. *Transportation*, doi:10.1007/s11116-017-9804-0

Peer-reviewed conference proceedings paper:

Sharma, B., Hickman, M., & Nassir, N. (2016). Park-and-ride lot choice model with corrected endogeneity. Paper presented in the *14th World Conference on Transport Research*, Shanghai, China, 10-15 July 2016.

Sharma, B., Hickman, M., & Nassir, N. (2016). Park-and-ride lot choice model using random utility maximization and random regret minimization. Paper presented in the *Transportation Research Board 95th Annual Meeting*, Washington, DC, United States, 10-14 January 2016.

Sharma, B., Hickman, M., & Nassir, N. (2016). A study on the utilization of Park-and-Ride lots in South East Queensland. Paper presented in *38th Australasian Transport Research Forum*, Melbourne, Australia, 16-18 November 2016.

Publications included in this thesis

None.

Contributions by others to the thesis

None.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Research Involving Human or Animal Subjects

None.

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Mark Hickman for giving me an opportunity to pursue PhD under him. It has been an honor to be his PhD student. I appreciate all his contributions of time, ideas, and funding to make my PhD experience productive and stimulating. I could not have imagined having a better supervisor and mentor for my PhD study. Also, I am thankful to Prof. Carlo Prato for being my co-supervisor. His insightful comments and encouragement have helped me widen my horizon on research from various perspectives. My sincere thanks also goes to my another co-supervisor Dr. Neema Nassir who provided me with guidance and suggestions during my PhD.

A special acknowledgement goes to my office mates of many years and friends for the rest of my life: Elnaz Irannezhadd and Svitlana Pyrohova. I am grateful to all my other great office mates who have been supportive in every way.

I wish to thank the members of my dissertation committee: Associate Prof. Alexander Scheuermann, Prof. Luis Ferreira, and Dr. Jiwon Kim for generously offering their time, support, guidance for my thesis milestones reviews. I am also thankful to the administrative staffs from the School of Civil Engineering and Graduate school, UQ. A big “Thank you!” also goes out to everybody who helped us conduct the park-and-ride survey study.

Words cannot express how grateful I am to my parents for raising me with love of science and for supporting me in all my pursuits. Special thank you for helping me take care of my daughter during my PhD. A special thanks to my both families: the one I was born to and the one I was married to, for their continuous help and support during my study. Also, I am grateful to my brother who helped me understand and solve my computer related problems during my PhD, even from a distance. And most of all for my loving, supportive, encouraging, and patient husband Anurodh Khanal whose support during all the stages of this PhD is so appreciated. Thank you.

Keywords

Park-and-ride, Discrete choice model, Multinomial logit model, Endogeneity, Random utility maximization, Random regret minimization, Discrete hazard model, Parking space availability, Dynamic utilization, Time-dependent parking demand, Mode choice model, Mobility portfolio, Cross nested logit model

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 090507 Transport Engineering 50%

ANZSRC code: 120504 Land Use and Environmental Planning 20%

ANZSRC code: 140302 Econometric and Statistical Methods 30%

Fields of Research (FoR) Classification

FoR code: 0905 Civil Engineering 50%

FoR code: 1205 Urban and Regional Planning 20%

FoR code: 1403 Econometrics 30%

Contents

List of Figures

List of Tables

List of Abbreviations

CNL	Cross Nested Logit
DTMR	Department of Transport and Main Roads
EMP	Eight Mile Plains
GNL	Generalized Nested Logit
HTS	Household Travel Survey
KNR	Kiss and Ride
MNL	Multinomial Logit
NL	Nested Logit
PNR	Park and Ride
PTOD	Public Transport Origin Destination Survey
RRM	Random Regret Minimization
RUM	Random Utility Maximization
SEQ	South East Queensland
t-o-d	Time of the day

1 Introduction

Through this dissertation, my PhD research explored travellers' behaviour in using park-and-ride services. I applied the strengths of discrete choice models to understand the demand at park-and-ride (PNR) facilities. Though the application of discrete choice models for travel behaviour studies is well established, I explored various aspects of cutting edge discrete choice models to obtain new findings in the context of PNR facilities. PNR is a mode that promises a lot, but at the same time there are several controversies surrounding it. On one hand, PNR is considered to reduce congestion in the CBD, while on the other, investments in PNR are considered money that might be better used for transit oriented development. PNR is also questioned for not attracting enough users. The literature on PNR demand is scarce. I studied PNR travel behaviour by modelling the PNR lot choice behaviour and by modelling the mode choices of South East Queensland's travellers. I made use of existing travel surveys as well as a self-conducted PNR lot utilization survey.

PNR is considered an access mode to the public transit network and can effectively improve the accessibility and the mode share of public transit, especially in the central core of urban areas. By providing facilities close to transit stations for travellers to park their cars, PNR allows travellers to connect with public transportation more conveniently. PNR also plays a vital role in maximizing the reach of transit to low density areas where users otherwise cannot access transit by walking. Also, when car drivers become PNR passengers, there is a reduction in vehicle-kilometres of travel (VKT) (Turnbull et al., 2004). Consequently, PNR helps reduce congestion in the central business district (CBD) by encouraging users to travel by transit inside the CBD while parking their car at the fringes of the CBD. Similarly, Kiss-and-ride (KNR) allows users to be dropped off by car to continue their remaining journey via public transport. PNR and KNR are used whenever a car user decides that changing to public transport is advantageous to their journey in terms of travel time, travel expense, or both (Dickins, 1991).

Despite the admitted advantages that are associated with introducing the PNR mode in urban transportation networks, there are some controversies around this topic that motivate careful modelling practices. For one, park-and-ride facilities are often criticised for being expensive investments, as providing PNR facilities requires costly parking structures (Engel-Yan et al., 2014). In addition, from an urban planning point of view, PNR facilities may also restrict opportunities for transit oriented development (TOD), as they may cause local area traffic congestion and reduce station access by other modes such as local transit and bicycle. Underutilization of PNR lots, small reduction in vehicle kilo-meters of travel (VKT) compared to drive alone, attraction of existing public transport users, and influx of induced traffic to replace new PNR users, are the main challenges of PNR. Regarding these existing controversies, PNR planning has become a complex issue. Therefore, careful strategic planning of PNR initiatives is required to justify high investments (Tsang et al., 2005). Oversupply of PNR lots has undesirable economic, environmental, and communal effects; whereas under-supply causes spillover into adjacent streets, making it unacceptable to the local community (Bolger et al., 1992).

To understand the utilization of PNR lots it is necessary to analyse passengers′ PNR choice mechanisms. Prevalent literature suggests that the choice of PNR as a mode, and the choice of a PNR lot among many, depends upon the time-of-day (t-o-d). However, there are very few other explanations for it. There is no established understanding of how t-o-d affects a traveller's PNR lot choice decision, how t-o-d affects a mode choice decision, or how t-o-d affects the utilization of PNR lots.

1.1 Motivation

The intention of this research is to investigate the possibility of understanding the PNR travel behaviour of users using discrete choice models. Modelling PNR trips is complex.

When a person makes a travel decision, they make numerous choices. One can decide mode first and, based on the mode decision, s/he can choose a route, or vice versa. Normally the first decision is fixed and leads to the second decision. For example, Person *A* wants to go to work. S/he can choose a mode first, say public transport or car, and then choose which route to use. Or, s/he can choose a route first and then decide whether s/he prefers public transport or car. But it is not likely that, after travelling half way to her/his destination, s/he will change her/his main mode. However, this is not the case for PNR trips. For PNR trips, one has the flexibility to change her/his mind at any time during the trip. Person *A* decides to use her/his car in the first decision and chooses a particular route. During her/his journey from home to work, at any PNR lot, s/he can change her/his earlier decision. S/he can choose to continue travelling in her/his car or switch to public transport by parking her/his car in a PNR lot. Thus a person with knowledge of PNR lots to his/her destination has the potential to change mode, route, and lot choice decisions at any step of her/his travel.

Let's say a person with perfect knowledge of PNR lots and routes available to his destination is travelling from node 1 to node 10 (represented by prefix N) as depicted in figure 1.1. From the point where he encounters the first PNR lot (say, node 3) to the point the last PNR is seen (say, node 7) where the PNR lot is represented by prefix P, he has the potential to change mode and route. Hence, in his journey from node 1 to node 10, there are multiple points at which he can change mode, route, or lot. This peculiar nature of PNR trips demands a detailed understanding of the choice behaviour of PNR users.

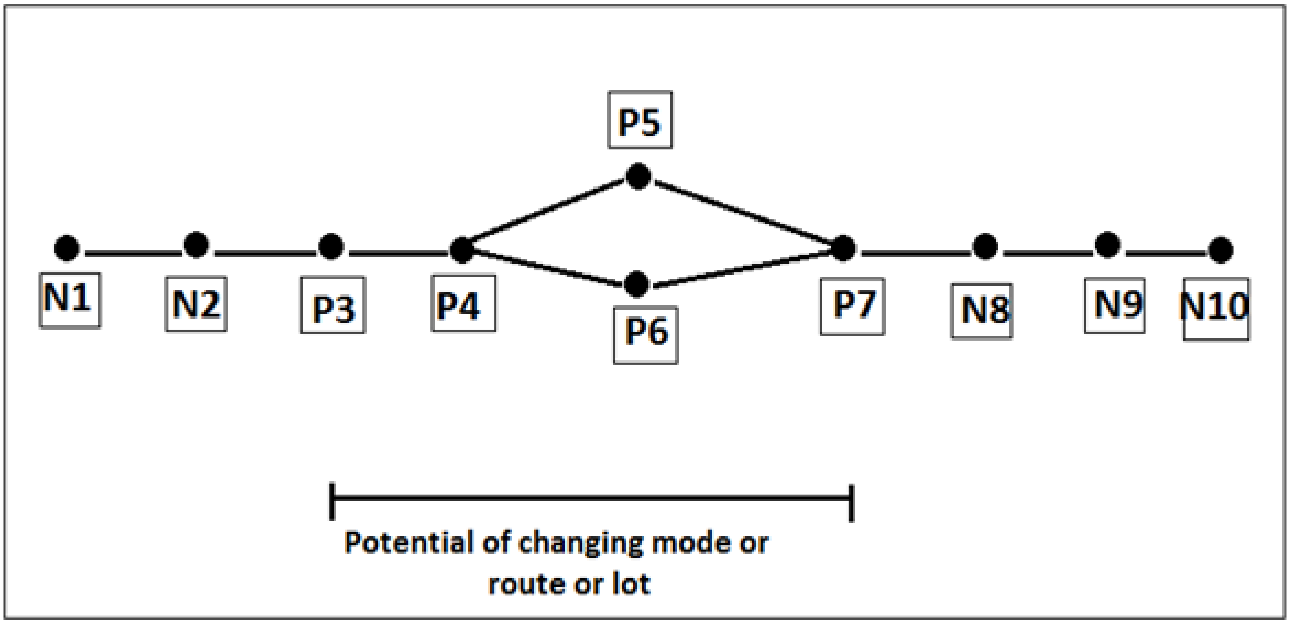

Figure 1.1: Decision making process for PNR users

There are three main factors that motivate this thesis:

Over-and-under utilization of PNR lots

Following heavy investments in PNR, it becomes essential to examine the reasons why some PNR lots are used more and some less. This question could be understood by analysing the travel behaviour of PNR users and finding why travellers choose one PNR versus another. For example, in figure 1.1, what is the reason a traveller would choose one lot versus another? (PNR lot 3, 4, 5, 6, or 7)

Incorporating time dynamics in PNR travel behaviour

For PNR planning, it is imperative to understand the exact demand at PNR lots. In order to correctly quantify this, apart from estimating a traveller's response to changes in characteristics of PNR lots only, it is equally important to understand the t-o-d factor. Parking lots for PNR differ from each other in many ways, in terms of their distance from the CBD, their capacity, and the availability of public transport services, to name a few. In figure 1.1, what is the probability that if person *A* chooses PNR lot 3 (or 4 ,5, 6, 7), s/he will get a parking space when s/he arrives there?

These motivations led to my research objectives. The overarching objective of this research is to understand PNR travel behaviour based on different times-of-day during a morning peak period. The sub-objectives are:

1. To model PNR lot choice behaviour based on different decision heuristics
2. To analyse the dynamic utilization of PNR lots
3. To model the mode choice behaviour of SEQ travellers that includes the PNR mode

Based on the objectives, a map of this research is presented in figure 1.2.

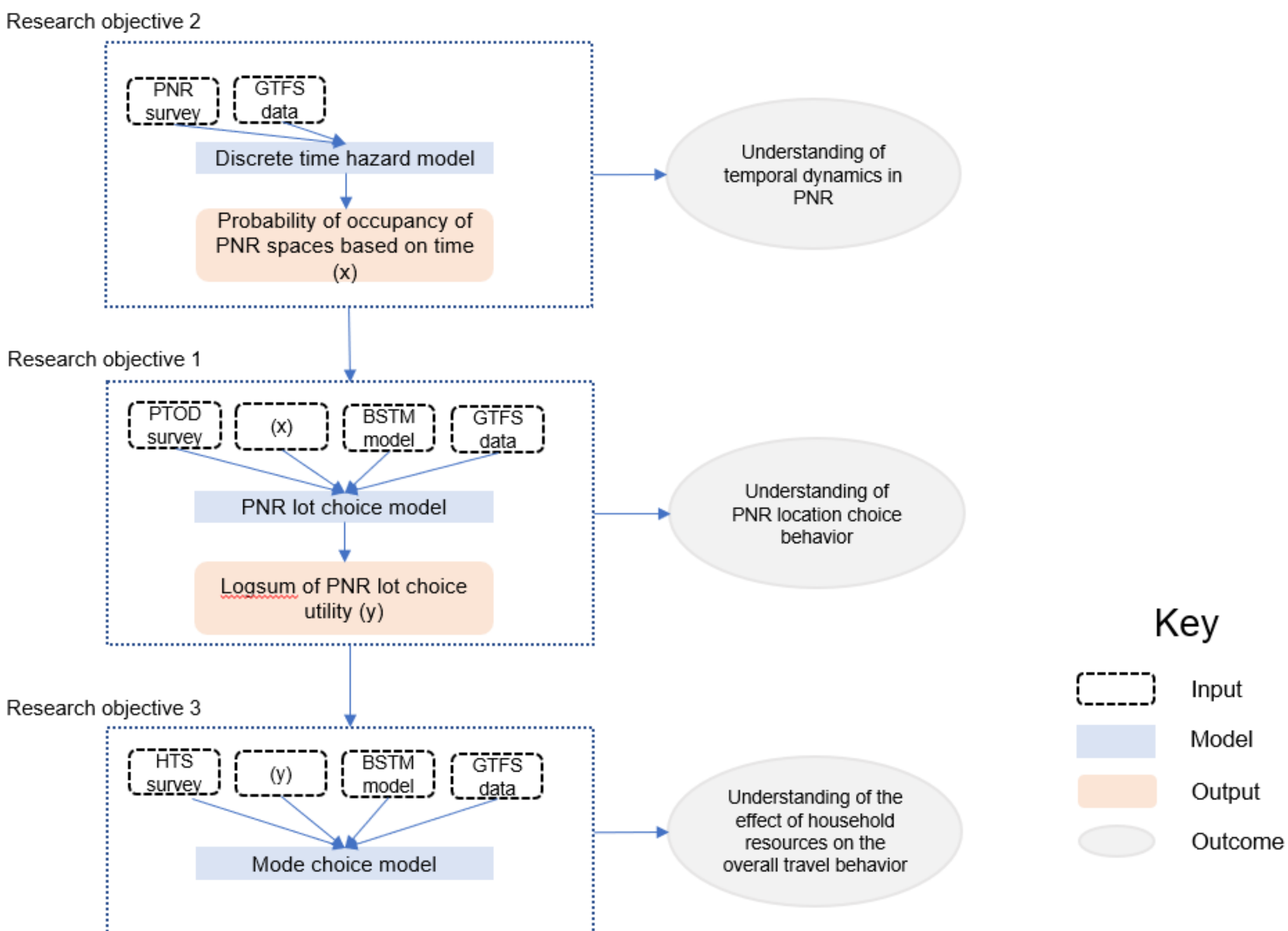

Figure 1.2: The overall workflow of thesis

The second objective of the research is reflected in the first block, i.e., the analysis of dynamic utilization of the PNR lots. For this, I studied the occupancy of parking spaces within PNR lots in the morning period in selected PNR lots in SEQ and developed a discrete time hazard model. The model provides an understanding of the temporal dynamics for the parking occupancy within PNR lots. The cumulative hazard obtained from this model feeds in to the first objective i.e., the analysis of PNR location (lot) choice behavior. For the analysis of location (PNR lot) choice behavior, I developed PNR lot choice models based on the multinomial logit framework. The logsum of the utility of PNR lots further serves as an input into the last objective i.e., the analysis of mode choice, which gives insight into the effect of household resource holdings on the travel mode decisions. Collectively, these models provide an understanding of the PNR travel behavior and address the knowledge gap. The details on the inputs to these model are discussed in the upcoming chapters.

1.2 Contribution of thesis

This thesis contributes to the area of PNR travel behaviour. The developed PNR lot choice models explore the temporal aspects of PNR travel. Though PNR commuters are understood to depart earlier than other mode users to secure parking spaces in PNR lots, the relation between the use of PNR lots and their dependence on the t-o-d was not clear in the literature. Also, the effect of the time-dependent use of PNR lots on the choice of PNR lots was not known. This research explores temporal and location choice aspects of PNR and uses these results to set up a mode choice model as shown in figure 1.2.

In terms of modelling technique, this research applies the regret minimization approach of modelling PNR lot choice, for the first time in PNR context, to investigate whether regret minimizing behaviour also explains PNR users' choice behaviour. Also, this research identifies and corrects the endogeneity in PNR lot choice models.

Other contributions of this thesis are the exploration of mode choice behaviour using a portfolio-based framework that addresses a housheold's resource holdings. Also, this research proposes a simple GNL model for mode choice, that is as competent as the commonly used intrinsic portfolio-based model.

1.3 Structure of the thesis

This thesis contains 6 chapters. I address the background of this research, aims, and objectives in chapter 1. Chapter 2 reviews the literature on PNR that is relevant to the study. Chapter 3 talks about PNR lot choice models based on the lot choice behaviour of travellers. In chapter 4, in an attempt to understand the dynamic utilization of PNR lots, the PNR survey conducted during this research is explained, and a discrete time hazard model derived from these survey data is explained. The model gives the probability of a parking space being filled by a certain time of day. Chapter 5, describes a mode choice model based on the fixed or more permanent transport resources that South East Queenslanders hold. Finally, chapter 6 marks the end of the thesis by presenting the conclusions drawn from this thesis, highlighting major contributions to the existing literature and discussing the outcomes and their implications for transport policy.

2 Literature Review

2.1 Introduction

This chapter provides a review of previous literature. Broadly, the literature on two main areas, *Park-and-ride* and *Discrete Choice modelling*, are reviewed as shown in figure 2.1. Section 2.2 argues that the study of PNR is important for transport planning and policy. The arguments draw upon evidence from the literature.

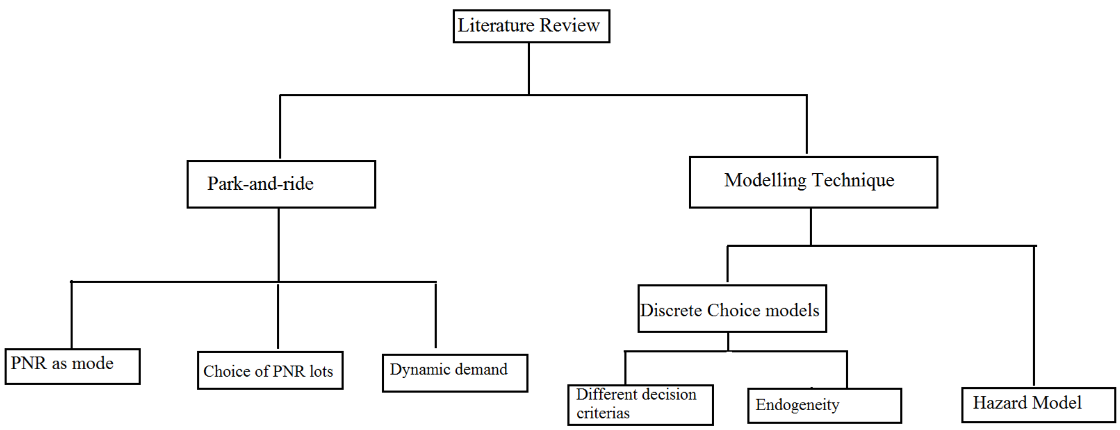

Figure 2.1: Branches of Literature reviewed

2.2 PNR definition

Park-and-Ride (PNR) is an alternative to door-to-door car use aiming to shift travellers from the exclusive use of the car to a combination of car and public transport. This shift results in an increase in public transport patronage, decrease in vehicle kilometres travelled, and most importantly a reduction in the number of cars that enter already congested cities (Zijlstra et al., 2015). The literature on PNR provides evidence of unintended effects: (i) not all users of PNR lots would travel door-to-door with the car, but there might be a substitution effect with a sustainable travel mode (e.g., bicycle, walk, public transport); (ii) not all users of PNR lots will use public transport (Mingardo, 2013), and (iii) the detour for reaching PNR lots might occasionally increase the vehicle kilometres travelled (Parkhurst, 2000). The result of these effects is that PNR lots might be overused in some instances and underused in others (Shirgaokar and

Deakin, 2005). Over-supply of PNR lots is not desirable; empty parking spaces are a waste of resources. Also, empty parking lots occupy space which is not environment and community friendly (Bolger et al., 1992). On the other hand, under-supply causes overspill parking in nearby streets and also potentially encourages car drivers to drive all the way to their destination (Bolger et al., 1992). While there exist unintended effects, there is relatively high investment (Tsang et al., 2005) on PNR infrastructure. PNR investment deserves careful planning and a detailed understanding of the reasons why a PNR lot may be over or underutilized.

2.3 PNR travel behaviour

In order to analyse the different aspects of PNR, the choice of PNR lots, and the choice of other modes, three branches of PNR literature as shown in figure 2.1 are reviewed. In section 2.3.1 I present the analysis of existing literature that explored the choice behaviour of travellers when it comes to choosing the PNR lot or train station etc. In section 2.3.2, literature on the choice of mode is studied. Section 2.3.3 is the discussion on mode choice studies taht are based on household resource holdings. Section 2.3.4 is about the investigation of those studies that are related to the time dynamics associated with PNR. Sections 2.4 and 2.5 are about discrete choice modelling and hazard modelling respectively.

2.3.1 PNR lot choice behaviour

Previous studies have analysed the choice of PNR lots explicitly (Mahmoud et al., 2014) as well as the access station choice of travellers (Debrezion et al., 2007) and have identified the access station choice process to be complex (Mahmoud et al., 2014). In my thesis I concentrated on studies of PNR lot choice.

The choice of PNR lot depends on how users perceive various PNR lots. Lot attributes are assessed by users not just to decide at which PNR lot to park, but also to decide if they want to use any PNR lot at all. Thus, lot attributes are significant factors in competition with other lots as well as with the drive alone mode. A great deal of research has been conducted to uncover the underlying attributes that affect a user's choice of a PNR lot. Users put a high value on the connecting public transport services at PNR lots and relative travel time by other modes while choosing PNR versus drive alone or door-to-door public transport (Bos et al., 2003).

Mahmoud et al. (2014) studied PNR access station choice and revealed that access distance and relative station direction (towards the workplace) influence transit station choice for PNR. The access distance variable used in their study is airline distance from home location to the PNR station location and represents static distance. Use of static distance does not fully capture the time-dependent details of the access trip. Users may perceive the same travel distance differently based on the roadway network structure and the traffic congestion in the network. For instance, during the peak hour the same path between an origin and an alternative PNR facility can take a longer travel time than during off-peak hours.

A study in the San Francisco Bay Area used three types of surveys: a user survey, an occupancy survey on PNR lots, and a focus group survey (Shirgaokar and Deakin, 2005). This study tells us that users prioritise lots with security, lighting, and a high-quality and reliable transit service. The lot survey explained that underused lots are located far away from a freeway or do not have active land uses nearby.

A study in the Toronto region (Mahmoud et al., 2014) used a multinomial logit (MNL) to model PNR access station choice. PNR trips from the morning peak period were analysed to find that access distance from origin to station and relative direction of station towards work were the parameters most affecting transit station choice for PNR. This study used travel distance as a parameter for the choice model which is a static parameter.

In another study, Bos et al. (2004) examined the attributes of PNR lots that affect the decision-making process of a car driver for choosing a PNR option. They made use of a hierarchical information integration approach, where different PNR lot attributes were grouped under five decision constructs: ‘parking’, ‘PNR facilities’, ‘connecting public transport’,‘time’ and ‘cost’. The study also used context-dependent variables like weather, travel purpose, heavy luggage, carpooling, and time of day. They found that social safety, quality of connecting public transport, and relative travel times by other transport modes were significant to the success of PNR facilities, whereas context variables have lesser impact. The time of day variable was not seen to affect the decision of choosing PNR versus drive alone. However, it is to be noted that the time of day variable referred to daytime or evening only.

A study in King County, Washington used PNR lot choice as an intermediate point choice in an overall mode choice decision between auto and transit and forecasts demand for PNR lots (Hendricks and Outwater, 1998). The demand forecasting model used a capacity constraint while assigning lots such that when a lot reaches its capacity, demand is redistributed to other lots based on the relative utilities of all lots. Further, the study investigated how users and non-users perceive lot attributes and acknowledged that even though lot security was considered an important attribute, its improvement did not necessarily attract more PNR users. The study used a combination of a demand model and user survey to understand the effect of capacity and user fees; users are more willing to use PNR if there are no capacity restrictions and lower user fees.

In order to gain a more accurate understanding of PNR choice behaviour, in addition to considering time-dependence of the access leg in the traffic network, the spatio-temporal limitation of the transit service has to be taken into consideration also. Due to these limitations, the travel experience in the transit network is very time-dependent and can effectively influence the PNR choice process. The effect of such limitations on passengers' choice of the fastest PNR travel option(s) was previously studied in a trip-based structure (Khani et al., 2012) and in a tour-based structure (Nassir et al., 2012).

In this thesis, I accounted for these time-dependencies in the discrete choice modelling of the PNR choice process, by using time-dependent travel time variables for calculating the travel attributes in the transit network. Also, for the auto network, I made use of travel times specifically from the am peak interval out of four available intervals: am, pm, mid-day off peak (dop), and night off peak (nop).

There are several other dimensions of PNR studies which deal with where to locate PNR lots (Yushimito et al., 2012) and when to build a new PNR lot (Coffel et al., 2012). Engel-Yan et al. (2014) applied the station access planning guidelines given by Coffel et al. (2012) to determine when to build new PNR lots and how many spaces to provide for a PNR lot. Moreover, the size of a PNR lot should be governed by the estimated parking demand, connecting public transport service and frequency, street system capacity, availability of reasonably priced land, and environmental constraints (Coffel et al., 2012).

