Modelling the Effect of the Number of Stop-&-gos on the Route Choice Behaviour of Car Drivers

Author:
Saxena, Neeraj

Publication Date:
2017

DOI:
https://doi.org/10.26190/unsworks/20379

Modelling the Effect of the Number of Stop-&-gos on the Route Choice Behaviour of Car Drivers

By

Neeraj Saxena

M. Tech. (Transportation), B. Tech. (Civil)

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

School of Civil and Environmental Engineering

Faculty of Engineering

The University of New South Wales

August 2017

ॐ

जननी जन्मभूमिश्च स्वर्गादपि गरीयसी ।

Mother and Motherland are superior to Heaven.

गुरु गोबिन्द दोउ खडे काके लागूँ पाँय ।

बलिहारी गुरु आपने गोबिन्द दियो बताय ॥

Guru and God both are here to whom should I first bow. All glory be unto the guru path to God who did bestow.

This thesis is dedicated to my

Parents, Gurus and Motherland

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Saxena**

First name: **Neeraj** Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Civil and Environmental Engineering** Faculty: **Faculty of Engineering**

Title: **Modelling the effect of the number of stop-&-gos on the route choice behaviour of car drivers**

Abstract 350 words maximum: (PLEASE TYPE)

Driving in stop-&-go (S&G) traffic can be a frustrating experience and can lead to an increase in fuel emissions, driver distress and safety risks. Past studies have modelled the time spent in S&G traffic as one of the factors for the disutility of a route. These studies did not account for the effect of the number of S&G waves, which has been found to bear a closer relationship with driver discomfort. This thesis tests the research hypothesis that an increase in the number of S&Gs on a route increases its disutility for a driver.

As a proof of concept, a Stated Choice (SC) experiment was initially conducted on a sample of university staff and students. A D-efficient pivot design technique was used to generate the set of choice tasks. The competing routes were defined in terms of the number of S&Gs along with other travel attributes. The collected data was analysed using a Random Parameter Error Component Logit model to account for the taste heterogeneity and serial correlation among the choice tasks. The results showed that the number of S&Gs negatively impacted the utility of a route, thus validating the proposed hypothesis.

Another follow-up SC study was then conducted on a sample of general commuters in Sydney. The data from this study was analysed using a Latent Class Choice Model which relaxed the distributional assumption associated with the previous statistical model. The obtained results showed a negative effect for the number of S&Gs, except for nearly one quarter of the sample which was indifferent towards this attribute.

Lastly, a driving simulator experiment was conducted to further understand the role of S&G traffic characteristics on driver frustration. Participants were made to drive through virtual scenarios depicting varying S&G traffic conditions. A Structural Equation Model was estimated on the data which indicated that the number of S&Gs had a positive effect on the level of frustration and was also observed to influence route choice.

The findings from this thesis not only extend the body of knowledge on intricate route choice behaviour of drivers, but also inform policies aimed at reducing traffic congestion and the resulting S&G traffic.

28/08/2017

Signature Witness Signature Date

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

ORIGINALITY STATEMENT