2.3.2 PNR as mode choice

The second category of literature concerns the choice of mode including PNR. "Park-and-ride, as a mode choice option, often has been treated under the umbrella of access to transit" (Mahmoud et al., 2014). There are studies explaining the choice of PNR mode versus drive alone mode and other modes (Hendricks and Outwater, 1998),(Khan et al., 2007),(Hamer, 2010),(Qin et al., 2013),(Yamamoto et al., 2002). Risa Hole (2004) used a binary logit to model choice of PNR versus drive alone based on the stated choice experiment conducted with commuters. Results indicated that factors like income, gender, how many vehicles a household owns, and parking availability at work affected the choice of PNR service. Lower income people and females were more likely to use PNR. However, the more cars available in a household, the lower the chances of using PNR. Also, parking availability at work was studied in conjunction with an *arrival time at work after 9 am* variable. If the arrival time at work was later than 9 am and there was limited parking at work, people were more likely to use PNR. Also, walk time and wait time had a negative effect on choosing a PNR alternative, as anticipated.

Li et al. (2007) found that the number of parking spaces at the station, and the quality or frequency or fare of transit service, play an important role in determining the choice of PNR over other alternatives like auto and transit with walk access.

Likewise, another study in Shanghai also used binary logit and concluded that users are more likely to use PNR to save time rather than to save money when compared to the drive alone mode. However, the travel time variable used in the study does not reflect real travel time from the origin to the destination, as it was just a value chosen by the user as a measure of acceptable travel time say, 30 minutes or 45 minutes etc. given that it takes 1 hour to reach the destination. Other insights from the study were that lower income people and young people were more keen on using PNR. It is interesting to observe that people who owned a motorcycle or had a child or an elderly person at home were also likely to use PNR Liu et al. (2012).

Additionally, there is a study on mode choice behaviour of users based on decision field theory (Qin et al., 2013). They used eye movement tracking technology to calculate the deliberation time required by users to answer questions relating to mode choice. The study suggested that incentives, like providing more information on Park-and-ride, a parking fee, and the number of car parks available, may help to improve the deliberative process which in turn may help people choose park-and-ride. Further, to analyse the choice of users for a dynamic PNR system or drive alone mode, Yamamoto et al. (2002) used a decision tree as a production rule of the driver's decision, and compared it with a semi-ordered lexicographic model, and found that the latter performs better. The study also revealed that information on parking congestion in the central business district (CBD) has a significant effect on dynamically choosing dynamic PNR versus drive alone.

In their mode choice model including car, metro, PNR and bus, Li et al. (2017) found a significant relationship between in-vehicle time, crowding of public transport services, and travel time, but no effect between in-vehicle crowding and travel time reliability. Recently Choudhury et al. (2018) used a stated preference survey to study choice of mode, departure time, and occupancy (a multi-dimensional choice framework) in Libson for shared taxi, one-way car rental, and a combination of PNR and school bus facilities. Results suggested that for non-commute trips, users prefer one-way rental car and shared taxi, while for commute trips, they prefer the improved version of traditional public transport which are complemented by smart mobility options.

2.3.3 Mode choice based on household resource holdings

The choice of mode is a short term decision that travellers make. A person's mode choice decision could be based on whether s/he owns a vehicle at home (Miller and Roorda, 2003)(Katoshevski et al., 2015), the availability of PT services, and mobility enablers/restictions. Le Vine et al. (2013) defined *perceived activity set* (PAS) as "the array of activities which a person views (at a particular point in their life) as encompassing their potential travel needs, when making decisions that structurally affect their accessibility". They further investigated the concept by analysing people's "*mobility resource holdings*" at their household. A mobility resource is something that facilitates travel; it could be a product, service, information or anything that supports one or more travel options.(Le Vine et al., 2013). In this thesis, mode choice models are prepared in which long-term decisions of holding a mobility resource(s) may influence the short-term (mode choice) decision.

2.3.4 Dynamic Utilization of PNR lots

An emerging branch of PNR research is concerned with dynamic PNR demand. The attractiveness of PNR depends on the time of arrival at a PNR lot (Tsang et al., 2005). The later the arrival time to the lot, the lower the probability of getting a spot close to the platform (or the spot itself); this lengthens the transfer time and hence, the lot is less attractive. This idea corresponds with another finding on commuter parking in the morning peak (Arnott et al., 1991) that parking lots are used in the order of increasing distance from the CBD. Further a study (Nurul Habib et al., 2012) identified that there is a relation between parking type choice and the starting time of travel. To illustrate, if users need to park their car for a longer time, they prefer the morning time to start their travel activities. Afternoon time caters to those activities that require a shorter duration of parking. Also, users with longer travel time requirements are likely to avoid travelling in the peak period. This study (Nurul Habib et al., 2012) can be taken as a reference for understanding the relation between the parking choice and activity start time. However, this study included drive alone modes only.

To my knowledge, there is no literature that elucidates the relationship between parking type choice, parking duration, and activity starting time for PNR trips. Especially for PNR trips, the start time of travel is expected to play a vital role because decisions for PNR trips rely on transit frequency and reliability (Bos et al., 2003; Shirgaokar and Deakin, 2005; Bos et al., 2004). Also, transit attributes change for peak and off peak periods.

The literature inferred that time dynamics can be captured in a model by introducing time related variables or by using different time slots as choices themselves. In the binary logit model, Risa Hole (2004) used a dummy variable *arrive at work later than 9 am* to address late trips. Apart from this, Nurul Habib et al. (2012) used a nested parking type choice within the trip start time choice in a generalized extreme value (GEV) model. Start time choices included time slots of certain hours: before 6:30 am, 6:31 to 7:30 am, and so on to after 7:30 pm.

Traditionally, utilization of PNR lots was simply calculated as the overall usage of a PNR lot in a day divided by it's capacity. In contrast, dynamic utilization refers to a time-dependent utilization of PNR lots. This means a single PNR lot's dynamic utilization varies throughout a day depending on it's attractiveness to users. This thesis aims to study dynamic utilization for the morning time period only.

Study within a PNR lot

The literature clearly shows that there is an association between time-of-day and PNR use. (Hamid, 2009) and (Chen et al., 2014) examined the utilisation profile of PNR lots by conducting surveys. However, the factors that contribute to different usage profiles within a PNR lot and among PNR lots, to the best of my knowledge, has not been explored. There are many studies (Foote, 2000)(Shirgaokar and Deakin, 2005)(Hamid, 2009)(Syed et al., 2009)(Chen et al., 2014) that performed counts and occupancy surveys in PNR lots. Among those who conducted car arrival and duration surveys (Chen et al., 2014) (Hamid, 2009), the number of PNR lots they surveyed was limited; Chen et al. (2014) surveyed 4 lots and Hamid (2009) surveyed 2.

Hamer (2010) made a comparative study of seven PNR lots in Melbourne before and after their capacity expansion. Two types of surveys were conducted: a count survey and an interview survey. The count survey was performed twice, once prior to the expansion of PNR facilities and a second time a year after the upgrades were completed. The count of all cars parked in and around the selected stations was made twice in a single day, at 10 am and 2 pm. An in-person questionnaire was conducted at each upgraded station as the parking usage survey during the weekday AM peak. The questionnaires targeted users' travel mode before and after the car parks were upgraded. The results suggest that the demand for PNR exceeded parking supply at all surveyed locations. Even after the upgrade, the parking supply was not sufficient as the demand increased further. This study gives many insights on the PNR situation in Melbourne; however, as they counted cars only at two specific times of the day, it gives little information on the usage of the PNR lots for other times of the day.

In their study, Wiseman et al. (2012) highlighted the disadvantages of PNR by studying the PNR activity in the Adelaide Entertainment Centre PNR facility on the fringes of the Adelaide CBD. They conducted an observational study on the PNR facility and a face-to-face questionnaire with PNR patrons. The observational study was performed on two days from 7 am to 9:30 am, a week prior to the face-to-face interviews. The face-to-face interview was performed on a single day from 7 am to 9 am. The impacts of the PNR facility focussed on three aspects: car interception, public transport abstraction and additional trip generation. Car interception occurs when there is a reduction in car trips due to the introduction of PNR, as people no longer drive into CBD but use the combination of car-mass transit. For car interception, they found that 29.8% of the new PNR lot users previously travelled to the CBD by car. PT abstraction occurs when users who use PT for their entire journey start using the car and mass transit combination due to the introduction of PNR. In terms of the public transport abstraction, it was found that prior to the PNR facility opening, 62.7% of PNR lot users made their entire journey to the city by bus or train. An interesting result is that unlike other studies, this new PNR facility did not generate any new trips. Further, they found that people who used to walk/cycle to the city started using PNR. Though these results are very interesting, as their study was based on only one PNR lot, generalization of the results can be risky.

Another study by Lin et al. (2014) in Perth examined access by the elderly to train stations, where they studied PNR as one of the modes of access alongside BNR (Bike and ride) and

WNR (Walk and ride). They conducted an intercept survey at 3 train stations to get the perceptions of people (including the elderly) on the accessibility to train stations. For measuring accessibility they used network connectivity, distance, facility and service quality, mixed landuse within 800m buffer around a railway station, and information on intermodal connections like PNR capacity at and outside a staion. Also, perceived accessibility was obtained based on the evaluation of elderly on the overall accessibility of each station. The perceived accessibility was found to be higher relatively than the measured accessibility. Olaru et al. (2014a) also studied the effectiveness of PNR in Perth by examining the difference between old and new PNR lots using a stated preference survey. They found that the basic facilities at stations are important to users.PNR users were classified to be of two types; ones that are interested in stations with good amenities, and others that are more interested in quick access to the stations (Olaru et al., 2014b).

Another study (Chen et al., 2014) on PNR in Perth developed a location-based service (LBS) application to provide PNR users with real time information on the best departure station, based on their current location and planned departure time. A weighted linear combination was applied to determine the best location based on four criteria:(i) the availability of parking spaces at a chosen station, (ii) the travel time from the user's location to a target train station, (iii) the frequency of trains, and (iv) the service quality of stations which is measured by users' perceived services and facilities provided by stations. In order to find the availability of parking spaces in the PNR facility, a fuzzy logic forecasting model was used; the data was obtained from a traffic flow survey conducted in PNR facilities. In those surveys, researchers stood in the entrance of the car park lot from 6:00am to 9:00am and recorded the time at which each car entered and left the car park, for a period of 5 days. In total, 4 PNR lots were surveyed. Further, in order to understand the service quality of stations, an intercept survey was performed from 3 pm to 7 pm, where the questionnaire included questions on users' perception on the parking availability, travel time, train frequency, and service quality of the station.

Apart from Chen et al. (2014), there is no significant research on time-dependent PNR utilization in Australia. Chen et al. (2014) studied only four PNR lots over a set time, i.e., 6 am to 9 am. Count surveys with a single snapshot of lot utilisation provide no information on the temporal dynamics of PNR lots, whereas intercept/car arrival surveys provide information on the temporal dynamics but they are more expensive than count and occupancy surveys. In this research my focus is on studying the temporal aspects associated with the occupancy of parking spaces in PNR lots. For understanding the details of the time-dependent utilisation of parking spaces in PNR lot, I chose to conduct a car arrival survey at PNR lots.

Among studies that investigated aspects of parking spaces within a parking lot, Caicedo et al. (2006) mentioned that while many studies discuss the relative advantage of one parking facility over another, only a few discuss the relative advantage of one parking space over another within a given parking facility, in terms of comfort and user aid. Further, Caicedo et al. (2006) argued that a traveller might prefer to park near an exit in order to reduce the time spent waiting to join the exit queue as well as to reduce the amount of time spent in queue itself. Caicedo et al. (2006) differentiated parking spaces within a general parking lot to the level of a parking floor only, but not specifically across parking spaces within each floor. In the context of PNR, the position of a parking space within a lot is found to play a role. Lai and Shalaby (2007) found that parking spaces in PNR lots which were closest to the subway entrances were filled first. The probability of getting a spot close to the platform influences the transfer time (car to transit) and hence the attractiveness of a lot (Tsang et al., 2005).

Thus, I make a distinction of parking spaces within parking lots based on their location in the lot, as this is expected to add value in this study.

2.4 Discrete choice modelling

Because I analyse the choice behaviour of travellers, the backbone of this thesis is Discrete Choice Modelling. Discrete choice modelling is a very broad topic. As a result, this section presents only those areas under Discrete Choice Modelling that are touched by this thesis.

The modelling framework used in this thesis is a simple Multinomial logit model for PNR lot choice analysis and a Generalized nested logit (GNL) model (Wen and Koppelman, 2001) for the mode choice analysis.

Popular mode choice models include the multinomial logit (MNL) and nested logit (NL). The MNL model suffers from the independence from irrelevant alternatives (IIA) assumption, meaning that the relative odds for any two alternatives (the ratio of probabilities of any two alternatives) is unaffected by the systematic utility of any other alternative(McFadden, 1978; Ben-Akiva and Lerman, 1985). The limitations of this assumption gave rise to the NL model where correlated alternatives can be treated in a nest of alternatives, and the effect of this nest can be incorporated in an upper-level logit model. However, even using the NL model, we are still unable to account for similarities between the PNR and car alone modes. To account for similarities among different pure and combined modes, a cross nested logit (CNL) can be used (Vovsha, 1997). Detailed literature on CNL is found in Bierlaire (2006). Bierlaire et al. (2001) used MNL, NL, and CNL for the mode choice model and concluded that allowing the use of more complex structures like the CNL model improves the quality of an estimated NL model. I use a GNL model (Wen and Koppelman, 2001), which is a generalised form of CNL, in my mode choice study.

2.4.1 PNR lot choice model based on Random Regret Minimization (RRM)

In the literature regarding of PNR utilization, there are many studies explaining the choice of PNR mode versus drive alone mode and other modes (Hendricks and Outwater, 1998; Khan et al., 2007; Hamer, 2010; Qin et al., 2013; Yamamoto et al., 2002). The majority of the existing literature is based on discrete choice models with a utility-maximization rule for describing the travellers' choice behaviour. A utility-maximization rule states that, when a decision maker faces a number of choice alternatives, s/he perceives a certain level of utility from each alternative, and s/he chooses the option that provides them with the greatest utility. Using the utility maximization concept, the choice probability of each alternative is then derived (Train, 2009).

Recently, the concept of regret minimization has been gaining popularity (Chorus et al., 2008; Chorus, 2010). A random regret minimization technique has been used to model travel mode or route choices and parking lot choices as well (Chorus et al., 2008). The regret minimization technique is based on the decision rule that travellers try to minimize their anticipated regret rather than maximizing expected utility, where regret is the experience a person feels when a foregone alternative performs better than the chosen one (Chorus, 2010). Both random utility maximization and RRM models are equally parsimonious; when choice sets are binary, the RRM model reduces to the RUM's linear additive binary logit model (Chorus, 2010).

The PNR lot choice is a complex process. It involves travel in both auto and transit networks. Therefore, when choosing a PNR lot, not just the attributes related to the PNR lot itself, but attributes associated with the auto network and with the transit network, are considered too. With a typical fully-compensatory random utility maximization (RUM) model, one assumes that the loss of utility due to a certain attribute can always be compensated by strong performance of other attributes. Alternately, one aims to explore the semi-compensatory assumption using the RRM model specification. Semi-compensatory behaviour suggests that improvement in one attribute of an alternative does not always compensate for an equally large reduction in the performance of another attribute (Chorus, 2010). Also, another property of the RRM model is its ability to capture the “compromise effect”. According to Chorus (2010), the compromise effect means that alternatives with an ‘in–between’ performance on all attributes in relation to all other alternatives in the choice set are likely to be chosen by choice-makers, when compared to extreme alternatives that have very poor performance on some attributes and a very strong performance on others. Especially in the context of PNR lot choice, I anticipate that passengers may display behaviours based on this compromise effect. This is because the PNR choice process resembles a trade-off situation between the auto and transit legs of travel. Similar to any typical human choice behaviour in trade-off situations, it is reasonable to assume that an “in-between” alternative could have a better chance of being chosen. For instance, reducing the origin-to-PNR lot time alone may not be a priority for typical travellers. It is reasonable to assume that travellers aim to minimize their travel impedance for the overall trip, while making sure they will get a parking space in their chosen lot. In this choice process, reducing (increasing) the network distance from origin-to-PNR (which is dependent on the time from origin-to-PNR) may mean increasing (reducing) PNR to-destination time. Therefore, a person is likely to choose a compromised option with reasonable travel attributes in both the auto and transit legs. In order to address the compromise effect, Chorus and Bierlaire (2013) compared three approaches; namely, adding a compromise variable, using the RRM form, and using a contextual concavity model (CCM). They found the RRM model to be the ‘most parsimonious’ among these models. Recently, Guevara and Polanco (2016) introduced the decoy effect in the context of transportation and explained the compromise effect as one type of decoy. Therefore, I anticipate that a model specification that can capture such effects, such as RRM (Chorus and Bierlaire, 2013) (Guevara and Polanco, 2016), may perform better in this choice situation. As a result, RRM model specifications are also included in this research.

2.4.2 PNR lot choice model with endogeneity

I have hypothesized that PNR lot use is related to the utilization, of the lots. The greater the utilization the more it is likely to be attractive to travelers. The probability of choosing a lot is a function of utilization and utilization in turn is a function of the traveler demand (that is attributable to travelers' choices). Even one single traveller's choice of a PNR lot affects the utilization and consequently the demand for that lot. When a person chooses a particular PNR lot, a space is taken up and the available capacity of the lot decreases by one space. Also, a change in utilization at a particular lot plays a role in making other lots more or less attractive. As a result, the utilisation of a PNR lot is endogenous with the traveller's choice of that lot.

Guevara (2010) studied endogeneity for a residential location choice model and suggested that endogeneity is an unavoidable problem for all spatial choice models. His study identified the price variable as an endogenous variable which is correlated with the error term. His thesis further indicates that in case of spatial choice models, there can be generally three causes of endogeneity: error in the variables, simultaneous determination, or the omission of relevant attributes that are correlated with the observed attributes.

The presence of endogeneity gives incorrect parameter estimates and can be misleading. Correction of endogeneity is essential and thus a model is needed that can correct for this effect.

The popular methods to correct for endogeneity, when the endogenous variable is continuous, are the BLP method (Berry et al., 1995) and the control function method (Guevara, 2010). 'BLP is a fixed-effects procedure, by product and market, to solve market-level endongeity' (Guevara, 2010). I used the control function (CF) method to correct the endogeneity in the model. The BLP method has its own limitations. Petrin and Train (2003) use both the BLP and the control function method on the choice of television options and concludes that both of the methods give fairly similar results in their study. The BLP method is unable to handle an individual level of endogeneity (Petrin and Train, 2003) (Guevara and Ben-Akiva, 2006) (Petrin and Train, 2010). Also, the BLP method is not consistent in settings where there are zero, one, or just a small number of purchase observations per product, because the constants for these product market shares are not identified (Train, 2009). In comparison to BLP, the CF method is better suited in my case as I was interested in individual level endogeneity. Specifically, I had 20 alternatives with attributes that are different for different observations and observations, are not separated by any market. Also, the number of observations for some alternatives may be few.

Studies have used the control function approach to correct price endogeneity in residential location choice models (Guevara and Ben-Akiva, 2006)(Ferreira, 2010). For the correction of the endogenous attribute price, the instrumental variables (IV) used are transaction cost (tax) paid by households (Ferreira, 2010) and average price of adjacent residential locations (Guevara and Ben-Akiva, 2006). Also, the control function method has been applied in other contexts like choice of television options, where again price is the endogenous variable and the instrument used is the average price in other markets that are served by the same operator (Petrin and Train, 2003) (Petrin and Train, 2010).

The CF method has its limitations too. One criticism is its dependence on IVs which are often very hard to find, and even when found they may suffer from the weak instrument problem (Guevara and Polanco, 2016)(Guevara, 2017)(Fernandez-Antolin et al., 2014). Recently, a multiple indicator solution (MIS) method has been proposed (Guevara and Polanco, 2016) which does not require IVs. However, the MIS method addresses only those types of endogeneity that arise from the omission of an attribute that could be measured by an indicator, whereas the CF method addresses the correlation with the error term in a broader sense, regardless of the source of endogeneity (Guevara and Polanco, 2016). My aim is to correct endogeneity caused by omitted attributes and the simultaneity of demand estimates, and thus I chose the CF method.

2.5 Hazard Modelling

The availability of parking spaces in the lot is likely to change with the time-of-day. PNR location choice depends not only on the geographical location but also on the departure time in terms of finding a parking space (Weiss and Habib, 2017). The tendency of PNR commuters to depart earlier than commuters using other modes indicates a high correlation between parking availability and the departure time (Huang et al., 2016). I aimed to study the times and rates at which parking lots tend to fill by observing the times at which commuters arrive to park at a PNR lot. In a parking lot, some spaces get filled while some remain empty past the time at which a survey of the lot may end. The parking spaces which are not filled by the end of the study period (here in the lot survey) are censored data. General statistical tools do not account for parking records that are left or right censored. Censored data cannot be ignored; empty parking spaces carry important information. An empty parking space does not necessarily mean it will remain empty forever; it can be filled up after the end of the lot survey. In order to account for these censored data((records of parking spaces that are left or right censored), I used survival analysis; survival analysis incorporates data from both the non-censored and censored cases.

I used a discrete-time method (DTM) of survival analysis. There are several benefits (Keiley et al., 2007) of using a DTM of survival analysis instead of continuous-time method (CTM). The first reason why I used DTM here is because the data were collected for discrete time intervals of five minutes. Further, I planned to study the effect of the frequency of public transport service available near the PNR lot as one of the covariates in the model. This covariate is time-dependent; i.e., the value of the frequency of public transport service was different for different time intervals. Addressing this in the continuous-time method would be difficult (Keiley et al., 2007). I used DTM as it allows the inclusion of both time-dependent and time-independent covariates. Several other studies that used DTM include, as examples, (Chen and Xie, 2014) (Singer and Willett, 2003b) (Singer and Willett, 1993)(Allison, 1982). To specify and calibrate the DTM, the important task is to specify how this hazard rate depends on time and the explanatory variables. The most popular modelling approach as summarized by Allison (1982) is the logistic regression function.The models of discrete-time survival analysis can be fit using standard logistic regression analysis software (Singer and Willett, 1993).

3 Park-and-ride lot choice model

3.1 Introduction

This chapter presents PNR lot choice models developed for undestanding PNR users' lot choice behaviour. Models were developed based on the simple multinomial logit (MNL) framework using two different decision frameworks/paradigms, namely random utility maximization (RUM) and random regret minimization (RRM). Further, the issue of endogeneity in the PNR lot choice model was studied, and is discussed here.
Firstly, section 3.2 presents the study area of this thesis, which is South East Queensland (SEQ). Section 3.3 includes description of the datasets I used for the preparation of the model. Section 3.4 sheds light on the choice set generation process. In section 3.5, the variables in the model are explained. The framework used in developing a PNR lot choice model based on the RUM concept are shown in section 3.6. PNR lot choice behaviour was also explored from the perspective of RRM concept; the framework which constitutes section 3.7. Further, in section 3.8, the results obtained from the RUM model and RRM model are discussed. The sections up to 3.8 report the models that were based on both formal and informal PNR lots. In order to understand the effect of a PNR lot's *utilization* on its attractiveness, a lot choice model based only on the formal PNR lots was developed; *utilization* of informal lots is tricky to measure. Therefore, section 3.9 is about the PNR lot choice model that was developed using formal PNR lots only. From the developed PNR lot choice model, the *utilization* variable was found to be endogenous. Thus, the endogeneity in the PNR lot choice model is discussed and corrected using a *Control Function Method*. Lastly, in section 3.10, the conclusions developed from this research are presented.

3.2 Study Area

I chose South East Queensland (SEQ) as the study area for this thesis. SEQ is home to approx. 3.04 million people. There are over 170 formal PNR facilities across SEQ providing around 27,000 parking spaces, including 21,500 spaces on the city rail network. The cost of construction and land combined adds up to A$ 30,000 per space. Also, the annual maintenance cost per space is A$ 250. Average utilization of PNR facilities is around 80 %. 35 % of the facilities are over capacity. It is interesting to note that Brisbane of SEQ, has more PNR spaces per person than some of Australia's other major cities like Adelaide, Melbourne, Perth, or Sydney (Government, 2014).

3.3 Dataset

The datasets used were taken from the public transport origin destination (PTOD from here on) survey and the PNR number plate survey conducted by the Department of Transport and Main Roads (DTMR here on) as described in sections 3.3.1 and 3.3.2 respectively.

3.3.1 PTOD survey

The PTOD survey was a survey of public transport trips throughout the Translink service area within SEQ conducted for a period of four weeks in May 2010. Translink is an agency of the DTMR which plans and manages bus, ferry, and train services and integrated ticketing within SEQ. Within the PTOD questionnaire, public transport access modes were classified as PNR driver, KNR/PNR passenger, cycle, and walk.

3.3.2 PNR number plate survey

The PNR number plate survey was conducted by the DTMR at various parking lots in SEQ on random days at different times to investigate the usage of PNR lots. The data included the number plates of vehicles parked both in the designated Translink lot (formal) as well as vehicles parked in neighbouring (informal) parking areas.

During preliminary analysis of the number plate survey, I observed that, while many PNR facilities were full to capacity and have spillover parking in adjacent areas, some facilities remained underused. The ratio of usage to capacity from this survey was found to vary from 10 % to over 100 % (at certain times of the day) for different lots as shown in figure 3.1. A ratio greater than 1 (or 100 %) indicates spillover parking in the adjacent areas. It became interestingly relevant to study and understand why some PNR lots are underused and some are overused. Thus, in order to understand why PNR users would choose one PNR lot versus another, I developed PNR lot choice models based on two different decision frameworks. They are explained in the following sections.

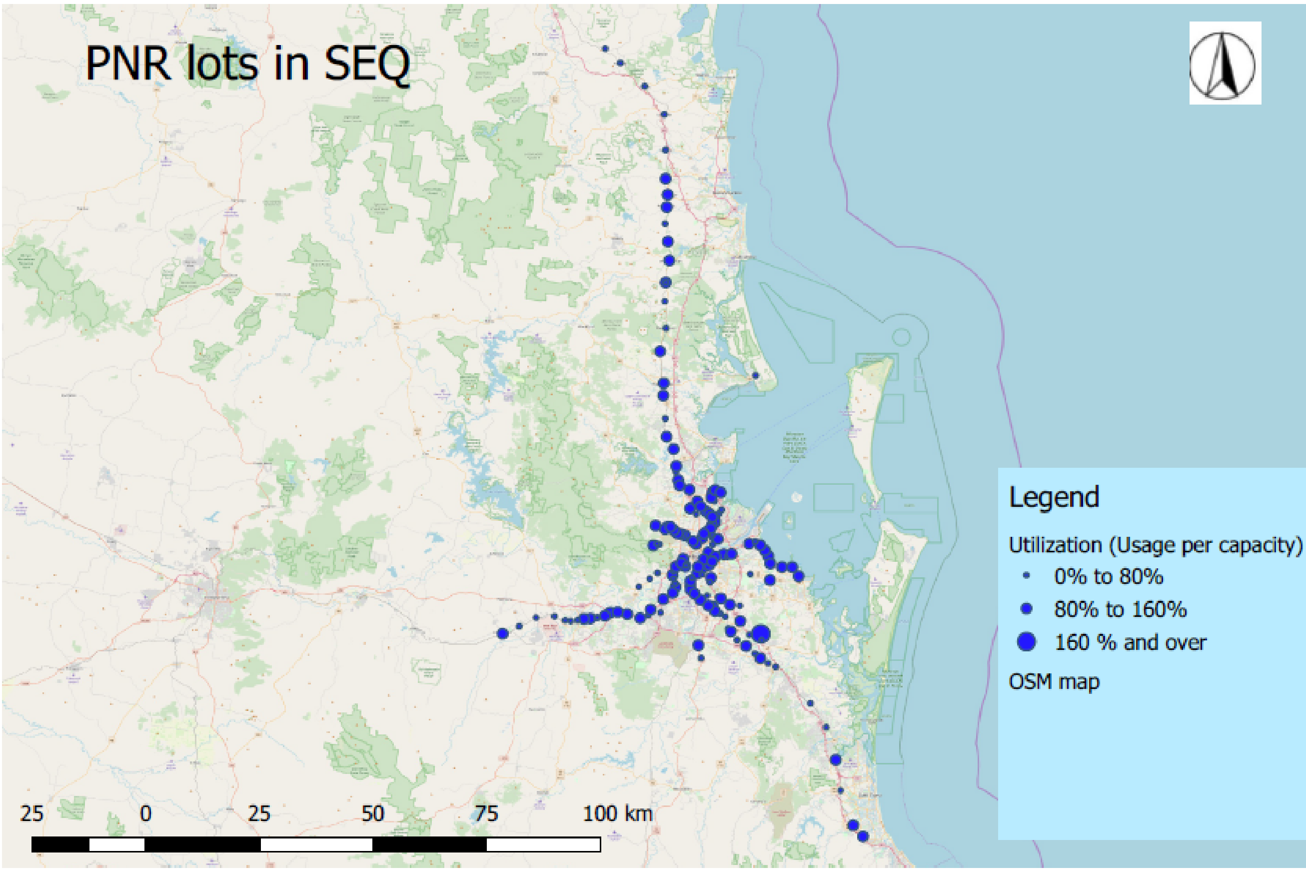

Figure 3.1: Utilization of PNR lots in SEQ

3.4 Choice set generation

Defining a choice set for the model was the first step for the lot choice model. There were altogether 418 PNR locations identified from the PTOD survey where people parked their cars to transfer to public transport. Some of these locations were formal (defined by Translink) PNR lots and some were informal. Not all 418 PNR locations could be considered in the universal choice set, as some would be infeasible in terms of distance/time. Thus, the individual choice set is comprised of a subset of facilities. As a first step in choice set generation, I considered only those PNR lots where the shortest travel time from the origin to the PNR lot (by automobile) was less than the shortest travel time from the origin to the destination (by automobile). The rest were excluded from the possible choice set. This was a data-driven decision; in the PTOD data all travelers chose PNR lots which took less time (by automobile) to reach than it took to reach their destination (by automobile).

The shortest travel time calculation was computed in EMME/4 software using the Brisbane strategic transport model (BSTM). The BSTM gave the travel time in each link on the roads of Brisbane. I calculated the shortest path from the origin to a PNR lot (auto network) based on travel time. I treated the origin of an individual as the origin (in EMME) and PNR lots (from choice set) as the destinations (in EMME) while finding the shortest travel time path. The total observed trips from the PTOD survey 201- were classified into four time slots: am, pm, mid-day off peak (dop) and night off peak (nop), based on the time of travel. Also, the BSTM results had the same four sets of travel times for each link. In this thesis I aimed to investigate the choice of PNR lots, so I focused on access PNR trips using am (morning) observations only.

After exclusion of unreasonable PNR lots, there remained a plausible number of alternative PNR lots for a given trip. Out of these lots, a random sample of 19 PNR lots was created for each traveller. For random sampling in MNL (Nerella and Bhat, 2004), a minimum sample should be an eighth of the size of the full choice set. These 19 lots, along with the one chosen lot, made 20 alternative PNR lots for each individual. As the attributes and existence of the remaining lots would not affect the relative probability of choosing a particular PNR lot within the subset of 20 alternatives, there would not be any influence of excluded alternatives on the consistency of the maximum likelihood estimator (Train, 2009).

I classified all the observations seen in the PTOD to be either access PNR trips or egress PNR trips. Access PNR trips refer to those where the first half of the journey is made in a car and then users transfer to public transport, whereas egress PNR trips refer to those where the car is used in the second half of the journey. In access PNR trips, travellers choose among various PNR lots to park their car and transfer to public transport. However, for the egress PNR trip there is a constraint wherein users have to return to the same lot where they parked their car earlier, as explored in Nassir et al. (2012). In this context, access PNR trips are mostly home-based trips and egress PNR is actually not a choice (strictly deterministic). However, one of the limitations of this random sampling procedure could be that it does not consider spatial auto-correlation among the PNR lots if it exists.

3.5 Explanatory attributes

Travelers' sensitivity to changes in lot characteristics, accessibility, and availability were addressed using a wide range of variables. Travel time was one of the main attributes in choosing a PNR lot. Given that the origin to PNR lot travel occurs in the auto network and that the PNR to destination travel occurs in the transit network (Nassir et al., 2012) (Khani et al., 2012), travel time in the auto network and in the transit network were calculated separately. In the auto network, the shortest travel time from the origin to a PNR lot was found using an all-streets network with EMME/4, conditioned on the time of day (am, pm, dop, or nop).

Finally, in the transit network, the time-dependent shortest travel times from all alternative PNR facilities to the destination were found using a backward time-dependent shortest path (BTDSP) algorithm (Nassir et al., 2012). The BTDSP algorithm uses Google's general transit feed specification (GTFS) dataset for SEQ, taken from 2010 to match the PTOD survey results, meaning that the algorithm considered the actual bus, train, and ferry schedules.

In the execution of the BTDSP algorithm, I used a backward calculation. Instead of finding shortest path trees from PNR lots to the destinations, BTDSP was run from each destination backwards to find the shortest path tree to all of the PNR lots. At the initiation step of the BTDSP algorithm, since the start time of a trip at the origin was known from the survey, the destination of the trip could be assigned an approximate time label after the start time, such as 2 hours after the departure from the origin. This way, the algorithm calculated the shortest time-dependent transit time from all PNR lots to the destination, using one execution per observation in the survey. The shortest travel time in the transit network was separated into in-vehicle travel time, walk time, and wait time to capture the effect of a user's sensitivity to different components of transit travel time.

Another attribute that I anticipated would influence a user's lot choice was the cost of travel. To calculate cost in the auto network, I needed to know the fuel price and distance travelled, which in turn are dependent on the type of vehicles used. Since I did not have information on the type of vehicle used in the PTOD survey, I could not measure the exact cost of travel in the auto network. I used the network distance *netdist*, which is the distance travelled in the shortest travel time path, as a proxy for auto cost. In the transit network, public transport fares in SEQ are based on zones; the whole SEQ is divided into 23 concentric zones. Based on this zoning, the distance travelled by travelers was known. Apart from the zones travelled, the fare system of TRNSLINK further depends on the age (young adults or seniors), education type (students or other) and availability of *Go Card* (Go card or paper ticket). There was a concession for seniors, students, and *Go Card* users. Thus, I categorized all observations based on age, education type, availability of *Go Card*, and number of zones travelled to calculate the fare from each PNR lot to the destination. The variable was represented as *totfare* which is the cost of travel in the transit network. Figure 3.2 represents the travel times for a typical PNR trip in the auto and transit networks respectively.

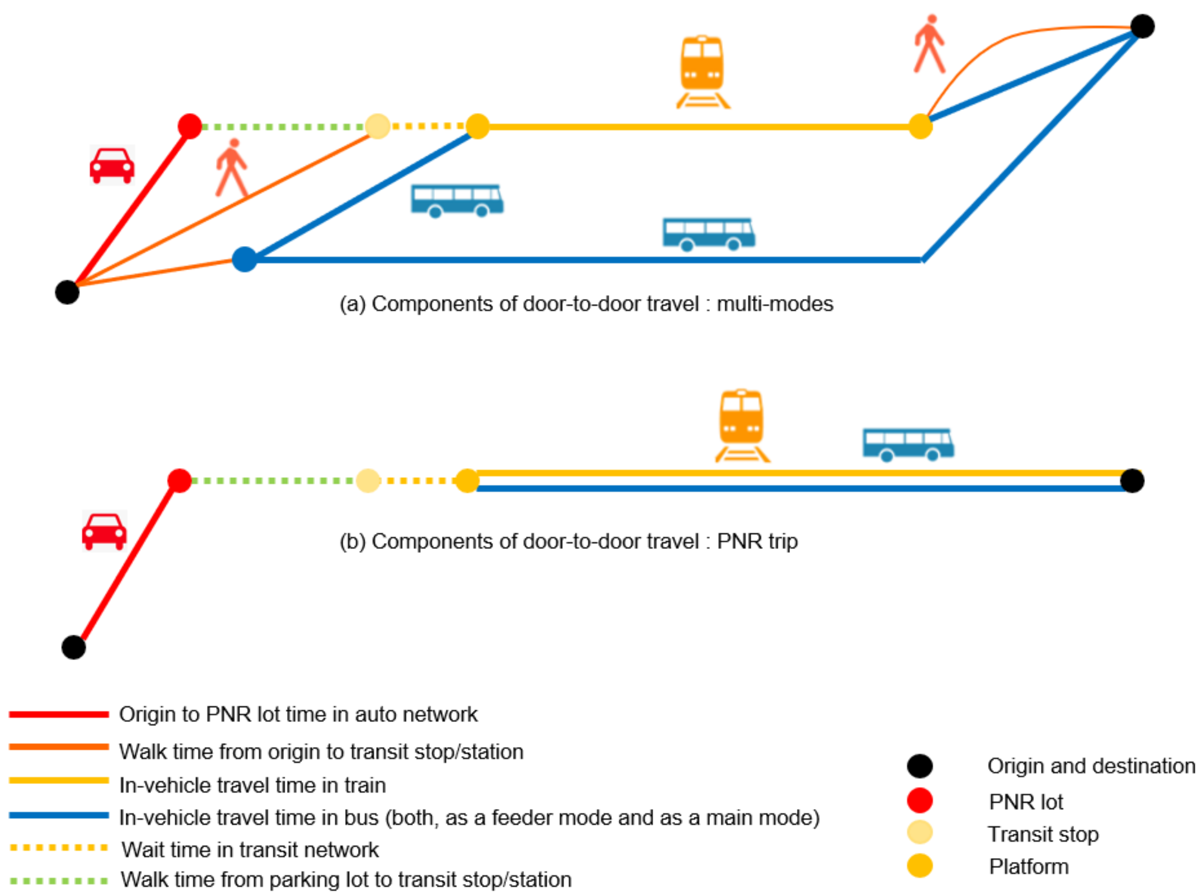

Figure 3.2: Components of door-to-door travel time

Further, I explored the proximity to a freeway or main transit route (Shirgaokar and Deakin, 2005). Most PNR lots in SEQ are near train stations, so I used a dummy variable called *train* to represent PNR lots which were served by train routes (versus only bus routes). The train lines in SEQ run nearly parallel to major highways so I used a variable called *freeway* to address those PNR lots that were within 1 km of freeway. Further, since the cost of travel is higher in Brisbane CBD (congestion), I used *cbdP* to represent PNR lots lying within the Brisbane CBD. Also, I used *for* to represent PNR lots that had formal parking. Use of variables like *freeway* and *cbdP* were expected to incorporate land use and network features within a PNR lot's proximity. The list of all variables used in the preparation of different lot choice models in this thesis, are as presented in table 3.1.

Table 3.1: Variables defining utility of a lot

Variables	Description
Capacity	Continuous variable, capacity of PNR lots
CbdP	Dummy variable, 1 for PNR lots lying in zone number 1 or 2 (CBD, and 0 otherwise
For	Dummy variable, 1 for PNR lots who are identified as formal parking lots by DTMR's Number Plate Survey and 0 otherwise
Free	Dummy variable, 1 for lots situated at a distance less than or equal to 1 km from major freeway/motorway and 0 otherwise
Invehtime	In vehicle travel from PNR lot to destination, in the shortest (travel time) path
Netdist	Distance traversed in the shortest travel time path in auto, from origin to PNR lot
Optime	Shortest travel time in auto from origin to PNR lot
Totfare	Total fare required to travel from PNR lot to destination, in shortest (travel time) path
Train	Dummy variable, 1 for PNR lot that serves a train station and 0 otherwise
Utl	Utilization of PNR lots; ratio of Usage to Capacity of PNR lots
Waittime	Waiting time spent from PNR lot to destination, in shortest (travel time) path

In different lot choice models I prepared, which are presented in the following sections, not all the variables shown in table 3.1 became significant. As previously stated, I set up a PNR lot choice model based on the MNL framework (Train, 2009),(McFadden, 1978). Biogeme (Bierlaire, 2003) was used to model the lot choice behaviour.

3.6 PNR lot choice model based on RUM

MNL based on RUM suggests that the user always chooses the alternative with the highest utility.

3.6.1 Model construction based on RUM

Utility is composed of observed (deterministic) and unobserved (random) components.

$$U_i = V_i + \epsilon_i \tag{1}$$

Where *i* is any PNR lot from 1 to 20

U is overall utility

V is deterministic utility

and ϵ is error term for unobserved utility

According to the definition of MNL (Train, 2009), the unobserved factors are uncorrelated and have the same variance across alternatives; i.e. they are independently and identically distributed extreme value random variables. Also, the theory of the extreme value distribution states that the difference between two extreme values follows a logistic distribution.
Thus, the density of the unobserved component of utility for user *A* is

$$f(\epsilon_{Aj}) = \mu e^{-\mu(\epsilon_{Aj})} \; * \; e^{-e^{-\mu(\epsilon_{Aj})}} \tag{2}$$

and the cumulative distribution is

$$F(\epsilon_{Aj}) = e^{-e^{-\mu(\epsilon_{Aj})}}, \mu > 0 \tag{3}$$

where μ is a positive scale parameter. Confirming the idea of utility maximization, alternative i is chosen out of all alternatives j, if the utility of i is higher than j. The probability that PNR user *A* chooses PNR lot i is

$$\begin{aligned} P_{Ai} &= P(V_i + \epsilon_i > V_j + \epsilon_j) \\ &= P(\epsilon_j < \epsilon_i \; + \; V_i \; - \; V_j) \end{aligned} \tag{4}$$

From equation 3 and 4, we obtain the following equation for the logit choice probability (Train, 2009):

$$P_{Ai} = (e^{\mu(V_{Ai})})/(\sum_{1}^{j} e^{\mu(V_{Aj})}) \tag{5}$$

We often assume the scaling parameter μ to be 1. Hence, equation 5 becomes:

$$P_{Ai} = (e^{V_{Ai}})/(\sum_{1}^{j} e^{V_{Aj}}) \tag{6}$$

The predicted probability of say, PNR lot 1 is given by

$$P(lot_1) = e^{V_1}/(e^{V_1} + e^{V_2} + e^{V_3} + ...e^{V_{20}}) \tag{7}$$

Predicted probabilities for all other PNR lots could be calculated similarly.The PTOD survey provided the PNR lot choices by all users, hence the market share is known. Likelihood is simply the probability of the chosen alternative, for example if the probability of person *A* choosing PNR lot 5 and person *B* choosing PNR lot 18 is known (i.e., the probability of jointly selecting lots 5 and 18 by two independent travellers), then given a set of coefficients on variables in the function V_i, the likelihood of two users making these choices is

$$Likelihood = [e^{V_5}/e^{V_1} + e^{V_2} + e^{V_3} + ...e^{V_{20}}]_{personA} * [e^{V_18}/e^{V_1} + e^{V_2} + e^{V_3} + ...e^{V_{20}}]_{personB} \quad (8)$$

Instead of using likelihood, it is a common practice to use log-likelihood as it makes calculation of coefficients easier; it makes the function concave. The log-likelihood of the whole sample is the product of individual log-likelihoods. The values of βs (the coefficients) that maximize the log-likelihood function are the estimated values of the βs. Biogeme software (Bierlaire, 2003) uses a conjugate gradient method of optimization and estimates these βs. The estimates of coefficients from the RUM model are as presented in table 3.2.

3.7 PNR lot choice model based on RRM

The regret minimization technique is based on the decision rule stating that travellers try to minimize their anticipated regret rather than maximize expected utility, where regret is the experience a person feels when a foregone alternative performs better than the chosen one (Chorus, 2010). The park-and-ride lot choice decision is a complex process. It involves travel in both auto and transit networks. Therefore, when choosing a PNR lot, attributes associated with the auto network and with the transit network are considered along with attributes related to the PNR network itself. Along with using a typical, fully-compensatory RUM model, which assumes that the loss of utility due to a certain attribute can always be compensated by strong performance of other attributes, I aimed to explore the semi-compensatory assumption using the RRM model specification. Semi-compensatory behaviour suggests that improvement in one attribute of an alternative does not always compensate for an equally large reduction in the performance of another attribute (Chorus, 2010). Also, another property of the RRM model is its ability to capture the “compromise effect”. According to Chorus (2010), the compromise effect means that alternatives with an ‘in–between’ performance on all attributes in relation to all other alternatives in the choice set are more likely to be chosen by choice-makers when compared to extreme alternatives that have very poor performance on some attributes and a very strong performance on others. Especially in the context of PNR lot choice, I anticipated that passengers would display behaviours based on this compromise effect. This is because the PNR choice process resembles a trade-off situation between the auto and transit legs of travel. Similar to any typical human choice behaviour in trade-off situations, it is reasonable to assume that an "in-between" alternative would have a better chance of being chosen. For instance, reducing the origin-to-PNR lot time alone may not be a priority for typical travellers. It is reasonable to assume that travellers aim to minimize their travel impedance for the overall trip, while making sure they will get a parking space in their chosen lot. In this choice process, reducing (increasing) the network distance from origin-to-PNR (which is dependent on the time from origin-to-PNR) could mean increasing (reducing) PNR-destination time. Therefore, the person could be likely to choose a compromised option with reasonable travel attributes in both the auto and transit legs. Therefore, I anticipated that a model specification that can capture such effects, such as RRM, might perform better in such a choice situation. As a result, RRM model specifications were also included in the discrete choice modelling in this research.

Independence from Irrelevant Alternative (IIA) issue in RRM

One of the properties of the RUM MNL logit model is that the ratio of any two alternatives is independent from the existence of another alternative, i.e., independence from irrelevant alternatives (IIA). It is due to this property that RUM MNL can provide unbiased estimates for the set of alternatives containing a chosen alternative and a random sample of non-chosen alternatives. However, RRM MNL does not have this desired property (Chorus, 2012). In the choice set, I had a chosen alternative and 19 randomly sampled non-chosen alternatives. With the knowledge of this potential caveat, and in order to test the credibility of RRM in this case, I performed an out-of-sample validation in which I calibrated the models (RUM and RRM) using 70 % of the data (table 3.2) and validated their performance in predicting the remaining 30 % of the data (table 3.3). I then compared the performance of the RRM model in relation to the RUM model using the validation data. While validation is an integral part of analysis, it does not resolve the IIA issue. However, a validated RRM model suggests that the model is credible despite IIA issues.

3.7.1 Model construction based on RRM

From section 3.6.1, the probability of choosing an alternative *i* by user A was given by equation 6. The underlying assumption of this probability expression is based on the decision rule that alternative *i* is chosen out of all alternatives *j* if and only if the overall utility of *i* is higher than that of *j*. Similarly, for the RRM model specification (Chorus, 2010), the overall regret is also composed of observed (deterministic) and unobserved (random) components:

$$R_i = \hat{R}_i + \epsilon_i \tag{9}$$

Where, i is any choice alternative. R_i is overall (or perceived) regret, $\hat{R}_i$ is the deterministic (or systematic) regret estimated by the modeller, and ϵ_i is the error term for unobserved regret. The regret can be evaluated for alternative i using:

$$\hat{R}_i = \sum_{j \neq i} \sum_m ln(1 + e^{[\beta_m(X_{jm} - X_{im})]}) \tag{10}$$

Where, X_{im} is the attribute value of attribute m of the alternative *i*, and β_m is the coefficient of attribute *m* to be calibrated.
Conceptually, deterministic regret could be calculated through a pairwise comparison of alternatives' attribute values, i.e. as the sum of all binary regrets that were obtained by comparing the considered alternative with each of the alternatives in the choice set. Further, Chorus (2010) pointed out that minimizing random regret is mathematically equivalent to maximizing the negative of random regret. The probability of choosing alternative *i* out of all alternatives j is equal to:

$$P_i = e^{(-\hat{R}_i)} / \sum_j e^{(-\hat{R}_j)} \tag{11}$$

3.8 Lot choice models based on RUM and RRM model

The estimates of coefficients based on the RRM model are presented in table 3.2. Four models are presented. Model 1 presents models for both RUM and RRM without the variable *free* and *cbdP*and 2 represents models with the inclusion of these two variables. The magnitude of one parameter relative to another parameter in the RUM model and in the RRM separately gives the relative importance (marginal rate of substitution) of the variables in the respective models. It is important to mention that my results confirmed that those rates of substitution were almost the same among the RUM and RRM models, whereas the coefficient values were different in the these models by a factor of about 10.

Table 3.2: Model estimation results from RUM and RRM models

	RUM				RRM			
	Model 1		Model 2		Model 1		Model 2	
	β	t	β	t	β	t	β	t
For	3.89	17.73	3.88	17.71	0.341	17.31	0.342	17.24
Invehtime	-0.041	-11.7	-0.041	-11.70	-0.004	12.82	0.00	-12.53
Netdist	-0.526	-36.04	-0.528	-35.83	-0.071	-26.61	-0.074	-25.47
Totfare	-0.879	-10.51	-0.939	-8.82	-0.058	-9.1	-0.042	-5.53
Train	1.50	19.75	1.50	19.73	0.134	18.06	0.136	18.15
Waittime	-0.03	-4.12	-0.03	-4.17	-0.003	-4.43	-0.003	-4.3
CbdP			-0.132	-0.96			0.058	4.31
Free			-0.029	-0.31			-0.006	-0.62
Observations	2575.000		2575.000		2575.000		2575.000	
Initial log-likelihood	-7714.011		-7714.011		-7714.011		-7714.011	
Final log-likelihood	-1701.908		-1700.573		-1685.604		-1675.906	
Adj.R^2 of the model	0.779		0.779		0.781		0.782	

Variables *free* and *cbdP* were found to be insignificant, but slightly improved the log-likelihood of the models, especially in the RRM case. This means that using the information on how a user perceives a lot near a freeway or a lot that lies in the CBD helps the model explain decisions, but from the travellers' point of view choosing a lot near a freeway or in the CBD is not much of a priority to them. Estimation results indicated that users preferred to use formal PNR lots (*for*) as well as lots that serve train stations (*train*). This may be because users feel that they have a greater chance of getting parking in lots with larger parking capacities or formal lots. Also, formal lots are usually more secure. Likewise, users seemed to be inclined toward lots served by train. This may be because this study was done for morning peak and many trips were work- or education-related and well served by the radially-oriented rail network in SEQ. From another point of view, it is also possible that PNR lots that serve a train station, may have more bus stops located nearby than other PNR lots. For the purpose of increasing accessibility to mass transit, train stations are mostly complemented by nearby bus stops and a user may find such PNR lots more attractive as they offer more transit options.

As anticipated, in-vehicle travel time (*invehtime*) had a negative coefficient. Users do not prefer PNR lots which have higher in-vehicle travel times in transit. Also, users seem to choose PNR lots that can minimize their wait time. The *netdist* (proxy of auto cost) and transit fare (*totfare*) variables were negative, showing that users want to minimize their travel cost in both auto and transit networks. Also, considering Model 1, the value of time (VOT) in transit for the RUM model is AUD 2.8 per hour while for RRM it is AUD 4.1 per hour, which is relatively lower than the average wage. According to Guevara (2017), the value of in-vehicle travel time savings in the case of mode choice models is lesser for public transport than private auto. Though my models do not give VOT in auto and no comparisons can be made, the relatively lower VOT for transit can be understood in this way.

These results provide some interesting insights. Users want to use PNR lots where they can minimize their transit travel time (suggested by *invehtime*). Also, users want to choose a PNR lot where their auto and transit travel costs are reduced (as suggested by the coefficients of *netdist* and *totfare*). By analysing travel time in two different networks for PNR users, it can be understood that, although users want to minimize travel time and travel expenses, they place different values on travel time in auto and transit and make a trade-off/compromise between these travel times. This understanding is based on the use of a time-dependent travel time as the impedance between the origin to the PNR lot and the PNR lot to the destination. In this way, developed models explain the choice of PNR lots based on time-dependent travel times.

It is important to note that direct comparison between results from the two models is not meaningful. Where RUM parameters signify the contribution of an attribute to an alternative's utility, RRM parameters signify potential contribution of an attribute to the regret associated with an alternative (Hensher et al., 2013). Further, according to Chorus (2012), even though one cannot compare the parameters from RUM and RRM models directly, one can compare the magnitude of one parameter relative to another within the same model to understand the relative importance of the attributes. In both of my models, the estimated parameters were intuitive and were statistically significant at a 95% confidence interval, except *cbdP* and *freeway*. Finally, the difference between the final log-likelihood values in the RRM and RUM models showed a larger improvement for RRM models. Hence, I can state that the RRM model fits the data better than the RUM model for the lot choice model.

Validation

Using the estimated models from the 70% sample, data validation was done using the remaining 30% of data for both the RUM and RRM models. The total number of observations is 3679, 2575 of which were used to estimate the models and 1104 were used for validation. Validation was done ten times. At first, a randomly selected 10% sample of the data was taken from the validation sample and validation was performed. The sample was then replaced. Again, a 10% random sample was drawn from the validation set and validation was performed. I repeated this procedure 10 times and the average of the 10 runs is as presented in 3.3.

Table 3.3: Predictive ability out-of-sample for RUM and RRM models

	RUM		RRM	
	Model 1	Model 2	Model 1	Model 2
Initial log-likelihood	−3310.28	−3310.28	−3310.287	−3310.28
Final log-likelihood	−1482.68	−1473.73	−1411.22	−1382.29
Adjusted $R^(2)$ of the model	0.552	0.555	0.574	0.582

The comparison of model performance in the validation dataset is shown in table 3.3, which suggests that RRM model is credible despite IIA issues.

Test of fit of the models

We use the Ben-Akiva and Swait test (Ben-Akiva and Swait, 1986) to compare RUM and RRM models, because RRM cannot be specified as a special case of RUM or vice versa when there are three or more alternatives in the choice set (Hensher et al., 2016). In case where the two models have the same number of parameters, the probability is calculated as:

$p = NormSDistr(-\surd[(2{\cdot}N{\cdot}ln(J){\cdot}(LL(B) - LL(A))/LL(0)])$

where, NormSDistr(x) is the probability that draw from a standard normal is less than x, N is the number of observations, and J is the number of alternatives in the choice set. In our case, $(p = NormSDistr(-\surd(2 * 2575 * ln(20) * ((-1700.573 + 1675.906)/(-7714.011))]$, where p of $1.7 * 10^{-12}$ value is much lower than 0.05 suggesting that the difference between the two models is significant in favour of model A (i.e., RRM) at a 5% significance level.

3.9 PNR lot choice model for formal PNR lots only

The analyses performed so far, both for RUM and RRM, included formal as well as informal PNR lots. For studying the nature of formal PNR lots specifically, I constructed a lot choice model using *utilization* as one of the variables. The nature of formal and informal parking is different; for informal locations like on-street parking it is difficult to quantify capacity, and thus *utilization* cannot be correctly known. Thus, as a further step, I included only observations in which the chosen PNR lot was a formal PNR lot. The excluded observations comprised 3 %

of the original observations. After excluding informal PNR locations, there remained 116 PNR lots in the choice set. Not all 116 lots were seen as a choice alternative by all travellers. Thus, similar to the process mentioned in 3.4, among 116 PNR lots, those where the shortest travel time from the origin to the PNR lot (by automobile) was greater than the shortest travel time from the origin to the destination (by automobile) were excluded from the possible choice set. From the remaining PNR lots, 19 PNR lots were selected randomly. 19 lots along with one chosen lot made a choice set of 20 PNR lots. The choice set generation process in the case of formal PNR lots is in line with that mentioned in 3.4.

3.9.1 Endogeneity in PNR lot choice model

I studied a variety of attributes that could possibly influence travellers' PNR lot choice behavior as mentioned in table 3.1. The resulting lot choice model (table 3.6, which has a downward bias and an insignificant utilization parameter) tempted me to perceive that travellers do not place a high value on utilization of PNR lots. However, it is not possible for a researcher to identify all attributes considered by travellers, and I might have missed some important attributes. For example, travellers might have considered 'availability of informal off street parking', 'number of access points for the lots', etc. That information was not available to me but does affect the utilization of lots. These omitted attributes, which could have ended up in the error term, may have been correlated with the utilization, causing endogeneity in the model and resulting in a downwardly biased coefficient on utilization.

Models with endogeneity give incorrect parameter estimates and can be misleading. For example, PNR facility providers, based on the PNR lot choice model, may consider utilization as an insignificant attribute and plan to locate future PNR lots in areas where utilization is low. In this case the model is useless in representing actual behaviour, and planners may need to pay a high cost for this misspecification in the model. As a result, correction of endogeneity is imperative. The part of endogeneity caused by omitted attributes and the simultaneity of demand estimates is corrected.

3.9.2 Correction of endogeneity in the PNR lot choice model

As discussed in section 2.4.2, I used CF method for the correction of endogeneity. The CF method has its limitations too; one criticism is its dependence on IVs which are often very hard to find, and even when they are found they may suffer from the weak instrument problem (Guevara and Polanco, 2016; Guevara, 2015; Fernandez-Antolin et al., 2014). Recently, a multiple indicator solution (MIS) method was proposed (Guevara and Polanco, 2016), which does not require IVs. However, the MIS method addresses only those types of endogeneity that arise from the omission of an attribute that could be measured by an indicator, whereas the CF method addresses the correlation with the error term in a broader sense, regardless of the source of endogeneity (Guevara and Polanco, 2016). My aim is to correct endogeneity caused by omitted attributes and the simultaneity of demand estimates, and thus I chose the CF method.

3.9.3 Application of control function for correcting endogeneity

The first step in applying the CF method was to find the relevant IVs. I constructed two instruments (IV1 and IV2) for each observation. The first and second instruments were the average utilization of PNR lots located within 2.5 to 5km and the average utilization of PNR lots located within 5 to 7.5 km respectively from the concerned PNR lot. I excluded PNR lots within 2.5 km to avoid the ‘reflection bias’ problem and limited the distance to 7.5 km to avoid the ‘weak instrument problem’ as described in detail by Guevara (2010).
Correlation among IVs is shown in table 3.4.

Table 3.4: Correlation among utilization and instrumental variables

	Utilization	IV1	IV2	Log of utilization
Utilization	1	0.36	0.18	0.67
IV1	0.38	1	0.48	0.54
IV2	0.18	0.48	1	0.31
Log of utilization	0.67	0.54	0.31	1

The correlation was relatively low for both IV1 and IV2; IV1 was more correlated to utilization and the logarithm of utilization than IV2. As mentioned in Guevara (2010), the thresholds defined to construct IVs are not hard constraints and a slight change in threshold will not impact the estimates of the model parameters qualitatively . I continued with those values of IV1 and IV2. In applying the 2-stage CF (2SCF) method, the first stage is the regression of utilization on the instruments IV1 and IV2. The results of this auxiliary regression are shown in table 3.5. Model 1 represents the model where IVs are regressed against utilization and model 2 where they were regressed against the logarithm of utilization.

Table 3.5: Results from auxiliary regression

	Model 1		Model 2	
	Coefficient	Standard Error	Coefficient	Standard Error
Constant	-1.45	0.89	-0.76	0.16
IV1	2.51	0.59	0.65	0.1
IV2	-0.02	0.59	0.09	0.1
R^2 of the model	0.14		0.29	
Adjusted R^2 of the model	0.13		0.28	

One possible reason for the low explanatory power could be the nature of the dependent variables used in the regression; both were spatial variables and as the distance in space increased, the correlation with the dependent variable could increase too. Model 2, however, had better explanatory power than model 1. There is no existing study on the choice of good/weak instruments in a discrete choice framework (Guevara, 2010; Fernandez-Antolin et al., 2014). Hence, I continued with the regression models with adjusted R^2 value of 0.13 and 0.28 to the second stage of 2SCF. The residual obtained from the regression stage was used as an additional variable in the utility. The estimation results from both models i.e. with correction for endogeneity and without correction for endogeneity for the two sets of models (with utilization and log of utilization) are presented in table 3.6.

Table 3.6: Estimation results for the model with and without correction of endogeneity

	Without correction				With correction			
	Model 1		Model 2		Model 1		Model 2	
	β	t-test	β	t-test	β	t-test	β	t-test
Capacity	<0.01	9.56	<0.01	12.33	<0.01	10.36	<0.01	12.86
Invehtime	-0.13	-27.01	-0.12	-24.98	-0.12	-24.98	-0.11	-22.79
Optime	-0.45	-49.25	-0.44	-48.68	-0.45	-48.71	-0.44	-48.32
Totfare	0.33	4.04	0.31	3.77	0.33	3.95	0.29	3.45
Waittime	-0.03	-5.03	-0.03	-4.54	-0.03	-4.49	-0.03	-3.94
Walktime	-0.08	-6.86	-0.07	-5.86	-0.08	-7.07	-0.07	-6.00
Utl	0.01	0.99			0.23	8.37		
Log of utl			0.50	10.71			1.35	12.15
Residual					-0.24	-8.57	-0.90	-8.51
Observations	3364		3364		3364		3364	
Initial log-likelihood	$-10\,077.64$		$-10\,077.64$		$-10\,077.64$		$-10\,077.64$	
Final log-likelihood	-3967.42		-3909.23		-3930.53		-3872.69	
Adj.R^2 of the model	0.61		0.61		0.61		0.62	

All the coefficients showed expected signs in all models. As anticipated, in-vehicle travel time (*invehtime*), wait time, (*waittime*) and walk time (*walktime*) had negative coefficients. Travellers do not prefer PNR lots which have higher in-vehicle travel time, walking time, or waiting time within the transit network. Also, travellers want to minimize their auto travel (optime) from the origin to the PNR lot. The transit fare (fare) variable was positive, a result which may seem counterintuitive. However, in SEQ, fares are based on the number of zones travelled, and a positive coefficient on fare suggested that travellers choose PNR for longer transit journeys. The coefficient of utilization (utl) in model 1 was low in magnitude and also was not statistically significant at a 95% confidence level. Further, the variable that represented the capacity of a PNR lot (capacity) came out to be positive (0.001) and significant, suggesting travellers prefer lots with higher capacities. In model 2, the variable representing logarithm of utilization (log of utl) was positive and significant, suggesting that travellers are inclined to use PNR lots which have greater utilization. However, the insignificant coefficient of the utilization (utl) variable was not consistent with common sense and I suspected that some important variables were not addressed in the model, causing endogeneity. After applying the 2SCF correction to the model, the estimation results were as shown in the fourth and fifth columns of table 3.6. The coefficient of the auxiliary variable (which was the residual from the first stage of 2SCF) was statistically significant at a 95 % confidence level for both models 1 and 2. The negative sign of the residual suggests that the omitted attributes were negatively correlated with utilization, and the parameter estimation of utilization (before correction) was capturing both the positive impact of utilization and the negative impact of the omitted attributes, which had resulted in a less positive estimated value than the true one. This was the reason for the downwardly biased value of utilization before correction. In studies like residential location choice (Guevara and Ben-Akiva, 2006), the omitted attribute was found to have a positive correlation with the endogenous attribute, which was understandable because some features like comfort might positively influence the traveller in choosing a particular alternative. In this case, the omitted attribute was negatively correlated with utilization; this is due to the nature of the endogenous attribute in the PNR lot choice model. After correction of endogeneity, the magnitudes of other coefficients were also slightly changed, but the overall marginal utilities remain similar. The difference in final log-likelihood between the lot choice model and the corrected lot choice model showed a higher value for the corrected model. Also, using the likelihood ratio test $(-2(-3967.420 + 3930.529) = 73.782 > \chi_{21,95\%}$ for model 1) and $(-2(-3909.230 + 3872.689) = 73.082 > \chi_{21,95\%}$ for model 2), the residual improved the model as its coefficient was significantly different from zero at a 95% confidence interval.

The corrected models fit the data better than the initial PNR lot choice model. The best model in terms of the statistical fit was the one which was corrected for endogeneity and that used the variable with the logarithm of utilization.

Given the skewed nature of the values of utilization, it was tempting to enter the variable in its logarithmic form; the model with logarithmic form also performed better. However, if I had used the variable with log of utilization in the first place, instead of analyzing the underlying nature of the utilization of PNR lots, the endogenous nature of the utilization variable might have been overlooked. Overall, I obtained some interesting insights from both. Travellers want to use PNR for longer distances, as long as it is time-efficient in both auto and public transport (suggested by the coefficient of transit fare *fare*, auto travel time variables *optime* and the transit time variable *invehtime*).

Further, I applied the likelihood ratio version of the direct test presented in Guevara (2010) to check for the validity of the IVs. This test was straightforward. Utility contained not only the residual from the first stage of 2SCF, but also one of the IVs. The main idea was that an IV is correct if it is independent from utility. If the coefficient of the included IV was not statistically significant, the null hypothesis that both instruments were valid would be accepted. I included the first IV, the average utilization of PNR lots within 2.5 to 5 km, in the utility for the model 1 along with the corresponding residual. The estimation result is presented in table 3.7.

Table 3.7: Direct test with Instrumental variable

	β	t-test
Capacity	0.00	10.35
Invehtime	-0.12	-24.75
Optime	-0.45	-48.59
Totfare	0.32	3.93
Waittime	-0.03	-4.47
Walktime	-0.08	-7.05
Utl	-0.45	-0.22
Residual	0.45	0.22
IV1	1.72	0.33
Observations	3364	
Initial log-likelihood	$-10\,077.64$	
Final log-likelihood	3928.40	
Adj.R^2 of the model	0.61	

The coefficient of IV1 was not statistically significant (with 95% confidence) and thus the null hypothesis that both instruments were valid is accepted. Another part of the direct test was a likelihood ratio test. It was noted that, $(-2(-3930.539 + 3928.397) = 4.284 < \chi_{21,95\%}$), the likelihood ratio test did not agree with the validity of the instruments with 95% confidence. It was interesting that in the remainder of the three cases, (i) when IV2 was included in model 1, or (ii) when IV1 was included in model 2, or (iii) when IV2 was included in model 2, the IVs turned out to be statistically significant and slightly improved the model fit as well (not presented here for the sake of brevity). Though I performed the direct test, and IV1 looked like a valid IV for model 1, tests on other cases gave inconclusive results.

3.10 Conclusions

In this chapter I developed a PNR lot choice model for PNR lots in SEQ using two decision criteria: RUM and RRM. Also, I found that *utilization* of PNR is endogenous in nature. I then, corrected endogeneity using the CF method. The contributions of this chapter are three-fold.

First, the developed model explained the PNR lot choice behaviour of users using time-dependent travel time variables for trip legs by public transport. A PNR trip is divided into two parts, auto leg and transit leg, and travel times associated with each leg were used as variables in the model. This explained choice behaviour for varying travel times in the auto network and in the transit network during different time periods within the morning peak. Interesting insights into PNR lot choice behaviour were found, which improved the state of knowledge on PNR utilization in a real case study.

Second, this chapter is the first application of an RRM model to PNR lot choice and also the first comparison between the two models in this context. Through the acknowledgement of a trade-off relation behind the PNR choice mechanism in this chapter, it was anticipated that an RRM model can capture this effect. Results indicated that both RUM and RRM models gave reasonable coefficients with very close marginal rates of substitution. The out-of-sample validation results also suggested that despite the IIA issue, the RRM model still performed better (in terms of log likelihood) than the best RUM model. This study gives an additional viewpoint to policy makers in understanding PNR users' lot choice behaviour by developing PNR lot choice models based on two different decision frameworks.

Third, I identified the presence of endogeneity in a PNR lot choice model, which is the first of its kind. In the context of discrete choice models, endogeneity has been studied mainly for attributes like price and travel time. I found that the utilization of PNR lots is downwardly biased when not corrected for endogeneity. I applied the CF method for the correction of endogeneity; the corrected model gives better parameter estimations than the uncorrected one and reflects consistent values of all attributes. Also, I used the residual obtained from regression with an adjusted R^2 of 0.13 which was efficient in correcting the endogeneity. This R^2 value was much lower than the suggested minimum R^2 value of 0.4 to avoid a weak instrument problem for linear models (Hahn and Hausman, 2003; Guevara, 2010). In the absence of literature on weak instruments in the context of discrete choice models, this finding gives another viewpoint. Further, by testing log of *utilization* as a variable in the lot choice model to get a model that looked to be free from endogeneity, I highlighted the importance of analyzing the fundamental nature of variables before applying any non-linear transformations.

3.11 Further direction from this chapter

From the PNR lot choice model I developed on formal PNR lots, I found that the *utilization* variable of PNR lots is endogenous in nature. Also, from chapter 2 there is an understanding that the attractiveness of PNR could be dependent on the time of day. Hence, to understand more about the endogenous nature of the *utilization* variable, I aimed to study time-dependent *utilization* at PNR lots. To do this, I conducted a survey at PNR lots in SEQ; chapter 4 presents the survey I conducted and the results obtained from it.

4 Dynamic utilization of PNR lots

4.1 Introduction

In chapter 3, I presented the choice model for PNR facilities and the issue of endogeneity caused by *utilization*, which requires further examination. The literature review 2 and the practice highlighted that PNR *utilisation* depends on the time-of-day. Thus, this chapter is about the exploration of the *utilization* of PNR lots as a function of time of day. To understand the *utilization*, I collected data on twenty PNR lots in SEQ. Section 4.2 presents the details of the PNR survey procedure which is then followed by section 4.3, where I present the preliminary results from the PNR survey. Section 4.4 shows the analysis done in understanding the dynamic utilization of PNR lots. In this section, the results obtained by modelling the arrival of cars at parking spaces using the discrete-time hazard method are presented. The conclusions drawn from this chapter are included in section 4.6. Finally, the implications of the results from this chapter for PNR policy are outlined in section 4.7.

4.2 Survey at PNR lots in SEQ

4.2.1 Management of PNR survey

In August 2015, a PNR lot survey, combined with a questionnaire survey, was performed at 20 PNR locations in SEQ. August was chosen as it is commonly used as one of the appropriate months for transport data collection due to the absence of school holidays. The survey was conducted over two weeks on weekdays only, and each lot was surveyed only once. The framework of the survey is as shown in figure 4.1.

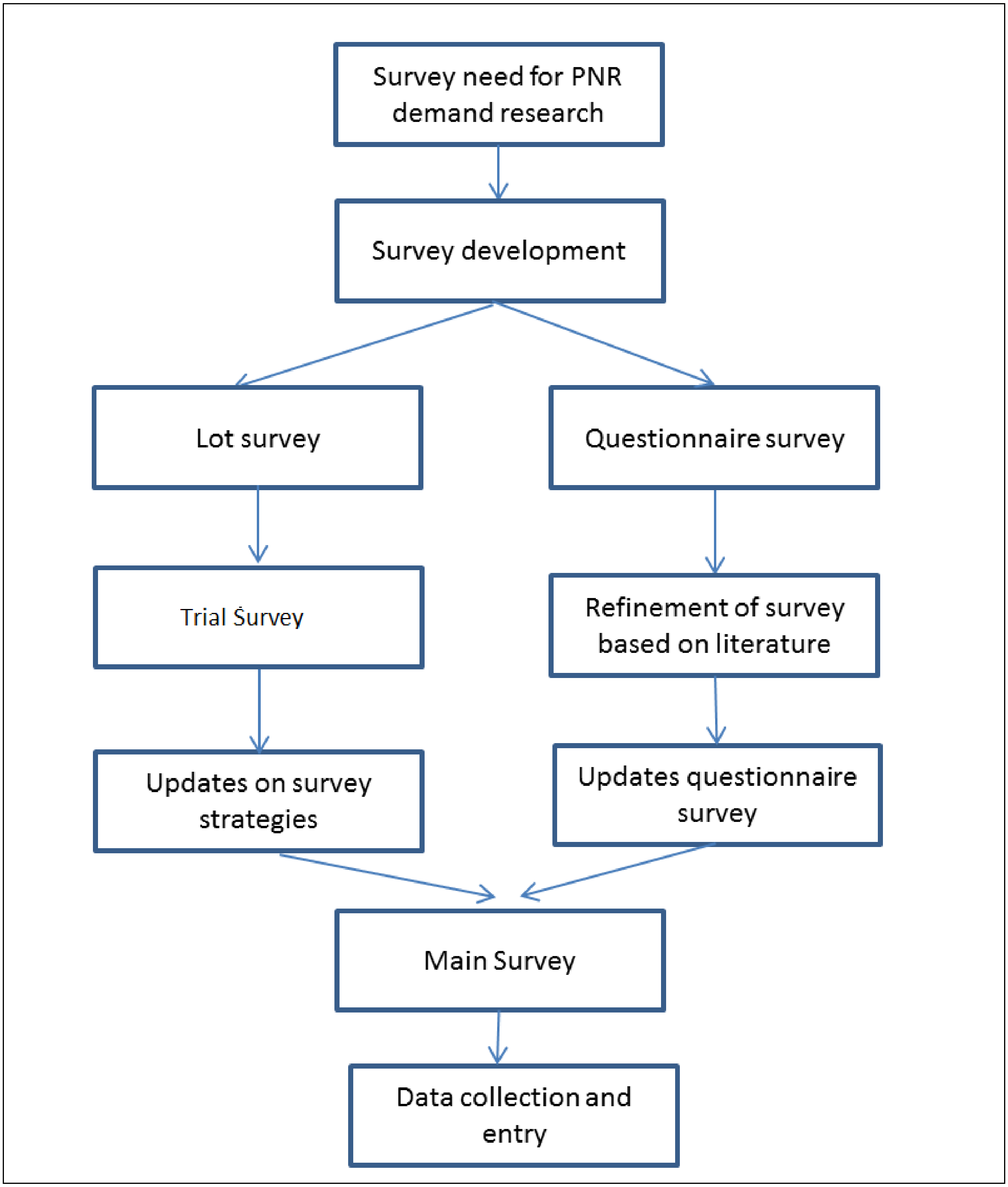

Figure 4.1: Survey Framework

There were two types of surveys done at the PNR lots. One was the PNR lot survey, which is the key survey; the other was a questionnaire survey which aimed to capture travellers' perceptions of the amenities of PNR lots. The data from the latter survey is not used in this thesis. For both surveys, field/survey assistants were hired and trained. An advertisement was published in the University of Queensland's job website calling for the position of casual PNR lot field/survey assistants. Among all applications, 23 people were chosen based on their time availability and their choice of PNR lot. 3-6 people were assigned a PNR lot depending on the capacity and the number of access points available in the lot. In every lot, a team leader, a questionnaire distributor/collector, and data collectors were assigned. Team leaders had freedom to reassign the task of data collection and questionnaire distribution as per need. Once, the arrival of cars became steady or the capacity was reached, the team leader would call the survey off. Before conducting the actual survey, a training session was conducted where field/survey assistants were introduced to the concepts of PNR and educated about the survey. A trial survey was done on the following morning, at the parking lots of The University of Queensland. In the pilot survey, the arrival time of cars approaching the parking lot were recorded. Feedback from the trial survey was carefully considered to update the real survey.

To select PNR lots for the survey, those lots which were seen to be more heavily utilized in the TMR PNR study (Government, 2014) were selected, with representation from all major radial routes into the CBD, representing different train lines and busways. Figure 4.2 presents the twenty PNR lots that were surveyed in SEQ.

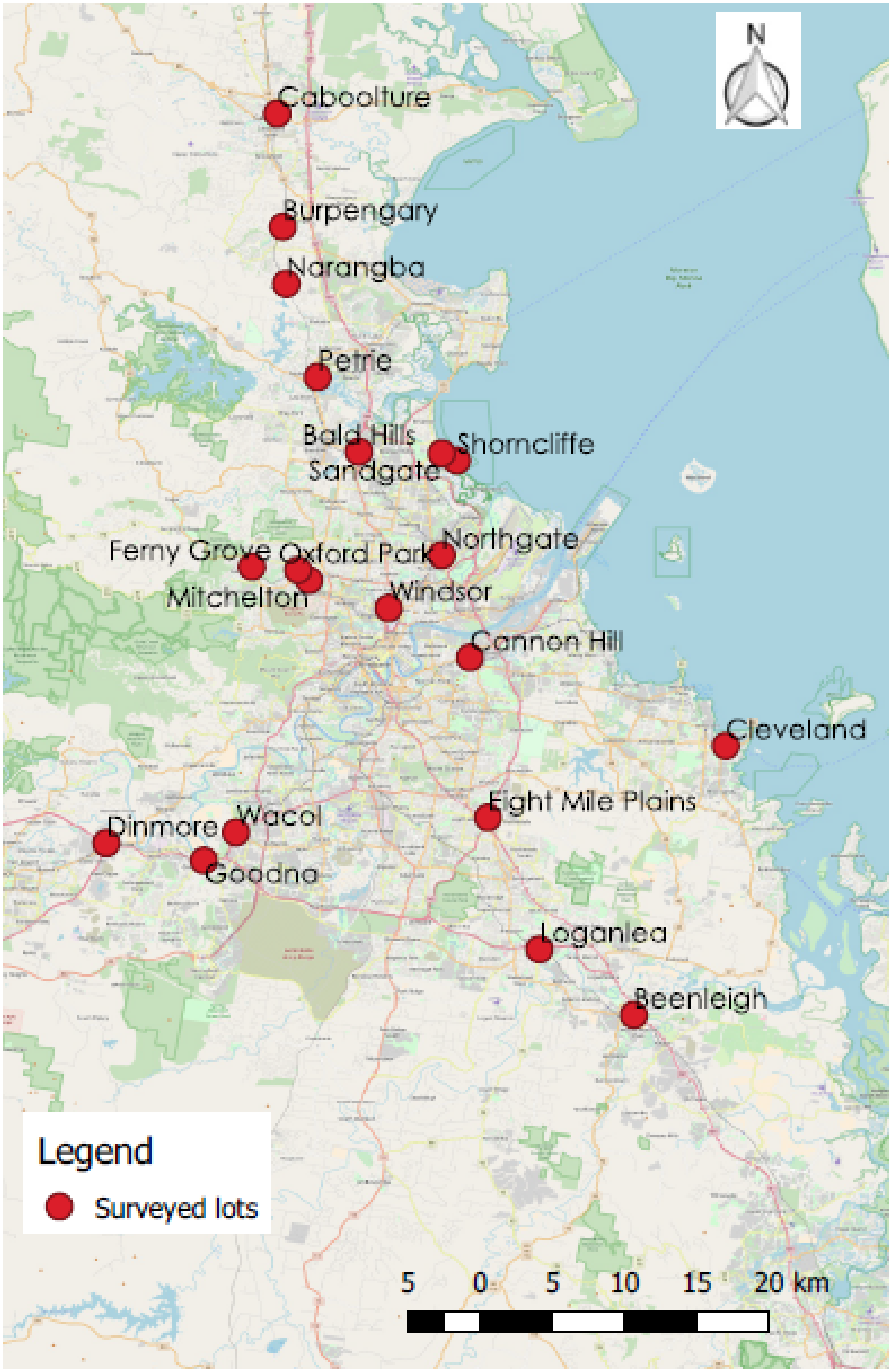

Figure 4.2: PNR lots surveyed

Data collected in the survey

The main purpose of data collection was to understand the *utilization* of PNR lots.The data collected was primarily concerned with the arrival of cars in PNR lots. The survey in each lot started 5-10 min prior to the arrival of the first public transport service at/near the lot and ended either when the lot was completely full or there were no further cars coming in for a significantly long time. Field/survey assistantss stood near the entrances of the car park and recorded the number of cars entering the lots and their arrival times. The first four digits of the license plate of each incoming car was recorded, in order to avoid double counting by another data collector. The use of the first four digits instead of the whole license plate number was to avoid unique identification and to maintain the privacy of PNR users. The total number of car spaces occupied and the total spaces available in the lot were recorded at the end of the survey. These data give the value of utilization of the PNR lot; utilization is the ratio of usage of PNR lot (cars) to the capacity of the lot at a given time. Also, audit surveys were done at each lot, on the same day, independently from the field/survey assistants' effort for quality control purposes.

Apart from the car arrival data, the area within the PNR lot where cars were parked was also recorded. This relative position of car park space within a parking lot was studied previously by Caicedo et al. (2006); Lai and Shalaby (2007); Tsang et al. (2005) as mentioned in chapter 2. I classified each parking space in the lot as being in a specific zone, based on the walking time from the parking space to the public transport platform; the categorization of zones is as shown in 4.1.

Table 4.1: Zoning in PNR lot

Walk time (minutes)	Zone
0-2	A
2-5	B
5-8	C
8-10	D
10-12	E
12-15	F
15+	G

4.3 Preliminary results from the PNR survey

The end time of the PNR survey was the time when the arrival of cars to the lot ended, either because the lot was full or because the lot reached more sporadic car arrivals (though some spaces were still remaining). When I analyzed the ending time of the survey for all PNRs, as shown in figure 4.3, I found that it does not necessarily follow the pattern seen in Nurul Habib et al. (2012); notably, PNR lots in SEQ do not get filled in order of increasing distance from the CBD.

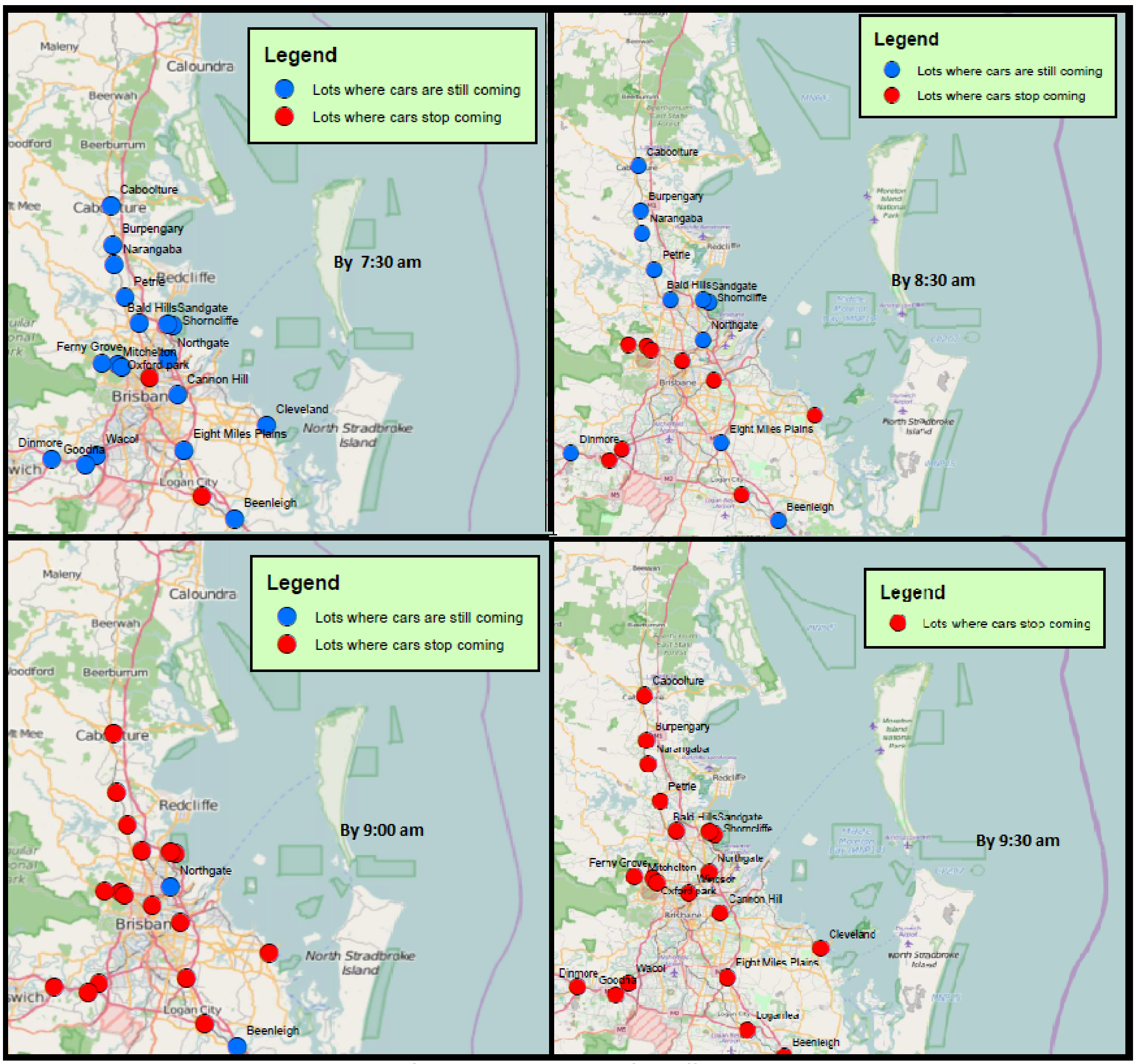

Figure 4.3: Survey closing time for different PNR lots

From all 20 PNR lots, it is understood that car arrivals increase and reach a peak during the period when there is a high frequency of public transport service. For example, Eight Mile Plains (EMP) is one of the largest PNR lots, and it has no train service. The car arrivals and the public transport service at EMP are shown in figure 4.4; where the y-axes represent the number of car arrivals and the number of public transport services. The public transport service schedule was downloaded for the survey day from the Translink website (Translink, 2016).

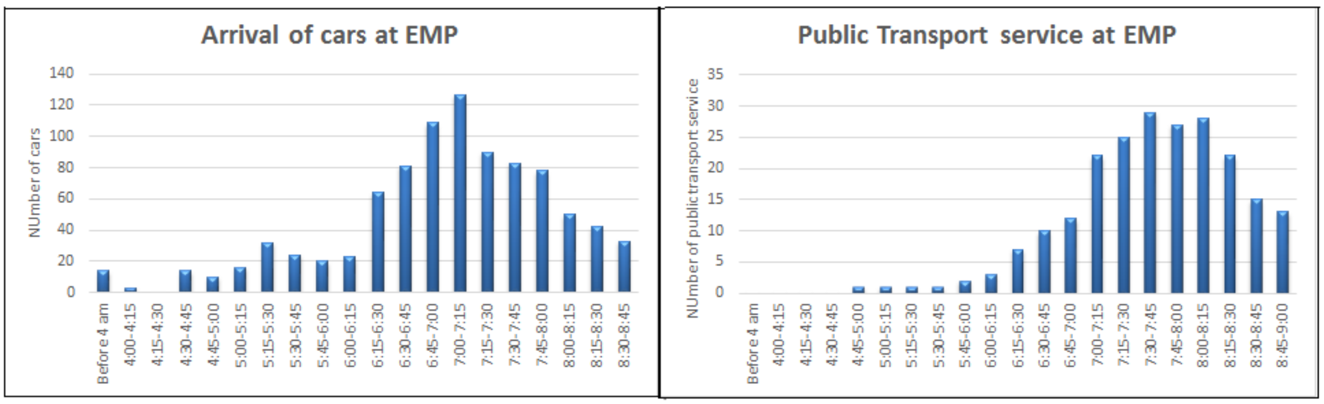

Figure 4.4: Car arrival pattern and the public transport service at EMP

Studying both car arrivals and public transport services together, it became clear that car arrivals in the PNR lot had a strong relationship with public transport services. The car arrivals reached their peak during 7:00 am to 7:15 am. Likewise, there was a public transport service peak during 7:30 to 7:45 am. The difference of 15 minutes between the two peaks was as expected; PNR users came some minutes earlier to the PNR lots than the arrival time of the public transport service they intended to catch. Another possibility for this strong relation could also be the response of PT service providers for commuting /education pattern seen for the respective time-periods.

Though the survey started 5-10 minutes prior to the first public transport service, in some of the PNR lots cars were found to be already parked before the commencement of the PNR survey. I simply treat such parking as parking before 4 am. They can be seen in figure 4.4.

Among the surveyed lots, Loganlea was filled by as early as 7:15 am, whereas Beenleigh filled up later, by 9:30 am. The car accumulation patterns of Loganlea and Beenleigh are shown in figure 4.5.

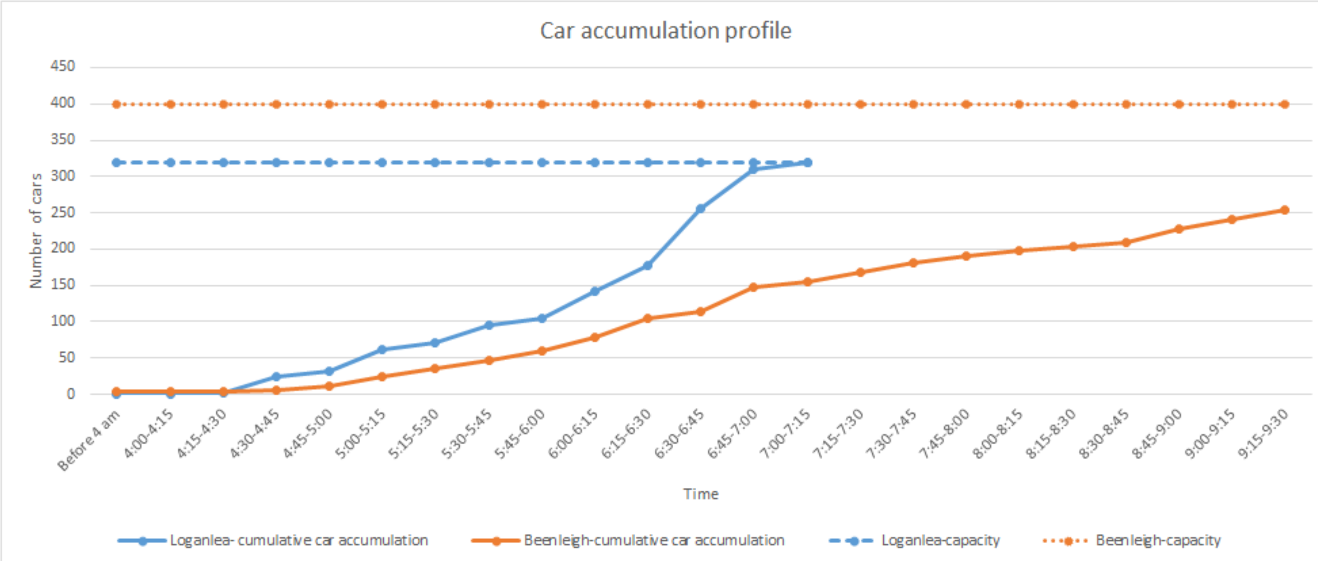

Figure 4.5: Car accumulation at Loganlea and Beenleigh

The Loganlea and Beenleigh PNR lots are both situated on the same train line. From 4 am to 10 am, Loganlea PNR lot served 68 public transport services, whereas Beenleigh which served 82 public transport services. From this I observed that users did not necessarily use a PNR lot only for the frequency of public transport services. The Loganlea station is closer to the CBD than Beenleigh.

The smallest lot in the survey was Windsor with a capacity of 63 cars, and the largest lot was Ferny Grove with the capacity of 931 cars. The accumulation profile of Windsor and Ferny Grove are shown in figure 4.6.

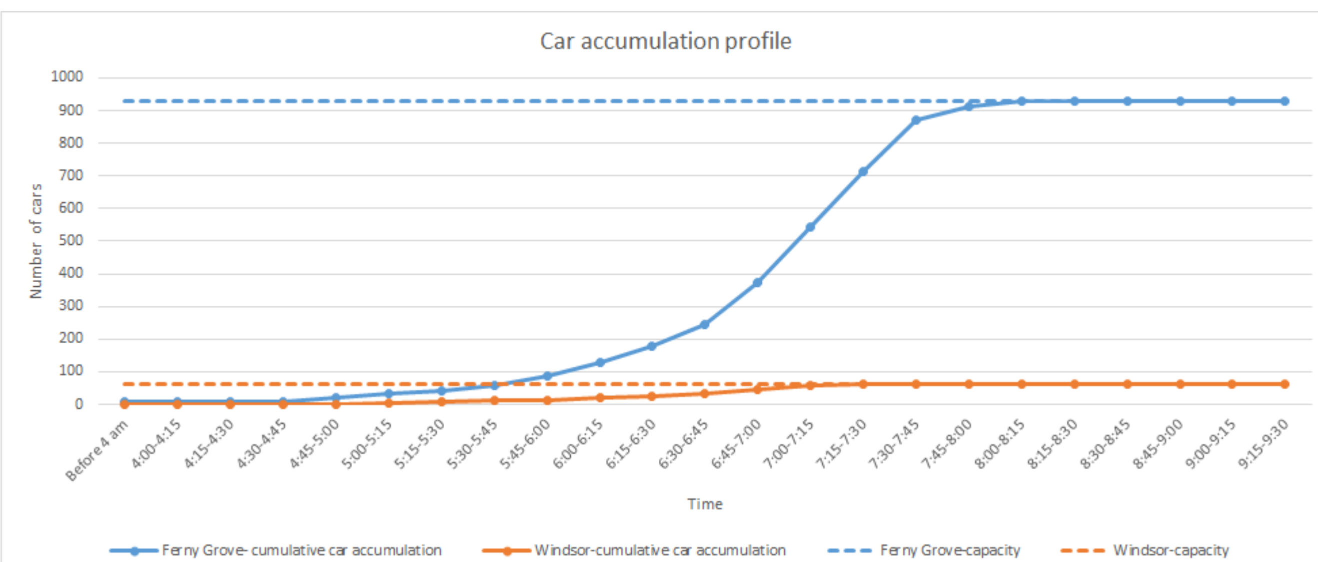

Figure 4.6: Car accumulation at Windsor and Ferny Grove

Windsor filled up by 7:30 am and Ferny Grove filled by 8:30 am. Windsor is the closest lot to the CBD among all surveyed lots. Further, the utilization of some other lots are presented in figure 4.7; Caboolture and Dinmore were two PNR lots that did not reach capacity in the am peak. Both lots are fairly large; Caboolture has a total capacity of 675 spaces and Dinmore has 490 spaces. The remainder of the lots were utilized at 100 percent by the end of the AM peak period.

Figure 4.7 presents the utilization of some of the PNR lots surveyed. It is observed that the accumulation profile of the PNR lots are dependent on time, however the rate of filling up of PNR lots differ widely and it is interesting to explore the determinants of the filling up of PNR spaces.

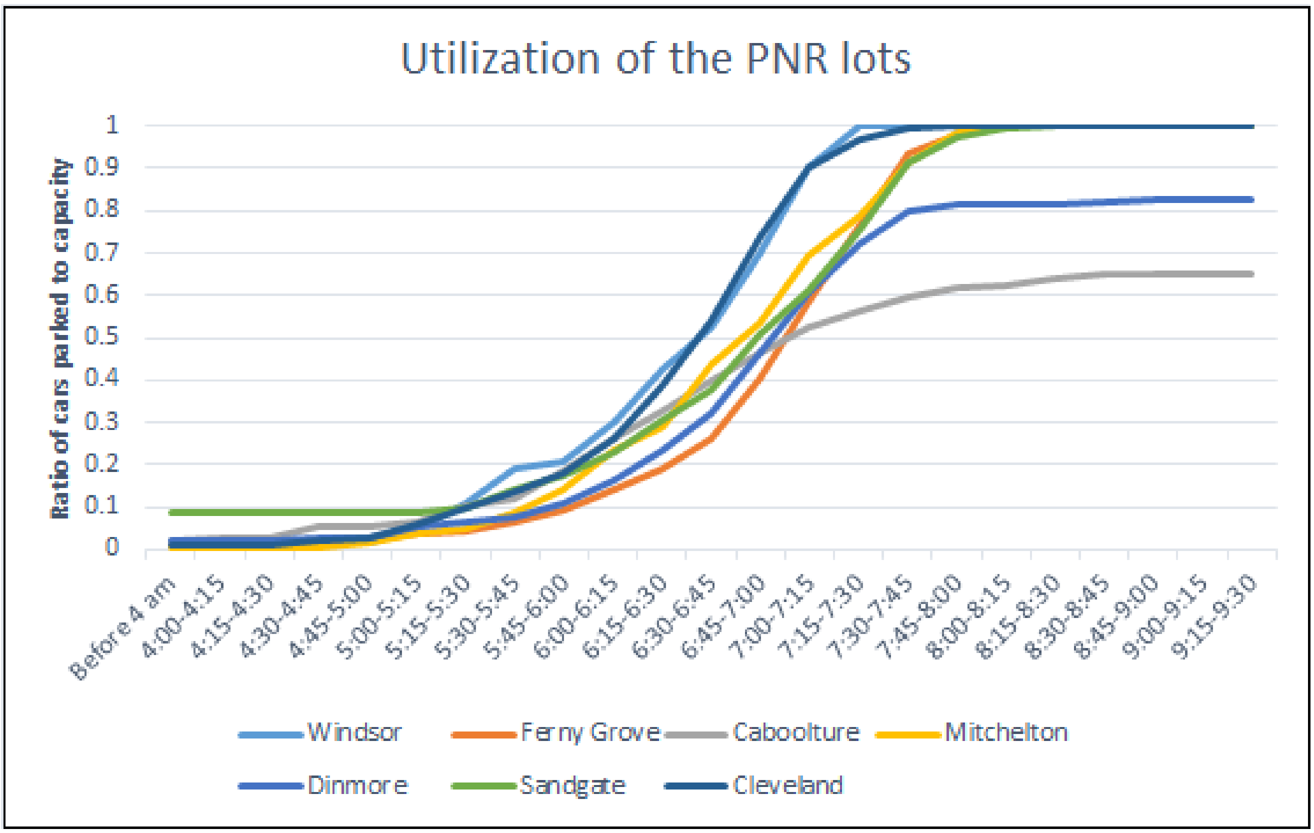

Figure 4.7: Temporal utilization of some PNR lots

KNR and informal parking

Kiss-and-ride (KNR) behaviour and informal parking were recorded in lots whenever and wherever possible. Though ample amounts of data were collected for those behaviours on many of the PNR lots, they were not exhaustive. While field/survey assistants recorded a number of KNR users (drop offs) within the vicinity of the PNR lot, many people could have been dropped off some distance away from the lot. Field/survey assistants were spread out across entrances, but there were no field/survey assistants at entrances with footpaths only; people who were dropped off or walked could not be easily recorded. Similarly, field/survey assistants made a note of informal parking in the adjacent streets, but informal parking could have occurred on distant streets too, making it impossible for field/survey assistants to account for all instances. KNR information was available in some lots, EMP being one of them. At EMP, it is possible to drop off people just in front of the gate leading to the platform. For EMP, recorded KNR trips could/not be the total KNR trips that happened at EMP. Though not complete, the collected data on KNR provided some sense of their demand. KNR trips recorded at EMP are presented in figure 4.8.

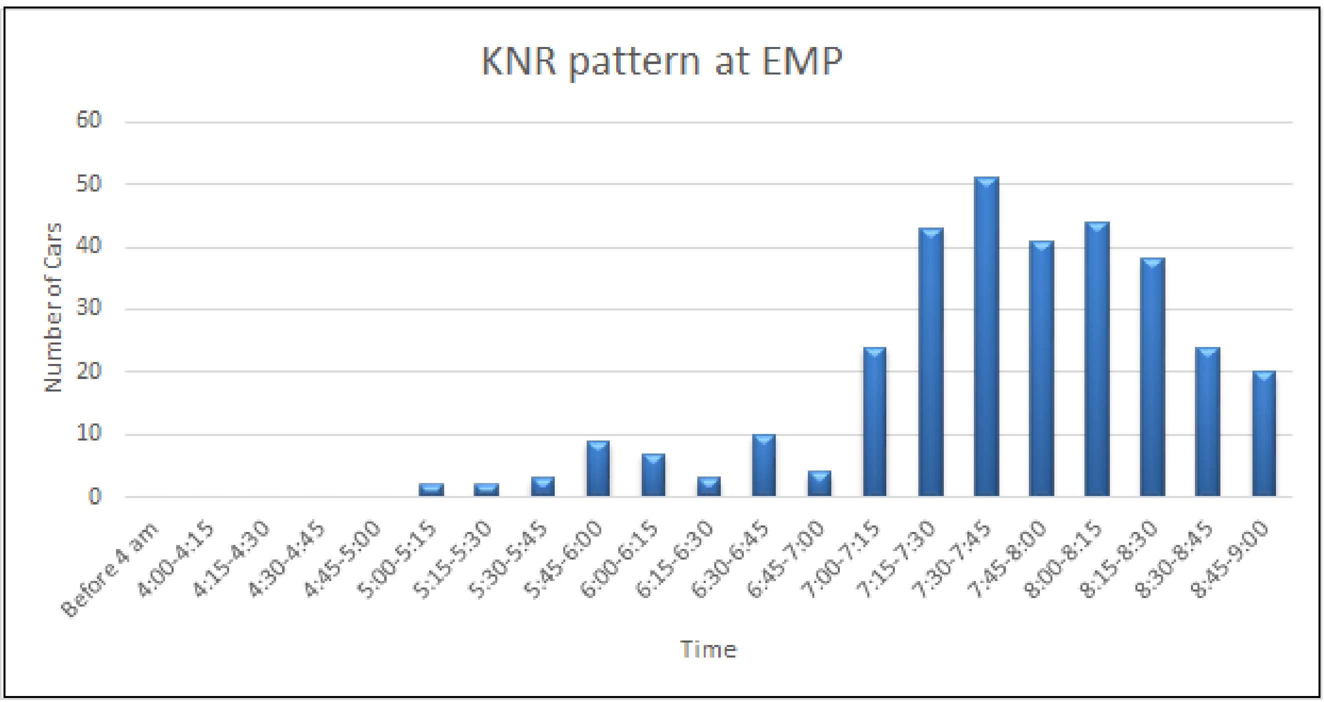

Figure 4.8: Number of KNR trips at EMP

KNR trips reached their peak at 7:30 am to 7:45 am, which was a little later than the time when the car arrivals in the PNR lot reached their peak, as shown in figure 4.4, i.e., 7:00 am to 7:15 am. The results complemented each other; people parking at PNR lots were likely to come some minutes prior to their intended public transport service, when compared with KNR users who had less activity to do (like searching for parking or parking the car) and less distance to walk.

Further, informal parking in SEQ has an interesting nature. Based on the observation study, it was found that people preferred to park on local streets (with no fees) than to park in PNR lots when the latter required a longer walk.

4.4 Dynamic Utilization of PNR lots

In the preliminary analysis I understood the nature and accumulation profile of car arrivals at PNR lots. For example, in figure 4.7, change in usage over time is displayed. However, the results were unable to shed any light on *why* usage changes over time. Also, by focussing on car arrivals (and hence parking spaces filled) only, the characteristics of the empty lots were sidelined. Unused/empty lots also carried information on *why* they were not used, which contributed largely to understanding the very important question, *Why do some PNR lots get filled while others remain empty ?* In order to the analyse question, I developed a hazard model. The incentive for using a hazard model was to understand the determinants of filling up of a PNR space related to location, facilities, and transport features in the area, which is not answered by the accumulation curve seen in figure 4.7. A hazard model is also able address the empty parking spaces, which are the censored data (parking bay records that are left or right censored), which is difficult to address using standard statistical tools. Hence, I used a hazard method to model the un/availability of parking spaces.

4.4.1 Terminology of hazard analysis and its meaning in PNR study

I used survival analysis to model the time until a single PNR parking space was filled. The terminology of survival analysis and the way it was understood in the PNR study are explained in this section. The *Start time* for the PNR study was the same for all PNR lots at 4 am. Then the time was discretized into 5-min intervals from 4-9:30 am. Thus, the 1st interval represented 4:00am to 4:05am; likewise, the last (66th) interval represented 9:25am to 9:30am. An alternative approach I could have considered was to assign a different start time for each PNR lot depending on the start of the survey at those lots. This should have been the case if there were a considerable number of cars parked between 4 am to the time when the survey began at a lot. Also, this case would have demanded to consider left censoring (Singer and Willett, 1993) in the data, as the exact time when a parking space was filled was not known. However, there were only a trivial number of cars parked before the survey began, which I regarded as overnight parking. Also, since all surveys started prior to the first public transport service, chances were that those parked cars were not representing park-and-ride travel; it was highly unlikely that people would park ten to twenty minutes prior to their intended transit's arrival. Further, Keiley et al. (2007) mentioned that a common start time could be the one that places all individuals in the population at risk of experiencing the event; I chose the start time to be 4 am across all PNR lots.

The *end time* of the study period was different for each PNR lot and each space. If I had considered a common end time for the survey, say 10 am for all PNR lots, then it could have been problematic for those lots where the study ended due to lack of arrivals at 9 am. Then those empty spaces were censored at 9 am but it would appear as if they were censored at 10 am. In any case, a common end time would have had no effect on the lots where the survey ended because the lot was full. Another integral term in survival analysis is the censored data; in my case, the parking spaces that were filled before the start of the study or were not filled by the completion of the study were considered as parking bay records that are left or right censored. Since the start time of the study was 4 am, any cars that were parked before 4 am are left censored spaces (Kleinbaum and Klein, 2006). In this study, those spaces are negligible in number and thus I assumed that any cars parked before 4 am would also follow the same arrival distribution as the rest of the cars. I excluded cars parked before 4 am and as such there were no left censored data in this study. The parking spaces that remain empty even after the completion of the survey period were regarded as right censored (Kleinbaum and Klein, 2006) and were explicitly considered in the study.

The *Event* in my analysis was *the filling of a parking space* (or failure in typical survival analysis language). I considered that all parking spaces were at an equal risk of being occupied from the beginning of the study. Also, I considered that for each space the event was non-repeatable; i.e., a parking space could be filled only once. In reality a parking space could be filled more than once, when the parked vehicle leaves a space and another vehicle takes that space, but this event was never observed in this study. Another aspect of hazard analysis was *Risk set*. In this study, the risk set contained only those parking spaces that remained eligible to be filled in the period. According to the definition by Keiley et al. (2007), at the beginning of the period, all parking spaces were members of the risk set. As parking spaces were filled or censored, the number of spaces in the risk set decreased. Tables A.1 and A.2 (in appendix) present the risk set in this study. At the beginning (the 0 interval), the total risk set was 7590 parking spaces and at the end (the last interval) it was 186.

4.4.2 Hazard probability, survival probability and the median lifetime

There are three essential statistical summaries of the life table: the hazard function, the survivor function, and the median lifetime (Singer and Willett, 2003b). The hazard probability in any particular discrete-time period is the conditional probability that a randomly selected parking space will be filled in that time interval, given it was not filled in any earlier interval (Keiley et al., 2007). Further, a hazard function is plotted in figure 4.9.

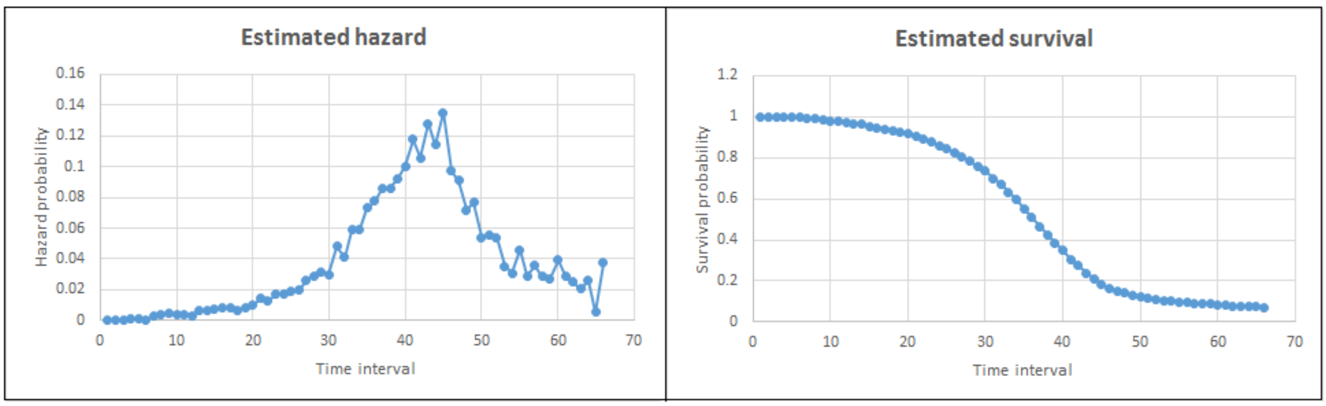

Figure 4.9: Estimated (baseline) hazard and survival probabilities

Likewise, the survival probability is the probability that a randomly selected parking space will remain empty (i.e., survive) beyond the current time interval without being filled. In Figure 4.9, the estimated median lifetime is the point in time when the value of the estimated survival function is 0.5. The median lifetime can be thought of as the average time to fill a space (Keiley et al., 2007)). The median lifetime of a parking space was between intervals 36 and 37 (around 7am). Using the survival probabilities corresponding to the intervals 36 and 37, I obtained the median life time as follows (Singer and Willett, 2003a):
The median lifetime $= 36 + (0.509 - 0.5)/(0.509 - .465) = 36.204$. The median suggested that an average parking space remained free until the 36.2th time interval; in other words, half of the parking spaces were filled by around 7 am. After summarizing event history data by constructing life tables, plotting sample hazard and survival probabilities, and estimating median lifetimes, I obtained answers to questions like: *How many parking spaces are filled in a given interval?* Or,

how many spaces remain empty at the end of the survey? Or, *what is the probability of a space filling by a certain interval?* Yet there was no answer as to why some parking spaces filled in earlier intervals while others remained empty. This question could be addressed by fitting statistical models of hazard to these data (Singer and Willett, 2003b)). For this purpose, I fitted a discrete time logistic regression model in which the unobserved hazard rate was influenced by the covariates, including both the time-independent covariates and time-dependent covariates. The model explained how the probability of a parking space being filled depended on different covariates. The model is described in the following section 4.4.3.

4.4.3 Modelling PNR occupancy as a function of predictors

Discrete time logistic regression model

Because the data were collected for discrete time intervals of five minutes, a discrete-time hazard model was suitable. In discrete-time models where time is segmented into uniform discrete intervals, exit probabilities in each time period can be estimated using logistic regression (Washington et al., 2010). The survival of a parking space in each time period was treated as a Bernoulli trial with two possible outcomes: 1 if the parking space was filled, and 0 when it was censored at the end of each time interval. If a parking space i with p different predictor values z_{1ij}, z_{2ij}…….., z_{1pj} had not been filled by the end of time interval j-1, the conditional probability for this space to be filled in time interval j was :

$$h_{ij} = Pr(T_i = j/T_i \geq j, Z_{1ij} = z_{1ij}, Z_{2ij} = z_{2ij}, ...Z_{pij} = z_{pij}) \tag{12}$$

"Because h_{ij} are probabilities, (Cox, 1992) proposed that they can be reparameterized so that they have a logistic dependence on the predictors and the time periods (Singer and Willett, 1993)". The proposed discrete-time hazard model is thus:

$$h_{ij} = \frac{1}{1 + e^{-[(\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + ...+ \alpha_p D_{pij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + ...+ \beta_p Z_{pij})]}} \tag{13}$$

Where [D_{1ij}, D_{2ij},……..,D_{pij}] are a sequence of dummy variables, indexing the time periods; [α1 α2.... αj] are intercept parameters that captured the baseline level of hazard in each time interval, and the slope parameters [β_1 β_2……… β_p] describe the effects of predictors/variables on the logistic scale (Singer and Willett, 1993).

Taking the logistic transformation on both sides of equation 13

$$log_e(\frac{h_{ij}}{1-h_{ij}}) = (\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + ... + \alpha_p D_{pij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + ... + \beta_p Z_{pij}) \quad (14)$$

When the values of predictors, $Z_1 - Z_p$ are set to zero, the population discrete-time hazard depends only on $[\alpha_1 \ \alpha_2....\alpha_j]$ and the values of the time dummies; $[\alpha_1 \ \alpha_2....\alpha_j]$ represent the baseline-logit hazard function.

Estimators for the parameters $[\alpha_1 \ \alpha_2....\alpha_j]$ and $[\beta_1 \ \beta_2......\beta_p]$ of the logistic discrete-time hazard model in equations 13 and 14, and therefore of h_{ij}, can be obtained by maximizing a likelihood function, which is the product of two distinct components; the first component is the probability for filled spaces and the second was for empty spaces. The probability that an uncensored parking space i will be filled in time interval j_i, given the parking space did not fill until the end of time interval j_{i-1}, is:

$$\begin{aligned} Pr(T_i = j) &= Pr(T_i = j/T_i \geq j) Pr(T_i \neq j-1/T_i \geq j-1).....Pr(T_i \neq 1/T_i \geq 1) \\ &= h_{ij}(1-h_{i(j-1)})......(1-h_{i1}) \\ &= h_{ij} \prod_{k=1}^{(j-1)} (1-h_{ik}) \end{aligned} \quad (15)$$

The probability that parking space i remains empty until the end of the time interval is:

$$\begin{aligned} Pr(T_i \geq j) &= Pr(T_i \neq j/T_i \geq j) Pr(T_i \neq j-1/T_i \geq j-1).....Pr(T_i \neq 1/T_i \geq 1) \\ &= (1-h_{ij})(1-h_{i(j-1)})......(1-h_{i1}) \\ &= \prod_{k=1}^{j} (1-h_{ik}) \end{aligned} \quad (16)$$

The likelihood function using equations 15 and 16, is written as:

$$L = \prod_{i=1}^{N} \left[h_{ij} \prod_{k=1}^{j-1} (1-h_{ik}) \right]^{\delta_i} \left[\prod_{k=1}^{j} (1-h_{ik}) \right]^{1-\delta_i} \quad (17)$$

Where N is the total number of parking spaces in the PNR study, $\delta = 1$ if space i was filled and $\delta = 0$ otherwise (Chen and Xie, 2014).

Predictor (variable) specification

From the data obtained from the survey and information I gathered from existing datasets on the surveyed PNR lots, I defined predictors for the discrete hazard model. In tables A.1 and A.2 (in appendix), the observations starting from the interval 0 up to 60 are shown. From observation 61 to 66, there were only 28 observations, so I excluded those intervals. Thus in the model development stage I considered the study time period up to interval 60 only (9 am). There were a total of 73 independent variables used in the model, with 60 variables representing the different time periods. Other variables were: walking time within the PNR lot (using the zone information), the capacity of the PNR lot, and the availability (schedule) of public transport service. Also, I used 10 variables that represented the effect of the parking lots' distance to the CBD (the most prominent commuting destination) for every half hour. I have described the preparation of the variables in detail in the following sections.

Time indicator variables using the space-period dataset

Following the concept described in (Singer and Willett, 1993), I prepared a PNR space-period data set. The space-interval dataset for discrete-time survival analysis has a separate record for each time period when an individual is at risk (Singer and Willett, 2003b).

Walking time inside PNR Lot

I investigated whether the position of a parking space within a PNR lot had a role in determining whether or when the space was filled. I used the zoning information I collected for each space in the PNR lot survey, as this variable was the average walking time required to reach the nearest platform, from table 4.1). For all the parking bay records that are left or right censored, the variable would take a value of walking time in the subsequent zone. E.g., if the survey ended at 8 am and the last parking space was filled in zone B (at 8am), then I assumed that all the empty (censored) spaces for that lot were in zone C. This was a time-independent variable.

Capacity

I explored the hypothesis that the rate of fill of a PNR lot was directly related to the size (capacity) of the lot. The more the lot fills, the less capacity is available (time-dependent capacity), and vice-versa. Hence, I considered one static value for the capacity of the PNR lot before the start of the study, to make sure that the variable was exogenous. This variable, capacity, takes the value of the total number of parking spaces available in the lot as observed during the survey. It is a continuous time-independent variable. Since I studied 20 PNR lots, 20 unique values for the capacity variable existed in the dataset.

Public transport service

Another variable I used in the model was the public transport service near the PNR lot. Both train and bus services within 500m walking distance of the PNR lot were calculated. The public transport variable took the value of the number of public transport services with scheduled departures in the subsequent interval. This leading effect was introduced because it was hypothesized that PNR users were likely to arrive some minutes prior to their intended public transport service (Sharma et al., 2016). Therefore, for a space that was filled in the 45th interval, the public transport variable took the value of the number of departing public transport services during the 46th interval. This was a time-dependent continuous variable because the number of departing public transport services changed for each time interval.

Distance to CBD

The variable of distance to CBD represented the distance of the parking lot from the CBD. I used ten dummy variables to represent the effect of distance to the CBD on the fill rate of parking spaces, possibly varying by half-hour period. Represented by $Dis_{CBD1} - Dis_{CBD10}$, those variables were expected to capture how the distance from the CBD changes a parking space's probability of being filled in every half hour from 4 am to 9 am.

4.4.4 Results and discussion from the developed model

Results for the baseline model and the fitted models are presented in table 4.2.

Table 4.2: Results from Discrete Time Hazard model

Model Variable	Estimates		Std. Error	Model Variable	Estimates		Std. Error
$Capacity$	0.001	*****	0.000	$Public$	0.125	*****	0.007
$Walk\_time$	−0.732	*****	0.013	Dis_CBD1	0.059	*****	0.016
Dis_CBD2	0.037	****	0.007	Dis_CBD3	0.024	*****	0.005
Dis_CBD4	0.028	*****	0.004	Dis_CBD5	0.035	*****	0.003
Dis_CBD6	0.034	*****	0.003	Dis_CBD7	0.026	*****	0.003
Dis_CBD8	−0.001	⋄	0.004	Dis_CBD9	−0.039	*****	0.007
Dis_CBD10	−0.062	*****	0.012	D1	−8.067	*****	0.167
D2	−8.587	*****	0.716	D3	−8.07	*****	0.617
D4	−7.885	*****	0.588	D5	−7.478	*****	0.539
D6	−8.311	*****	0.656	D7	−6.16	*****	0.290
D8	−5.708	*****	0.255	D9	−5.623	*****	0.248
D10	−5.788	*****	0.261	D11	−5.795	*****	0.261
D12	−5.956	*****	0.269	D13	−4.912	*****	0.191
D14	−4.962	*****	0.193	D15	−4.791	*****	0.183
D16	−4.679	*****	0.18	D17	−4.684	*****	0.18
D18	−4.918	*****	0.19	D19	−4.731	*****	0.163
D20	−4.608	*****	0.156	D21	−4.261	*****	0.141
D22	−4.326	*****	0.141	D23	−4.038	*****	0.136
D24	−4.064	*****	0.134	D25	−4.012	*****	0.12
D26	−3.996	*****	0.12	D27	−3.765	*****	0.112
D28	−3.613	*****	0.11	D29	−3.549	*****	0.108
D30	−3.068	*****	0.109	D31	−3.05	*****	0.09
D32	−3.119	*****	0.094	D33	−2.755	*****	0.087
D34	−2.734	*****	0.088	D35	−2.444	*****	0.085
D36	−2.426	*****	0.086	D37	−2.091	*****	0.087
D38	−1.896	*****	0.086	D39	−1.783	*****	0.087
D40	−1.598	*****	0.088	D41	−1.343	*****	0.087
D42	−1.324	*****	0.093	D43	−0.536	*****	0.102
D44	−0.501	*****	0.108	D45	−0.666	⋄ ⋄	0.11
D46	−0.356	***	0.149	D47	−0.206	*	0.134
D48	−0.337	**	0.149	D49	0.586	***	0.185
D50	0.541	**	0.213	D51	0.855	***	0.22
D52	0.986	*****	0.229	D53	0.623	**	0.262
D54	0.629	**	0.278	D55	1.767	*****	0.313
D56	1.457	*****	0.352	D57	2.214	*****	0.383
D58	2.212	*****	0.432	D59	2.304	*****	0.480
D60	2.926	*****	0.441				
⋄ p<1; ⋄ ⋄ p<0.8; *p<0.5; **p<0.1; ***p<0.01; ****p<0.001; *****p<0.0001							

The results in table 4.2 present the estimates of all 60 time indicator variables and other covariates. The gradual increase in the magnitude of the estimates over time $(D1 - D60)$ is consistent with pre-established understanding that the risk of a parking space getting filled increases over time. The results for $Walk\_time$ imply that the position of the parking space with respect to the platform plays an important role in its getting filled; the farther the parking space, the lesser the chance of getting filled. This result on walk time within the PNR lot is consistent with that of Tsang et al. (2005): the later the arrival time to the lot, the lesser is the probability of getting a spot close to the platform (or the spot itself), which lengthens the transfer time and hence, makes the lot less attractive. The anticipated decrement in $logit - hazard$ was -0.732 for a 1 unit difference in walking time. In my sample, walking time (representing by parking zone) had a standard deviation of 2.014 minutes, which suggested that any two parking spaces whose walking times were a standard deviation different would differ in the logit-hazard of getting filled by $(-0.732X2.01)$ or -1.47. Further, by anti-logging $(e^{-1.47} = 0.23)$, I found that the fitted odds that a parking space which had one standard deviation more walking time would be be about one-quarter of the odds of a parking space with less walking time, at each time interval. As Caicedo et al. (2006) indicated for general parking, there exists a similar trend in PNR; i.e., there is a relative advantage of one parking space versus another within a PNR lot just like the relative advantage of one parking facility over another.

The parameter estimate associated with the public transport availability (Public) was positive, indicating that higher public transport availability in the subsequent interval is likely to increase the probability of a space being filled at each interval. Note that the value of the public transport variable fluctuated over time. Further, the positive value of the parameter of the capacity variable (Capacity) suggests that in every time interval, parking spaces which belong to a high-capacity PNR lot are more likely to get filled than those in low-capacity lots.

The estimates of $Dis\_CBD1 - Dis\_CBD10$ show interesting results; as the distance to the CBD increases, the probability of a parking space getting filled increases from 4 am to 4:30am (Dis_CBD1), and so on until 8 am (Dis8). However, after 8 am, it is more likely that the parking spaces that are situated closer to the CBD will get filled. This finding suggests that the results from general parking could not be generalised to the PNR context. An explanation for these estimates could be that public transport services started early at stops located further away from the CBD, and thus, more parking spaces are filled earlier at distant lots. Another interesting insight from the model is that the shorter the distance to the PNR lot from the CBD, the greater the chance that the user would get a spot close to the platform (or the spot itself). For example, considering the distance to CBD in the first interval (Dis_CBD1), it is seen that a reduction in distance between the PNR lot and CBD by $(-0.732/.059 = -12.4)$ 12km could save a minute of walk inside a PNR lot.

Further, the results in table 4.2 were compared with the baseline model; the deviance of the baseline model and fitted model were 55436.16 and 49178.09, respectively. The better the fit of the model, the smaller the deviance (Singer and Willett, 2003b). A decrease in deviance was commendable and suggested that the model with covariates is a better fit. The model presented the variables/factors that affected the filling of parking spaces in PNR lots.

4.5 Lot choice model using *dynamic utilization* of PNR lots

4.5.1 Calculation of *dynamic utilization*

The aim of data collection in PNR lots was to understand the *utilization* of PNR lots better in association with time dynamics. The developed hazard model explained the factors that affect the un/availability of parking spaces dependent on the time-of-day in PNR lots and thus the dynamic *utilization* of lots. The lot choice model presented in chapter 3 used static utilization. A more profound understanding of the time dynamics in *utilization* of PNR lots could add value to the PNR lot choice model.

Using the hazard model, value of hazard for all PNR lots in SEQ could be calculated as presented in equation 13. The model presented in the table 4.2 was based on the 20 PNR lots for which I had information on the walk time, but I did not have such information for all PNR lots in SEQ. Hence, this time, I developed hazard models without using walk time. I developed two hazard models; (i) with the variables distance to CBD and capacity and (ii) with the variables distance to CBD, capacity and public transport service.

From the two sets of models, I had estimates for the coefficients of distance to CBD, capacity and public transport service. I needed to calculate cumulative hazard (*dynamic utilization*) for all the PNR lots in SEQ by simulating the estimates prepared from 20 models. In other words, for feeding in equation 13, all the αs were available, and Zs were needed. For Zs, the capacity of all PNR lots in SEQ were taken from the DTMR survey. The distance from the CBD to all PNR lots in SEQ were calculated using ARCGIS. To find the public transport service for all PNR lots in SEQ I used General Transit Feed Specification (GTFS) data. For each PNR lot I found the list of public transport stops within 500m using the stopfile from GTFS. Then, using the *stoptime file* from GTFS, I found the number of public transport services those stops catered in every 5 minutes starting from 4 am to 9:30 am on a typical weekday. Note here that I treated cumulative hazard for each PNR in each time interval as the proxy of *dynamic utilization* to prepare the new lot choice model. Figure 4.10 shows the average cumulative hazard of all PNR lots in SEQ.

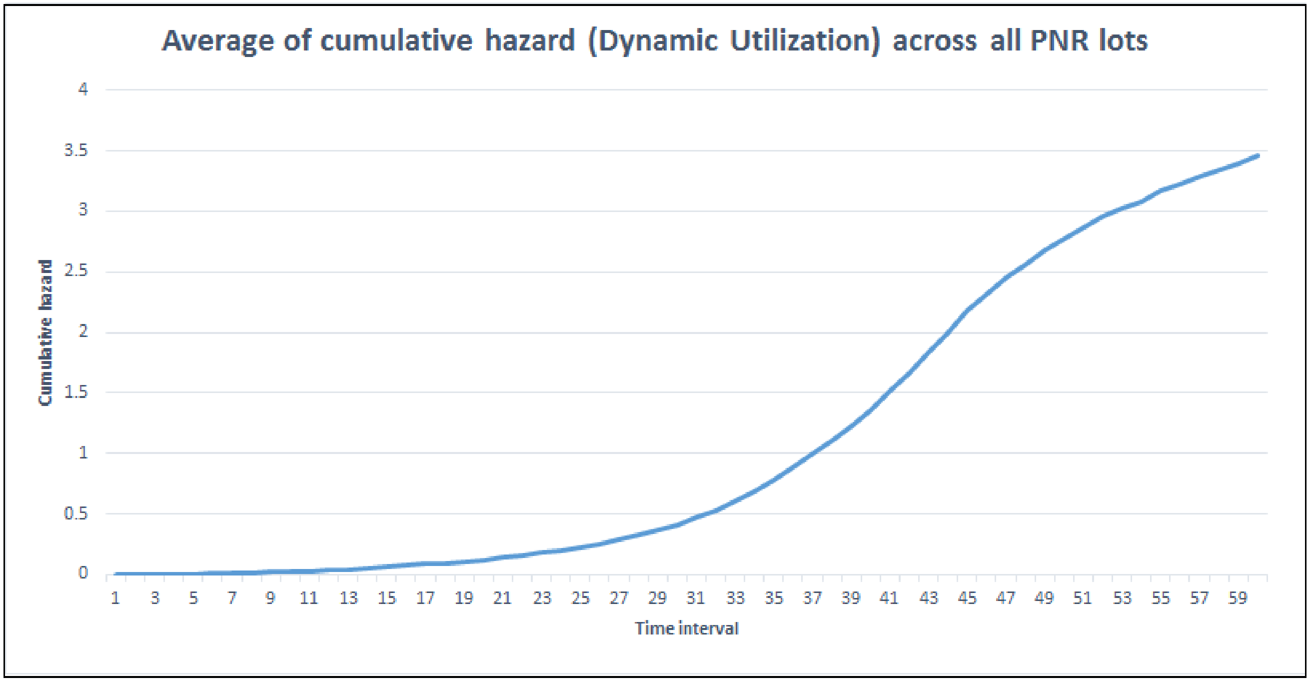

Figure 4.10: Average utilization of PNR lots in SEQ with time

4.5.2 Preparation of new lot choice model with dynamic utilization as a variable

I aimed to prepare a new lot choice model using *dynamic utilization* as one of the variables alongside with others described in Chapter 3. *Dynamic utilization* for each PNR lot is modelled at the level of 5-minute time intervals. For each observation in the PTOD survey, the value of *dynamic utilization* of PNR lots depended on the arrival time of users at the lot. The arrival time of each user at PNR lots was calculated using the origin to PNR lot time (close space as mentioned in previous Chapter, *optime*) and start time of the traveller. The arrival times were then classified in intervals, starting from 4 am, in the interval of 5 minutes. For example, if a traveller *A* was seen to start the journey at 6:42 am and it was found that it took her/him 10 minutes, 20 minutes, 30 minutes to reach PNR lot 1, PNR lot 2, PNR lot 3 respectively, then the arrival time of person *A* at PNR lot 1, PNR lot 2 and PNR lot 3 would be at 6:52 am (35th interval), 7:02 am (37th interval) and 7:12 am (39th interval) respectively. Also, as mentioned in section 4.5.1, for each PNR lot I had the value of *dynamic utilization* (cumulative hazard) for all time intervals 1 to 60. Thus, for traveller *A*, depending on her/his time of arrival at the PNR lots in her/his choice set, I used *dynamic utilization* (cumulative hazard) of PNR lot 1 in its 35th interval, PNR lot 2 in its 37th interval and PNR lot 3 in its 39th interval. Thus, *dynamic utilization* variables for all observations in their respective time intervals were prepared.

In the discrete hazard model, the time intervals used were up to the 60th interval, however in the PTOD survey there were some observations that were seen to reach PNR lots after the 60th interval. Out of the 3364 observations modelled in chapter 3, I could not find the time of start for 134 observations. Hence, the new model accommodated 3230 observations only.

For the purpose of making a clear comparison between the lot choice models due to the effect of *static* and *dynamic utilization*, I estimated two set (one using static utilization and another using dynamic utilization) models with same number of observations i.e., 3230. In total, three models are presented in table 4.3 : (i) a lot choice model that uses static utilisation (ii) a lot choice model that uses dynamic utilization coming from the hazard model with variables distance to CBD and capacity and (iii) a lot choice model that uses dynamic utilization coming from the hazard model with variables distance to CBD, capacity, and public transport service.

Table 4.3: Comparison of Lot Choice Models using Static versus Dynamic Utilization

	Using Static Utilization		Using Dynamic Utilization (1)		Using Dynamic Utilization (2)	
Variables	β	t	β	t	β	t
Capacity	0.001	9.61	0.002	13.82	0.002	13.39
Invehtime	-0.126	-26.64	-0.099	-18.52	-0.099	-18.38
Optime	-0.452	-48.32	-0.460	-47.94	-0.459	-47.90
Totfare	0.398	4.76	0.277	3.19	0.278	3.20
Waittime	-0.03	-4.99	-0.016	-2.52	-0.016	-2.54
Walktime	-0.081	-6.92	-0.077	-6.67	-0.011	-6.69
Utl	0.01	1.10				
Dynutl			2.12	10.94	2.150	11.08
Observations	3230.000		3230.000		3230	
Initial log-likelihood	-9676.215		-9676.215		-9676.215	
Final log-likelihood	-3780.727		-3720.590		-3717.378	
Adj.R^2 of the model	0.609		0.615		0.615	

As seen in table 4.3, the use of *dynamic utilization* (represented by *dynult*) was significant in both the models. The *dynamic utilization* not only became significant but also slightly improved the overall fit of the model. It improved the log likelihood by almost 60 units. All other estimates had intuitive sign and values. While one can argue that only a slight improvement in the overall fit of the data at the expense of data collection and survey and modelling might not be worth the benefits, the positive and significant estimate of *dynamic utilization* suggests a need for more research in this area.

4.5.3 Correction of endogeneity for the model with dynamic Utilization

Though the dynamic utilization variable is significant and intuitive, from the analyses presented in chapter 3, it is understood that utilization of PNR lots is endogenous in nature. Hence, for the correction of endogeneity in the lot choice model with dynamic utilization as a variable, I used the Control Function method Guevara (2010) as mentioned in the section 3.9.3 of Chapter 3. From the table 4.3, it is seen that the dynamic utilization (cumulative hazard) coming from the hazard model with the variables distance to CBD and capacity only, performed better than the one that included public transport service too. Hence, from this stage onwards I used the model 1 from table 4.3. The first and second instruments were the average dynamic utilization of PNR lots located within 2.5 km to 5km and the average dynamic utilization of PNR lots located within 5 km to 7.5 km respectively from the concerned PNR lot.

Correlation among IVs is shown in table 4.4.

Table 4.4: Correlation among Dynamic Utilization and Instrumental variables

	Dynamic Utilization	IV1	IV2
Dynamic Utilization	1	0.94	0.93
IV1	0.94	1	0.95
IV2	0.93	0.95	1

The correlation was relatively high for both IV1 and IV2; IV1 was a bit more correlated to the dynamic utilization than IV2. I then regressed IV1 an IV2 against the dynamic utilization. The results of this auxiliary regression are shown in table 4.5.

Table 4.5: Results from auxiliary regression

	Coefficient	t
Constant	0.00	8.87
IV1	0.60	210.90
IV2	0.36	29.53
R^2 of the model		0.89
Adjusted R^2 of the model		0.89

The residual obtained from the regression stage was used as an additional variable in the utility. The estimation result from both models, i.e. with correction for endogeneity and without correction for endogeneity, for the two set of models is presented in table 4.6.

Table 4.6: Estimation results for the model with and without correction of endogeneity

	Without correction		With correction	
	β	t-test	β	t-test
Capacity	0.002	13.82	0.002	13.39
Invehtime	-0.099	-18.52	-0.098	-18.38
Optime	-4.60	-47.94	-0.459	-47.90
Totfare	0.277	3.19	0.278	3.20
Waittime	-0.016	-2.52	-0.016	-2.54
Walktime	-0.077	-6.67	-0.077	-6.69
Dynamic Utilization	2.12	10.94	2.15	11.08
Residual			-0.197	-2.47
Observations	3230		3230	
Initial log-likelihood	-9676.215		-9676.215	
Final log-likelihood	-3720.590		-3717.378	
Adj.R^2 of the model	0.615		0.615	

From the table 4.6, it is seen that after applying the correction to the model, the residual is significant. Thus, the model is corrected for endongeneity.

4.6 Conclusions

In this Chapter I presented the survey conducted on the PNR lots of SEQ in the quest to understand the *utilization* variable better. *Utilization* is studied with regard to time-dynamics; based on the literature in Chapter 2, utilization depends on the time-of-day. I presented the details on the PNR lot survey done at the twenty PNR lots in SEQ. As part of preliminary analysis, I studied the utilization at twenty PNR lots in SEQ. It was interesting to observe how some PNR lots reached their capacity earlier in the morning while some filled later.

I studied an individual parking space's availability for a given time interval (itself a scarce topic in the PNR literature) and I developed a model that explained the factors affecting the un/availability of these spaces. The results suggest that the location of parking spaces within a parking lot is also important alongside the relative location of the PNR lot in the city. This result is in line with findings from another study on metro commuters' satisfaction where they concluded that PNR users are more sensitive to comfortable walking environments compared to other users who use bus and walk as access modes to metro (Yang et al., 2015). From the developed model, it was understood that public transport availability is an important factor influencing the filling rate of PNR lots. This also supports the point that was raised in Chapter 2; parking un/availability in

PNR lot is different to that of general lots. Studies on PNR lots have more dimensions to address than a general parking lot, as PNR lots are not the end of the trips , but rather are followed by public transportation.
Further, I used the cumulative hazard of PNR lots obtained from the discrete hazard model, as one of the *Dynamic Utilization* variables to prepare a new lot choice model, and it outperformed the model with *Static Utilization*. Also, the *Dynamic Utilization* variable is significant in the model and the model is corrected for endogeneity. This concludes that *utilization* of PNR lots is a dynamic property, and thus PNR demand is a dynamic process.

4.7 Policy Implications

The understanding of time-dependent parking un/availability, i.e. dynamic utilization of PNR lots, is beneficial for planners to comprehend the dynamic PNR demand at PNR lots. For most research, PNR demand has been confined to its static dimension. Apart from this, the finding that the relative position of a parking space within a PNR lot was found to be important alongside the relative location of a PNR lot in the city could initiate policies that address *How to locate PNR lots so that the relative distance/cost of parking spaces from public transport platform is minimized?*, apart from the existing focus on *Where to locate PNR lots?* The understandings from this chapter are useful for PNR planners in managing PNR availability during the morning peak more effectively. The results illustrate how improving access to platforms from parking spaces within a PNR lot, location of PNR lot, the distance from CBD, and increasing capacity of lots facilitates better management of park-and-ride availability during the morning peak.

4.8 Further direction from this Chapter

In this chapter I presented the details of the PNR survey and the discrete time hazard model that described the filling up of a parking space within a PNR lot. Further, by using the results from the hazard model, I measured the *dynamic utilization* of PNR lots and used it as a variable in PNR lot choice model. The developed lot choice model thus became a dynamic PNR lot choice model, which explained the reasons why travellers choose one PNR versus another. Itcontributed towards a more in-depth understanding of a traveller's decision to choose PNR as a mode versus other modes like drive alone or public transport. In Chapter 5, I present the mode choice model that I prepared.

5 Mode Choice Model for SEQ

5.1 Introduction

In chapter 4, I presented a complete dynamic PNR lot choice model which explained why users choose one PNR lot versus another. Following from this, it became of interest to understand the choice of PNR as a mode in the bigger picture. I thus modelled the mode choice behaviour of South East Queenslanders based on their household travel survey data. This chapter is concerned with presenting the mode choice model developed for SEQ. This model is based on the concept that a person's mode choice decision is based on the resources available to them in their household. The first section 5.2 presents the dataset used for this research. Section 5.3 is about the frameworks I used in estimating the model choice models. The concept I use states that people make their everyday short term decisions like mode choice based on their fixed long term decisions like owning a resource (car, bicycle, and motorbike). Based on this concept I present two sets of models. In section 5.4, the variables used for the models are explained. Then, in section 5.5, results from both models are discussed. Lastly, section 5.6 records the conclusions drawn from this Chapter.

5.2 Dataset

The study area remains South East Queensland. The first dataset used for this study was the Household Travel Survey (HTS survey) conducted by the Department of Transport and Main Roads (DTMR), which had rich information on peoples' trip making behaviour and their socio-economic information, and their households' vehicle holdings. Further, I used a Park-and-ride lot survey conducted by DTMR to observe the characteristics of PNR lots in SEQ such as whether the lot is a formal lot or not etc.

Data cleaning

The data cleaning process is as shown in table 5.1.

Table 5.1: Flowchart showing the trimming of sample

Process	Number of trips
Total from household travel survey	108,913
Include only work and education related trips	31,115
Include trips whose ODs are within SEQ	23,861
Exclude trips with purpose 'other'	23,850
Exclude intra-zonal trips	22,003
Exclude BNR and KNR trips	21,786
Exclude trips with high public transport travel times	19,038
Exclude trips with no time recorded	18,048
Include morning trips only (42% of the total)	7,580

Out of the total number of trips in the HTS survey, only those related to work were of interest to this study. Since my research is about PNR, and from previous chapters I understood that PNR users use them mainly for work-related trips, I chose to model only work related trips. Thus, out of 108,913 total trips, 31,115 trips were work and education related i.e., around 29% of the total household trips. Out of these trips, those whose origin and destination are within South East Queensland's Strategic Transport Model (SEQSTM) were 23,861. Among the 23,861 trips seen in the HTS survey, some trips were mentioned to have used mode 'other'. Thus, I excluded trips whose modes were not clear, and the remaining trips were 23,850 in number. Also, some trips were so short that their origin and destination were within same zone and thus represented the same node (centroid). For such trips I was not able to find travel time and other attributes. After deleting such trips there were 22,003 remaining.

I then removed Bike and ride (BNR) and Kiss-and-ride (KNR) observations from the total trips to get 21,786 trips, as modeling BNR and KNR trips is not relevant to the present study. Further, some observations had very high travel times, indicating that public transport mode was not available for those observations. If I were to conduct a simple mode choice analysis, I would have still considered those trips and rather indicated that their public transport availability or auto mode availability was 0. However, in this case the choice was for a portfolio and since they were negligible in number, I excluded them. The final remaining number of observations for the mode choice study was 19,038.

Out of the final 19,039 data observations used for the study, the majority of trips did not have any time recorded. For this study I used only trips which were made in the morning. Out of the remaining, almost 42% were performed in the morning which further reduced the number to 7,580 trips.

Share of modes and portfolios based on the observation

Out of 7580 total observations considered in the mode choice study, table 5.2 shows the share among modes:

Table 5.2: List modes and their share in the dataset

Mode	Actual Share	Share in percentage
Drive alone (D)	4338	58%
Passenger (P)	1819	24%
Bicycle (C)	139	2%
Motorbike (B)	73	1%
Walk (W)	326	4%
PT (T)	719	9%
PNR (N)	166	2%
Total	7580	100%

Further, I defined 'portfolios' (Le Vine et al., 2013) as the set of modes, enabled by certain resources. Based on the availability of resources in the household, I defined eight portfolios. The first portfolio is enabled by car, second by bicycle, third by motorbike, fourth by car and bicycle, fifth by car and motorbike, sixth by bicycle and motorbike, seventh by car, bicycle, and motorbike, and eighth by none. Table 5.3 shows the share of portfolios among the observations.

Table 5.3: Portfolios and their share in the dataset

Portfolios	Actual Share	Share in percentage
Portfolio 1	1969	26%
Portfolio 2	1551	20.5%
Portfolio 3	15	0.1%
Portfolio 4	3398	45%
Portfolio 5	72	1%
Portfolio 6	32	0.4%
Portfolio 7	314	4%
Portfolio 8	229	3%
Total	7580	100%

5.3 Framework of the model

I used two different frameworks to model the mode choice decisions. In the first framework, the alternatives are the portfolios, eg. portfolio 1, portfolio 2 etc., and in the second the alternatives are the modes, eg. Car, Bus, Walk etc.

Resource holdings

Mobility resources are not exclusive and should be "durable, market-traded, widely held, and substantial" (Le Vine et al., 2013). The three types of resources that I consider in this study are (i) car, (ii) bicycle, and (iii) motorbike. These resources will affect the following modes: namely, driving car alone, being a passenger in a car (pax), riding a bicycle, riding a motorbike, walking, using taxi, riding public transport (PT), and using PNR. In my setting, a traveller needed to own a car, bike, and bicycle in order to drive car, ride motorbike, and bicycle respectively. Another resource could have been Go-card, which is a prepaid integrated ticketing system in SEQ which allows users to travel in all three modes of public transport in SEQ: train, bus, and ferry. However, information on Go-card availability was limited in the HTS survey data, so it could not be used as a resource. Thus, public transport was treated as a mode that did not require any resource. Another type of resource could be the rent people pay every week/month to make sure that they live in areas that are accessible by public transport. But, in my analysis I did not go to that level of detail. Apart from public transport, other modes in this study that did not require any resource were walk, taxi, and car passenger. This is an assumption I made in this study to avoid further complicating the model, otherwise even walking requires a resource (Le Vine et al., 2013). With three resources, a person's portfolio options consists of (2^3=) 8 choice sets as shown in table 5.4.

Table 5.4: List of resource holding and portfolio for the mode choice study

Resources	Drive (D)	Pax (P)	Cycle (C)	MBike (B)	Walk (W)	PT (T)	PNR (N)	Name
Car	Yes	Yes	No	No	Yes	Yes	Yes	DPWTN
Bicycle	No	Yes	Yes	No	Yes	Yes	No	PCWT
Mbike	No	Yes	No	Yes	Yes	Yes	No	PBWT
Car+Bicycle	Yes	Yes	Yes	No	Yes	Yes	Yes	DPCWTN
Car+MBike	Yes	Yes	No	Yes	Yes	Yes	Yes	DPBWTN
Cycle+MBike	No	Yes	Yes	Yes	Yes	Yes	No	PCBWTN
All	Yes	Yes	Yes	Yes	Yes	Yes	Yes	DPCBWTN
None	No	Yes	No	No	Yes	Yes	No	PWT

Framework 1: Choice of portfolios

The framework for this analysis is based on Le Vine et al. (2013) which is as follows:

$$U_d^i = V_d^{i,non-travel} + V_d^{i,travel} + \epsilon_d^i \quad (18)$$

The concept is that, the utility of portfolio *d* to person *i* combines the systematic component of utility of holding and/or acquiring the portfolio $(V_d^{(i,non-travel)})$, and the systematic component of utility that is facilitated by the portfolio *d* to access the activities in a person's PAS $(V_d^{(i,travel)})$ and the error term (ϵ_d^i) (Le Vine et al., 2013). Among other notations, *r* is the mobility resources, *d* means the mobility resource portfolios, and *j* signifies the modes of travel. Thus, Equation 18 "shows a person's weighing between the dis(utility) (hassle expense, etc) of acquiring and/or maintaining the resources contained within a portfolio and the dis(utility) of using the travel modes which it enables to access activities in their perceived activity set". The travellers weighing is thus framed as:

$$U_d^i = \sum_{r=0,r\in d}^{R} V_r^{i,non-travel} + \Big(\sum_{j_i=1}^{j_1} \gamma_i * \frac{1}{\lambda^{travel}} * ln \sum_{m\in\mu_d}^{M} e^{(V_{m_{ji}}^{i,travel*\lambda^{travel}})}\Big) + \epsilon_d^i \qquad (19)$$

The term $\sum_{r=0,r\in d}^{R} V_r^{i,non-travel}$ is fixed, because it represents the fixed expense/hassle of owning the same portfolio of resources from adding up the expense/hassle of the resources within the portfolio. The term $(\sum_{j_i=1}^{j_1} \gamma_i * \frac{1}{\lambda^{travel}} * ln \sum_{m\in\mu_d}^{M} e^{(V_{m_{ji}}^{i,travel*\lambda^{travel}})})$ represents the specification for utility of traveler's expected travel needs. "It is hypothesised that a person views how well a portfolio would perform in providing access to a particular activity to be how well the optimal' mode enabled by this portfolio would perform to access the activity, which is then summed across all activities within the person's perceived activity set"(Le Vine et al., 2013). Moreover, the logsum form indicates the selection of the optimal mode from within the set of all modes enabled by portfolio *d* (μ_d). The terms γ_{ji} are defined as the *importance* terms to capture the likely higher priority that travellers might consider to access a certain type of activity over another. In my study, I analysed the behaviour of commute trips only and thus assumed that the importance placed by all travellers on their journeys is the same. I simply assumed γ to be 1. Thus, the model became the one shown in equation 20:

$$U_d^i = \sum_{r=0,r\in d}^{R} V_r^{i,non-travel} + \Big(\frac{1}{\lambda^{travel}} * ln \sum_{m\in\mu_d}^{M} e^{(V_{m_i}^{i,travel*\lambda^{travel}})}\Big) + \epsilon_d^i \qquad (20)$$

Availability of Portfolios

Not all portfolios are available to everyone. Therefore, I defined the availability of portfolios based on two categories. The portfolios without any car or motorbike in them are easily available to everyone. So, for all users, all portfolios without car or motorbike (PCWT, PWT) are always available (indicated by *portfolios always available*). Thus, availability is recorded as 1 for all of them. In the case of portfolios which include car or motorbike or both (DPWTN, PBWT, DPCWTN, DPBWTN, PCBWTN, DPCBWTN), persons under 16 years cannot have those portfolios in their options. Therefore, the availability of portfolios that contain car or motorbike to those users who are 16 or over 16 years have availability 1, and for the rest it is 0 (indicated by *license holders*).

While this framework worked nicely and provides detailed information, the availability of a portfolio is based only on the resource a household holds. One big issue with working in this framework is the uncertainty that the chosen portfolio (that we inferred) matches the actual chosen option. For example, based on the household resource car held in person *A*'s household, I would infer that person *A* chose portfolio 1, (i.e, DPWTN). In reality, s/he might have chosen public transport on the trip day. Then one can argue why is s/he wasnot inferred to have chosen portfolio 8 or other. Amidst all these issues, I decided to explore an alternative modelling framework, and developed a mode choice model using a simple framework, which is explained in the following paragraphs.

Framework 2: Choice of modes using GNL

In this framework I used a (GNL model). (Wen and Koppelman, 2001). I named 'nests' as 'portfolios' (for easy comparisons across the models). Nesting is done in the GNL framework based on the resource a household owns. The probability of choosing an alternative is :

$$P_n = \sum_m (P_{n/m} P_m), \tag{21}$$

where P_m is the probability of nest *m* which is,

$$P_m = \frac{(\sum_{n' \in N_m} (\alpha_{n'm} e^{v_{n'}})^{1/\mu_m})^{\mu_m}}{\sum_m (\sum_{n' \in N_m} (\alpha_{n'm} e^{V_{n'}})^{1/\mu_m})^{\mu_m}} \tag{22}$$

Where N_m is the set of all alternatives included in nest *m*, α_{nm} is the allocation parameter which characterizes the portion of alternative *n* assigned to nest *m* (α_{nm} must satisfy the condition $\alpha_{nm} \geqslant 0$, the additional condition $\sum_m \alpha_{nm}$=1, $\forall$ *n* provides a useful interpretation with respect to allocation of each alternative to each nest), μ_m is the logsum or dissimilarity parameter for nest *m* ($0 < \mu_m \leq 1$). Then $P_{n/m}$, the probability of alternative *n* if nest *m* is selected, is (Wen and Koppelman, 2001) :

$$P_{n/m} = \frac{(\alpha_{nm} e^{V_n})^{1/\mu_m}}{\sum_{n' \in N_m} (\alpha_{n'm} e^{V_{n'}})^{1/\mu_m}} \tag{23}$$

There are eight nests (portfolios) altogether as mentioned in table 5.4 and illustrated in figure 5.1

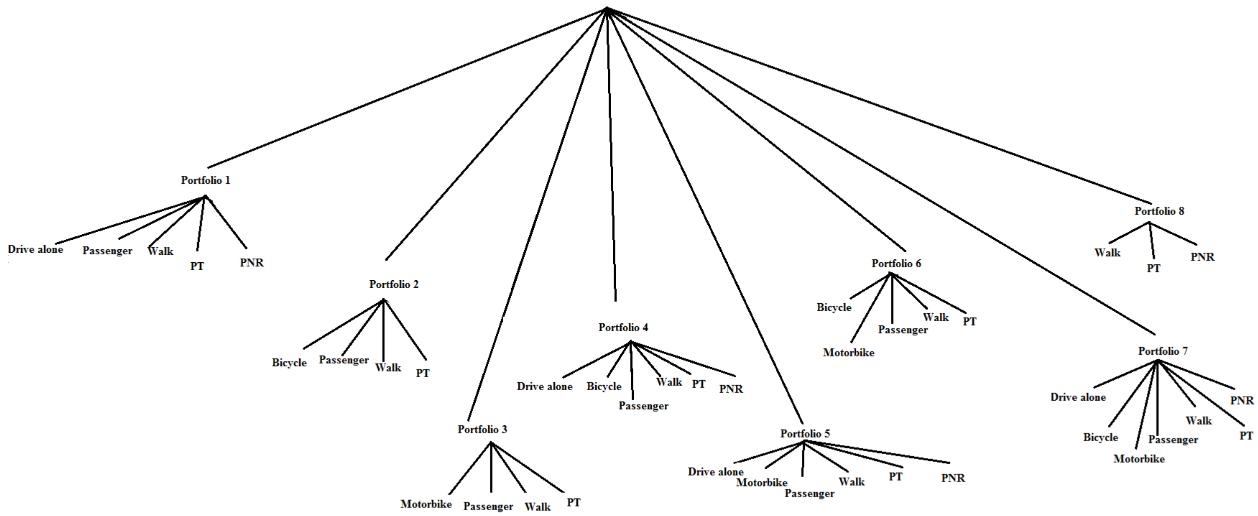

Figure 5.1: Generalized nested logit

Portfolio 1 represented the sub group of modes which are enabled by car, likewise, portfolio 2, 3, 4, 5, 6, 7, and 8 and those that are enabled by, bicycle, motorbike, car and bicycle, car and motorbike, bicycle and motorbike, all three, and none respectively. In this way, nesting captures the effect of holding a resource at the household level.

5.4 Attributes used in (both) the models

Travel time

The travel time for auto mode (*autotime*) was calculated using the South East Queensland Strategic Transport Model (SEQSTM) in EMME. In SEQSTM, travel times were available for four different time periods during the day. The travel time in a shortest (travel time) path from the morning period was chosen. Further, travel time (*transittime*) for public transport mode was obtained using the BTDSP algorithm using GTFS data, similar to what was mentioned in chapter 3. The transit travel time was categorised into inside vehicle travel time (*invehtime*), time required to walk in the transit network (*walktime*), and time spent waiting for transit in the transit network (*waittime*). For calculating the travel time using motorbike mode, I assumed that motorbike is allowed wherever auto mode is allowed. First, I found the distance traversed in the auto network while travelling in the shortest (travel time) path. Then using a motorbike speed value of 60kmph, I simply found the travel time (distance/speed) by motorbike mode. Likewise for the bike and walk mode, I again used SEQSTM to find the distance traversed in going from one origin to the destination. Then using the speed of 14.5 kmph and 4.5 kmph (Charles-Edwards et al., 2015) for bicycle and walk mode respectively, I found the travel times using bicycle and walk network. Note that the distances found in SEQSTM in the case of motorbike and bicycle/walk was different; in bicycle/walk mode the shortest path was found only in those links where bicycle/walk was feasible, whereas for motorbike, the calculation was done in those links where auto (hence motorbike) was allowed. Finally, travel time for passenger mode (Pax) was simply the *autottime*. The difference between passenger mode and auto mode was that in the former, people are not required to hold a car resource in their household, whereas in the latter they are.

Demographics

The socio-economic variables used were Income (Inc) and Millennial (Mil). *Inc* was a continuous variable that represented actual household income. Millennial was a dummy variable that was true if a person was below 30 (since the household travel survey was conducted in 2012). In the portfolio setting, the *Inc* variable was included in the utilities of all alternatives except alternative number 4. Alternative number 4 (those who are enabled by car and bicycle) was the dominant alternative and I kept it as a base alternative. *Mil* variable was associated with those portfolios which were enabled by bicycle and motorbike. Likewise, in the GNL setting, *Inc* was added in all utilities except to that of drive alone alternative and *Mil* was associated with the motorobike mode.

Composite utility from the lot choice model

I generated the PNR lot choice set for all observations both for observations; that chose PNR as the mode, and all others. Each person initially had to have all the PNR lots in SEQ in their choice set. Among them, only those PNR lots that would take less time to reach than the destination were retained in the choice set. Further, among these lots, 20 PNR lots were randomly selected and thus comprised the choice set for all observations. I then generated a table with observations and the node number of 20 PNR lots. For example, if table 5.5 represented a part of the observations from my actual table (after randomly sampling 20 PNR lots), then for Trip 1, Trip 2, and Trip 3 there were 20, 1, and zero PNR lot/s available in the choice set. It was understandable that not all trips would have all 20 PNR lots available to them as an option; I excluded all PNR lots that took a longer time to reach than the destination, which reduced the choice set for some trips. Also, Trip 3 could represent those trips which were made by walk mode or bicycle mode or even car mode, but were so short that even the nearest PNR lot took longer to reach than their destination.

Table 5.5: Choices of PNR lots for all observations

Observation	PNR 1	PNR 2	PNR 3	PNR 4	.....	PNR 20
Trip 1	400	576	345	890	.....	432
Trip 2	400	-	-	-	.....	-
Trip 3	-	-	-	-	.....	-

In the lot choice model prepared in Chapter 3, I used several variables as mentioned in table 3.1. In this case, I used the household travel survey, and unlike the PTOD survey, I did not have detailed information on certain characteristics and aspects of the travelers. For example, I had only a few observations with information on whether a traveler used Go-card, as this information was missing for the majority of travelers. This restricted me from constructing the *Totfare* variable (total fare in transit). So, in this section, I present a re-developed PNR lot choice model as developed in Chapter 3, using the PTOD survey but bearing in mind the availability of the variables in the HTS survey. The new model is as presented in table 5.6. Here, *pdtime* represents the overall travel time in the transit network including inside vehicle time, walk time, and wait time. Using the estimates from this model, I found the composite utility and then the logsum value, where composite utility is simply the denominator in the equation 6 and logsum is the natural logarithm of that denominator. The logsum captured the attractiveness of PNR as a mode by addressing the attractiveness of PNR lots for a PNR trip.

The PNR lot choice model that I chose to implement using the HTS survey data is as follows:

Table 5.6: Model estimation results from PNR lot choice model

Utility parameters	β	t-test
Netdist	-0.811	-44.86
Optime	0.293	32.32
Pdtime	-0.0627	-13.59
Train	1.98	26.66
Observations	3679.000	
Initial log-likelihood	-11021.299	
Final log-likelihood	-1957.717	
Adj.R^2 of the model	0.822	

The variables used in the mode choice models (both portfolio based and GNL) are as presented in table 5.7.

Table 5.7: List of variables used in the portfolio based model

Utility Parameters	Definition
HOLD_NONE	Alternative specific constant for holding no resource
HOLD_OWN_BICYCLE	Alternative specific constant for holding bicycle
HOLD_OWN_CAR	Alternative specific constant for holding car
HOLD_OWN_MOTORBIKE	Alternative specific constant for holding motorbike
HOLD_PAX	Alternative specific constant for passenger mode
HOLD_PNR	Alternative specific constant for PNR mode
HOLD_PT	Alternative specific constant for public transport mode
HOLD_WALK	Alternative specific constant for walk mode
MODECHOICE_BICYCLE	Alternative specific constant for mode choice bicycle
MODECHOICE_DRIVE	Alternative specific constant for mode choice drive
MODECHOICE_MOTORBIKE	Alternative specific constant for mode choice motorbike
MODECHOICE_PNR	Alternative specific constant for mode choice PNR
MODECHOICE_SHARED	Alternative specific constant for mode choice shared modes
FIXED_AT_ZERO	Parameter fixed at zero
AUTOTIME	Auto travel time from origin to destination
BICYCLETIME	Bicycle time from origin to destination
INCOME	Income of individual
LOTCHOICE	Logsum of PNR lot choice for PNR mode
MOTORBIKETIME	Travel time in motorbike from origin to destination
PAXTIME	Travel time as passenger mode from origin to destination
TRANSITTIME	Travel time in transit from origin to destination
WALKTIME	Travel time for walk mode from origin to destination
MIL	Dummy variable ‘Millennial’ to represent the generation that individual belongs to
ZONE	Zone variable which represents the number of zone traversed (as proxy of transit fare)
LAMBDA_MODECHOICE	Lambda of mode choice

5.5 Results from the developed models

5.5.1 Result form the portfolio based model

In this section I present the results from the model which treated portfolios as alternatives. The mode chosen by individuals was observed from the survey. Based on an individual's household resource holdings, portfolios were defined and I modelled the choice of portfolios as presented in table 5.8.

Table 5.8: Model estimation results from portfolio choice model

Utility Parameters	β	**t-test**
ASC_HOLD_NONE	0.00	
ASC_HOLD_OWN_BICYCLE	0.429	8.25
ASC_HOLD_OWN_CAR	-0.024	-0.07
ASC_HOLD_OWN_MOTORBIKE	-2.59	-40.42
ASC_HOLD_PAX	0.00	
ASC_HOLD_PNR	0.00	
ASC_HOLD_PT	0.00	
ASC_HOLD_WALK	0.00	
ASC_MODECHOICE_BICYCLE	-2.94	-4.95
ASC_MODECHOICE_DRIVE	0.00	
ASC_MODECHOICE_MOTORBIKE	-6.68	-7.13
ASC_MODECHOICE_PNR	-4.66	-6.91
ASC_MODECHOICE_SHARED	-4.78	-7.49
BETA_FIXED_AT_ZERO	0.00	
BETA_AUTOTIME	-0.300	-5.12
BETA_BICYCLETIME	-0.069	-5.96
BETA_INCOME	-0.091	-4.73
BETA_LOTCHOICE	0.019	1.95
BETA_MOTORBIKETIME	0.006	0.56
BETA_PAXTIME	-0.289	-3.17
BETA_TRANSITTIME	-0.139	-2.42
BETA_WALKTIME	-0.022	-3.90
BETA_MIL	-0.263	-4.88
BETA_ZONE	0.393	0.71
LAMBDA_MODECHOICE	1.00	
Observations	7580	
Initial log-likelihood	-11272.046	
Final log-likelihood	-6112.829	
Adj.R^2 of the model	0.473	

All the results from the portfolio choice model are intuitive. I fixed the drive mode to be the base mode. The negative sign of the *income* variable suggests that the probability of choosing modes other than drive alone decreases as income increases. This is understandable as high income is often associated with higher car use. All variables representing travel times in different modes are negative which makes sense; travellers prefer modes which minimize their travel time. The alternative specific constant of holding a car was found to be not significant, which could be explained by the fact that Australia is a highly car oriented society and a majority of households hold cars. Also, whether a person owns a motorbike or not is predicted more accurately than whether they own a bicycle or not. Further, whether a person owns a car or not is predicted less accurately. The logsum value of PNR lot choice is not significant, suggesting that the choice of PNR as a mode is not dependent on the choice of PNR lots. This result is consistent with the result from another study where they found that those who currently make their entire trip by car are hesitant to change to PNR (Smith et al., 2016). Apart from these, certain demographic characteristics like being a millenial make a person more likely to hold bicycle or motorbike resources.

The use of a portfolio based model allowed a greater understanding of the effect of holding car, bicycle and motorbike resources, which is an advantage over a simple MNL or NL model.

5.5.2 Result from the GNL model

Using the same data but analysing the actual mode choice of travelers this time, I estimated a GNL model as presented in table 5.9 and table 5.10.

Table 5.9: Model estimation results from mode choice model

Utility Parameters	β	**t-test**
ASC_MODECHOICE_BICYCLE	-1.27	-8.44
ASC_MODECHOICE_DRIVE	0.00	
ASC_MODECHOICE_MOTORBIKE	-1.06	-6.48
ASC_MODECHOICE_PAX	0.249	3.54
ASC_MODECHOICE_PNR	-2.54	-5.14
ASC_MODECHOICE_TRANSIT	-0.564	-2.79
ASC_MODECHOICE_WALK	0.945	9.92
BETA_AUTOTIME	-0.0312	-4.94
BETA_BICYCLETIME	-0.0590	-9.01
BETA_LOTCHOICE	0.001	2.51
BETA_MOTORBIKETIME	-0.0323	-4.38
BETA_PAXTIME	-0.0899	-8.65
BETA_TRANSITTIME	-0.0235	-9.49
BETA_WALKTIME	-0.0614	-7.36
BETA_INCOME	-0.0138	-0.98
BETA_MIL	-0.226	-1.09
Observations	7552	
Initial log-likelihood	-13186.714	
Final log-likelihood	-7844.414	
Adj.R^2 of the model	0.403	

Table 5.10: Model estimation results from mode choice model

Model Parameters	β	t-test	Model Parameters	β	t-test
Port 1	1.88	4.40	Port 2	1	
Port 3	4.89	2.30	Port 4	1.87	4.64
Port 5	1.93	4.45	Port 6	1.0	
Port 7	1.91	4.68	Port 8	1.20	1.84
Port1_Drive	0.250		Port1_Pax	0.125	
Port1_Walk	0.125		Port1_PT	0.125	
Port1_PNR	0.205		Port2_Pax	0.125	
Port2_Bicycle	0.250		Port2_Walk	0.125	
Port2_PT	0.125		Port3_Pax	0.125	
Port3_Motorbike	0.250		Port3_Walk	0.125	
Port3_PT	0.125		Port4_Drive	0.250	
Port4_Pax	0.125		Port4_Bicycle	0.250	
Port4_Walk	0.125		Port4_PT	0.125	
Port4_PNR	0.250		Port5_Drive	0.250	
Port5_Pax	0.125		Port5_Motorbike	0.250	
Port5_Walk	0.125		Port5_PT	0.125	
Port5_PNR	0.250		Port6_Pax	0.125	
Port6_Bicycle	0.250		Port6_Motorbike	0.250	
Port6_Walk	0.125		Port6_PT	0.125	
Port7_Drive	0.250		Port7_Pax	0.125	
Port7_Bicycle	0.250		Port7_Motorbike	0.250	
Port7_Walk	0.125		Port7_PT	0.125	
Port7_PNR	0.250		Port8_Pax	0.125	
Port8_Walk	0.125		Port8_PT	0.125	

The results from this GNL model are in line with the results from the model based on portfolio choice. While calculating the nesting parameter, based on several model estimations, the value of nest (portfolio) 2 was close to 1. Thus, I fixed the value to be 1. All other alphas were fixed such that they were equally represented in all nests. For example, passenger mode is present in all eight portfolios and hence its α fixed to be (1/8) 0.125, whereas since car mode is in four of the portfolios, its α is fixed to be 0.25. This model says the logsum of PNR lot is found to be significant here, whereas it was not significant in the previous model. Further, from the αs of portfolios it is seen that the nesting made sense. The exception for portfolios 2 and 6 can be interpreted as portfolio 2 is the one that is enabled by bicycle and portfolio 6 is the one that is enabled by both bicycle and motorbike. It was seen in the data that the majority of households own bicycles own a car also.

Though both models gave similar results on the mode choice behaviour of travellers, the latter one is clearly more straightforward. Also, the way the availability of an option is defined is close to reality in the latter one; alternatives are considered to be available if they are actually available to travellers (not just as a resource in the household but, available geographically also).

5.6 Conclusions

In this chapter I presented two mode choice models which I developed. In the first one, the alternatives are the portfolios (set of modes) which are enabled by certain resources owned by the household like car, bicycle, and motorbike. In the second one, the alternatives are the modes that are observed in the study area, and portfolios are the nests (based on the resources holdings) that contain alternatives enabled by those resources.

Apart from conforming to some established results (results from other mode choice models in SEQ, these models also shed light on the holding of resources. These results suggest an effect of long term household decisions like holding a resource on the everyday short term decision of choosing a mode to work. In terms of the modelling framework, I showcased a simple and straightforward model as an alternative to the portfolio based model (Le Vine et al., 2013). The chapter concludes noting that the unobserved ‘perceived activity set’ of travellers can be captured using a simple GNL model.

This chapter takes inputs from the results from earlier two chapters, chapter 3 and chapter 4. As mentioned in figure 1.2, the dynamic utilization of PNR lots are used as a variable in the PNR lot choice model. The logsum from the PNR lot choice model is used a a variable related to PNR mode in the mode choice model. The mode choice model thus gives an overall understanding of PNR travel behaviour since it incorporates the temporal and spatial dynamics of PNR travel behaviour too.

6 Conclusions

6.1 Summary of findings

In this dissertation I have presented the work that I have completed as part of my PhD, in the analysis of demand at PNR facilities using a series of discrete choice models. The research gaps and findings from this research are summarized in the table 6.1.

Table 6.1: Overall summary from the results

Gap in existing research	This research	Findings
-Research need to explore the temporal aspects of PNR	-Discrete time hazard model for parking space occupancy	-PNR utilisation is dynamic in nature
-Research need to explore reasons for under-and-over utilisation of PNR lots -Research need to explore different decision paradigm/framework to understand lot choice behaviour	-RUM and RRM models for PNR lot choice	-Understanding of the attributes that effect the choice of PNR lots -PNR lot choice behaviour can be explained by RRM framework too
-Endogeniety was not discovered in PNR travel models	-Correction for endogeneity in PNR lot choice model using CF method	-PNR lot utilisation is endogenous in nature -Endogeniety in lot choice model can be corrected using Control function method
-Lack of mode choice model conditoned on resource holdings in Australia	-Household resources/portfolios based model and GNL model for mode choice modelling	-Mode choice decisions are influenced by a household resource holdings -GNL model is straightforward and provides an alternative to the portfolio based model

While exploring the dynamic nature of *utilization*, from the developed hazard model, some interesting insights were found: the distance from parking space to platform matters more than the distance of the PNR lot from the CBD. This finding sheds light on an important aspect of PNR lots: the parking lot design. A lot of attention is given to optimising the location of PNR lots;

these findings suggest that the PNR lot design deserves attention as well.

Further, in an attempt to understand the mode choice decisions of travelers, I developed two set of models; one where alternatives were portfolios and another where they were modes. The idea behind the models is that a person with a car (or other resources) in a household perceives the overall choice set differently to that of a person who does not hold any resources. Modelling in this framework allowed an understanding of the underlying factor of not holding a resource (a long term decision) on the choice of mode (a short term decision). Further, I proposed a simple and straightforward method of using the GNL framework as an alternative to the portfolio based model that is much simpler to compute.

6.2 Contributions

The aim of this thesis has been to understand PNR travel behaviour based on different times-of-day during a morning peak period. Following the aim, the path taken was to understand PNR lot choice behaviour, to understand and analyse the dynamic utilization of PNR lots, and to analyse the mode choice behaviour of users, including PNR as one of the modes. The overall findings suggest that PNR demand is dynamic in nature. The time-of-day affects the *utilization* of PNR lots and which PNR lot is chosen. This is one of the main contributions of this research.

The identification of *utilization* as an endogenous variable is performed for the first time. Correction of endogeneity in the case of discrete choice transport models is a relatively new area. Specifically, the use of the control function is limited to a handful of studies on endogeneity. In this scenario, this work serves as additional evidence of the value of correcting for endogeneity using the control function method.

This research explored the choice mechanism of PNR users beyond the commonly used RUM concept and establishes that PNR users' choice of PNR lots could be better explained by the RRM concept. This research represents a milestone in PNR demand analysis.

By making use of portfolios as nests in studying the mode choice behaviour of travellers, this research addresses the long term decisions of a household (like holding car, bicycle etc) on making short term everyday decisions like mode choice. This mechanism makes it possible for researchers to address the effect of underlying factors (holding a resource) on mode choice, even for a trip based model.

6.3 Policy Implications

The findings from this research could be of high interest to policy makers. One of the biggest challenges in PNR is to understand why travelers choose one PNR versus another. This has been very well addressed by this research. Additionally, understanding that PNR users are constantly making comparisons among various PNR lots in terms of every aspect of PNR, and choosing one based on least regret could be another finding of interest to policy makers. Overall, the results could be used by PNR planners on better management parking lots within the morning period.

An important finding from this research is that the location of parking spaces within PNR lots (i.e., design of PNR lots) plays an important role alongside the location of PNR lots in a city. This sheds light on an often neglected area of PNR lot design. This finding further supports the possibilities of charging user fees based on the location of the parking space also. Having understood the role of fixed long term decisions on everyday short term travel decisions, policy makers could focus on acknowledging travellers' long term household decisions.

6.4 Future Directions

The addition of knowledge from this research existing PNR literature has opened up avenues for a wide variety of future work. Some obvious extensions could be the use of tour based models and activity based models, in place of trip based models to understand the role of long term decisions on short term everyday mode choice decisions. While using such tour based models, it might also be interesting to study the departure of cars from PNR lots. Future work could also include a route choice model for PNR trips (and tours). Apart from these, research can be further extended to using different decision heuristics apart from RUM and RRM while exploring PNR lot choice behaviour. In addition, the role of parking fees in PNR lots and in the choice of lots could be explored too. In the area of modelling technique, correction of endogeneity could benefit from recent developments in the area of endogeneity correction for discrete choice models (Guevara, 2015).

By establishing that PNR lot behaviour is different to that of general lots, this research opened up whole new areas of study such as the design of PNR lot, parking search mechanism in PNR lots, etc. In these areas, the study of PNR lots throughout a day, including arrival and departures and parking durations, could be investigated and could help in understanding PNR tours. Along these lines, it would be interesting to explore the role of parking fees at PNR lots where parking spaces have a time- and/or location-dependent parking tariff.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13:61–98.

Arnott, R., De Palma, A., and Lindsey, R. (1991). A temporal and spatial equilibrium analysis of commuter parking. *Journal of Public Economics*, 45:301–335.

Ben-Akiva, M. and Swait, J. (1986). The akaike likelihood ratio index. *Transportation Science*, 20:133–136.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*, volume 9. MIT press.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Bierlaire, M. (2003). Biogeme: a free package for the estimation of discrete choice models. In *Swiss Transport Research Conference*, pages TRANSP–OR–CONF–2006–048.

Bierlaire, M. (2006). A general formulation of the cross-nested logit model. In *Swiss Transport Research Conference*, pages TRANSP–OR–CONF–2006–054.

Bierlaire, M., Axhausen, K., and Abay, G. (2001). The acceptance of modal innovation: The case of swissmetro. In *Proceedings of the 1st Swiss Transportation Research Conference*.

Bolger, D., Colquhoun, D., and Morrall, J. (1992). Planning and design of park-and-ride facilities for the Calgary light rail transit system. *Transportation Research Record*, 1361:141–141.

Bos, I., Molin, E., Timmermans, H., and van der Heijden, R. (2003). Cognition and relative importance underlying consumer valuation of park-and-ride facilities. *Transportation Research Record: Journal of the Transportation Research Board*, 1835:121–127.

Bos, I. D., Van der Heijden, R., Molin, E. J., and Timmermans, H. J. (2004). The choice of park and ride facilities: An analysis using a context-dependent hierarchical choice experiment. *Environment and Planning A*, 36:1673.

Caicedo, F., Robuste, F., and Lopez-Pita, A. (2006). Parking management and modeling of car park patron behavior in underground facilities. *Transportation Research Record: Journal of the Transportation Research Board*, 1956:60–67.

Charles-Edwards, E., Bell, M., and Corcoran, J. (2015). Greening the commute: Assessing the impact of the eleanor schonell green bridge on travel to the university of queensland, australia. *Urban Policy and Research*, 33:61–78.

Chen, C. and Xie, Y. (2014). Modeling the safety impacts of driving hours and rest breaks on truck drivers considering time-dependent covariates. *Journal of Safety Research*, 51:57–63.

Chen, Z., Xia, J. C., Irawan, B., and Caulfied, C. (2014). Development of location-based services for recommending departure stations to park and ride users. *Transportation Research Part C: Emerging Technologies*, 48:256–268.

Chorus, C. (2012). Random regret minimization: An overview of model properties and empirical evidence. *Transport Reviews*, 32:75–92.

Chorus, C. G. (2010). A new model of random regret minimization. *European Journal of Transport and Infrastructure Research (EJTIR)*, 2:181–196.

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

Chorus, C. G. and Bierlaire, M. (2013). An empirical comparison of travel choice models that capture preferences for compromise alternatives. *Transportation*, 40:549–562.

Choudhury, C. F., Yang, L., e Silva, J. d. A., and Ben-Akiva, M. (2018). Modelling preferences for smart modes and services: A case study in lisbon. *Transportation Research Part A: Policy and Practice*, 115:15–31.

Coffel, K., Parks, J., Semler, C., Ryus, P., Sampson, D., Kachadoorian, C., Levinson, H., and Schofer, J. (2012). TCRP report 153: Guidelines for providing access to public transportation stations. *Trans—portation Research Board of the National Academies, Washington, DC*.

Cox, D. R. (1992). Regression models and life-tables. In *Breakthroughs in Statistics*, pages 527–541. Springer.

Debrezion, G., Pels, E., and Rietveld, P. (2007). Choice of departure station by railway users. *European Transport\Trasporti Europei*, 37:78–92.

Engel-Yan, J., Rudra, M., Livett, C., Nagorsky, R., Engel, J., and Nagorsky, B. (2014). Strategic station acess plannig for commuter rail: Balancing park and ride with other modes. In *Transportation Research Board 93rd Annual Meeting*, pages 14–1374.

Fernandez-Antolin, A., Stathopoulos, A., and Bierlaire, M. (2014). Exploratory analysis of endogeneity in discrete choice models. In *14th Swiss Transport Research Conference*, number EPFL-CONF-202581.

Ferreira, F. (2010). You can take it with you: Proposition 13 tax benefits, residential mobility, and willingness to pay for housing amenities. *Journal of Public Economics*, 94:661–673.

Foote, P. (2000). Chicago transit authority weekday park-and-ride users: Choice market with ridership growth potential. *Transportation Research Record: Journal of the Transportation Research Board*, 1735:158–168.

Government, Q. (2014). South East Queensland park and ride strategy.

Guevara, C. and Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record: Journal of the Transportation Research Board*, 1977:60–66.

Guevara, C. A. (2010). Endogeneity and sampling of alternatives in spatial choice models. *Unpublished doctoral dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA*.

Guevara, C. A. (2015). Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A: Policy and Practice*, 82:240–254.

Guevara, C. A. (2017). Mode-valued differences of in-vehicle travel time savings. *Transportation*, 44:977–997.

Guevara, C. A. and Polanco, D. (2016). Correcting for endogeneity due to omitted attributes in discrete-choice models: The multiple indicator solution. *Transportmetrica A: Transport Science*, 12:458–478.

Hahn, J. and Hausman, J. (2003). Weak instruments: Diagnosis and cures in empirical econometrics. *The American Economic Review*, 93:118–125.

Hamer, P. (2010). Analysing the effectiveness of park and ride as a generator of public transport mode shift. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, 19:51.

Hamid, N. A. (2009). Utilization patterns of park and ride facilities among Kuala Lumpur commuters. *Transportation*, 36:295–307.

Hendricks, S. and Outwater, M. (1998). Demand forecasting model for park-and-ride lots in King County, Washington. *Transportation Research Record: Journal of the Transportation Research Board*, 1623:80–87.

Hensher, D. A., Greene, W. H., and Chorus, C. G. (2013). Random regret minimization or random utility maximization: An exploratory analysis in the context of automobile fuel choice. *Journal of Advanced Transportation*, 47:667–678.

Hensher, D. A., Greene, W. H., and 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.

Huang, Y., Smith, B., and Olaru, D. (2016). Heterogeneous substitution patterns based on current mode and time of departure choices in the context of uncertainty. In *Australasian Transport Research Forum (ATRF), 38th, 2016, Melbourne, Victoria, Australia*.

Katoshevski, R., Glickman, I., Ishaq, R., and Shiftan, Y. (2015). Integrating activity-based travel-demand models with land-use and other long-term lifestyle decisions. *Journal of Transport and Land Use*, 8.

Keiley, M. K., Martin, N. C., Canino, J., Singer, J., and Willett, J. (2007). Discrete-time survival analysis: Predicting whether, and if so when, an event occurs. *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, pages 441–463.

Khan, O. A., Kruger, J., and Trivedi, T. (2007). Developing passenger mode choice models for Brisbane to reflect observed travel behaviour from the South East Queensland travel survey. *Proceedings of 30th Australasian Transport Research Forum, Melbourne, Australia*.

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

Kleinbaum, D. G. and Klein, M. (2006). *Survival analysis: A self-learning text*. Springer Science & Business Media.

Lai, L. and Shalaby, A. S. (2007). Cellular automaton simulation of vehicle dynamics in park-and-ride facilities. In *Transportation Research Board 86th Annual Meeting*, volume 07-1755.

Le Vine, S., Lee-Gosselin, M., Sivakumar, A., and Polak, J. (2013). A new concept of accessibility to personal activities: Development of theory and application to an empirical study of mobility resource holdings. *Journal of Transport Geography*, 31:1–10.

Li, H., Gao, K., and Tu, H. (2017). Variations in mode-specific valuations of travel time reliability and in-vehicle crowding: Implications for demand estimation. *Transportation Research Part A: Policy and Practice*, 103:250–263.

Li, Z.-C., Lam, W., Wong, S., Zhu, D.-L., and Huang, H.-J. (2007). Modeling park-and-ride services in a multimodal transport network with elastic demand. *Transportation Research Record: Journal of the Transportation Research Board*, 1994:101–109.

Lin, T. G., Xia, J. C., Robinson, T. P., Goulias, K. G., Church, R. L., Olaru, D., Tapin, J., and Han, R. (2014). Spatial analysis of access to and accessibility surrounding train stations: A case study of accessibility for the elderly in perth, western australia. *Journal of Transport Geography*, 39:111–120.

Liu, X., Yun, M., Chen, Z., and Yang, X. (2012). Investigation, analysis, and modeling of choice behavior of park and ride. pages 1643–1651.

Mahmoud, M. S., Habib, K. N., and Shalaby, A. (2014). Park-and-ride access station choice model for cross-regional commuter trips in the Greater Toronto and Hamilton Area (GTHA). In *Transportation Research Board 93rd Annual Meeting*, pages 14–3243.

McFadden, D. (1978). *Modelling the choice of residential location*. Institute of Transportation Studies, University of California.

Miller, E. and Roorda, M. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, 1831:114–121.

Mingardo, G. (2013). Transport and environmental effects of rail-based park and ride: Evidence from the netherlands. *Journal of Transport Geography*, 30:7–16.

Nassir, N., Khani, A., Hickman, M., and Noh, H. (2012). Algorithm for intermodal optimal multidestination tour with dynamic travel times. *Transportation Research Record: Journal of the Transportation Research Board*, 2283:57–66.

Nerella, S. and Bhat, C. R. (2004). Numerical analysis of effect of sampling of alternatives in discrete choice models. *Transportation Research Record: Journal of the Transportation Research Board*, 1894:11–19.

Nurul Habib, K. M., Morency, C., and Trépanier, M. (2012). Integrating parking behaviour in activity-based travel demand modelling: Investigation of the relationship between parking type choice and activity scheduling process. *Transportation Research Part A: Policy and Practice*, 46:154–166.

Olaru, D., Smith, B., Xia, J. C., Chen, C., Lin, T., Ying, H., and Han, R. (2014a). Ticket to ride?: factors affecting park-and-ride travel in perth, wa. In *AET Papers Repository*. Transport Research Society.

Olaru, D., Smith, B., Xia, J. C., and Lin, T. G. (2014b). Travellers? attitudes towards park-and-ride (pnr) and choice of pnr station: Evidence from perth, western australia. *Procedia-Social and Behavioral Sciences*, 162:101–110.

Parkhurst, G. (2000). Influence of bus-based park and ride facilities on users□ car traffic. *Transport Policy*, 7:159–172.

Petrin, A. and Train, K. (2003). Omitted product attributes in discrete choice models. Technical report, National Bureau of Economic Research.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47:3–13.

Qin, H., Guan, H., and Wu, Y.-J. (2013). Analysis of park-and-ride decision behavior based on decision field theory. *Transportation Research Part F: Traffic Psychology and Behaviour*, 18:199–212.

Risa Hole, A. (2004). Forecasting the demand for an employee park and ride service using commuters' stated choices. *Transport Policy*, 11:355–362.

Sharma, B., Hickman, M., and Nassir, N. (2016). A study on the utilization of park-and-ride lots in south east queensland. In *38th Australasian Transport Research Forum*.

Shirgaokar, M. and Deakin, E. (2005). Study of park-and-ride facilities and their use in the San Francisco Bay area of California. *Transportation Research Record: Journal of the Transportation Research Board*, 1927:46–54.

Singer, J. and Willett, J. (2003a). Describing discrete-time event occurrence data. *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. Oxford University Press, New York*, pages 325–356.

Singer, J. and Willett, J. (2003b). Fitting basic discrete-time hazard models. *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence. Oxford University Press, New York*, pages 357–467.

Singer, J. D. and Willett, J. B. (1993). It□s about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics*, 18:155–195.

Smith, B., Ying, H., and Taplin, J. (2016). *Public Transport Patronage Success. The Role of Park-and-ride*. Planning Boomtown and Beyond.

Syed, S., Golub, A., and Deakin, E. (2009). Response of regional rail park-and-ride users to parking price changes: Systemwide results and a detailed study of two stations. *Transportation Research Record: Journal of the Transportation Research Board*, 2110:155–162.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Translink (2016). Service updates. https://translink.com.au/.

Tsang, F. W., Shalaby, A. S., and Miller, E. J. (2005). Improved modeling of park-and-ride transfer time: Capturing the within-day dynamics. *Journal of Advanced Transportation*, 39:117–137.

Vovsha, P. (1997). Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607:6–15.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2010). *Statistical and Econometric Methods for Transportation Data Analysis*. CRC press.

Weiss, A. and Habib, K. N. (2017). Examining the difference between park and ride and kiss and ride station choices using a spatially weighted error correlation (swec) discrete choice model. *Journal of Transport Geography*, 59:111–119.

Wen, C.-H. and Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research Part B: Methodological*, 35:627–641.

Wiseman, N., Bonham, J., Mackintosh, M., Straschko, O., and Xu, H. (2012). Park and ride: An adelaide case study. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, 21:39.

Yamamoto, T., Kurauchi, S., and Morikawa, T. (2002). Comparison of non-compensatory models of driver's choice on dynamic park and ride. In *9th World Congress on Intelligent Transportation Systems, Chicago, IL, USA*.

Yang, M., Zhao, J., Wang, W., Liu, Z., and Li, Z. (2015). Metro commuters? satisfaction in multi-type access and egress transferring groups. *Transportation Research Part D: Transport and Environment*, 34:179–194.

Yushimito, W. F., Aros-Vera, F., Reilly, J. J., et al. (2012). User rationality and optimal park-and-ride location under potential demand maximization. *Transportation Research Part B: Methodological*, 46:949–970.

Zijlstra, T., Vanoutrive, T., and Verhetsel, A. (2015). A meta-analysis of the effectiveness of park-and-ride facilities. *European Journal of Transport & Infrastructure Research*, 15.

Appendix A Appendix

Table A.1: Distribution of time intervals of filling 7590 parking spaces

Interval	Risk Set	EVENT=1 (failed)	EVENT=0 (censored)	Estimated Hazard Rate	Estimated Survival Probability
0	7590	0	7590	0	1
1	7590	5	7585	0.001	0.999
2	7585	3	7582	0	0.999
3	7582	5	7577	0.001	0.998
4	7577	6	7571	0.001	0.997
5	7571	9	7562	0.001	0.996
6	7562	4	7558	0.001	0.996
7	7558	19	7539	0.003	0.993
8	7539	30	7509	0.004	0.989
9	7509	33	7476	0.004	0.985
10	7476	27	7449	0.004	0.981
11	7449	27	7422	0.004	0.978
12	7422	24	7398	0.003	0.975
13	7398	48	7350	0.006	0.968
14	7350	46	7304	0.006	0.962
15	7304	55	7249	0.008	0.955
16	7249	59	7190	0.008	0.947
17	7190	59	7131	0.008	0.940
18	7131	48	7083	0.007	0.933
19	7083	60	7023	0.008	0.925
20	7023	69	6954	0.01	0.916
21	6954	99	6855	0.014	0.903
22	6855	87	6768	0.013	0.892
23	6768	114	6654	0.017	0.877
24	6654	117	6537	0.018	0.861
25	6537	127	6410	0.019	0.845
26	6410	126	6284	0.002	0.828
27	6284	162	6122	0.026	0.807
28	6122	177	5945	0.029	0.783
29	5945	185	5760	0.031	0.759
30	5760	173	5587	0.03	0.736
31	5587	269	5318	0.048	0.701
32	5318	218	5100	0.041	0.672
33	5100	301	4799	0.059	0.632
34	4799	285	4514	0.059	0.595
35	4514	330	3530	0.085	0.465
36	4184	324	3860	0.077	0.509
37	3860	330	3530	0.085	0.465
38	3530	304	3226	0.086	0.425
39	3226	298	2928	0.092	0.386
40	2928	294	2634	0.1	0.347
41	2634	310	2324	0.118	0.306
42	2324	246	2078	0.106	0.274
43	2078	266	1812	0.128	0.239
44	1812	207	1605	0.114	0.211
45	1605	217	1388	0.135	0.183
46	1388	135	1253	0.097	0.165
47	1253	114	1139	0.091	0.150
48	1139	82	1057	0.072	0.139

Table A.2: Distribution of time intervals of filling 7590 parking spaces (..continued)

Interval	Risk Set	EVENT=1 (failed)	EVENT=0 (censored)	Estimated Hazard Rate	Estimated Survival Probability
49	990	81	976	0.082	0.128
50	909	49	860	0.054	0.121
51	860	48	812	0.056	0.114
52	812	44	768	0.054	0.108
53	768	27	741	0.035	0.104
54	741	23	718	0.031	0.101
55	718	33	685	0.046	0.096
56	600	20	665	0.030	0.093
57	550	23	617	0.042	0.090
58	344	15	512	0.044	0.086
59	314	9	320	0.029	0.083
60	219	12	293	0.055	0.079
61	207	6	201	0.029	0.076
62	201	5	196	0.025	0.074
63	196	4	192	0.020	0.073
64	192	5	187	0.026	0.071
65	187	1	186	0.005	0.071
66	186	7	179	0.038	0.068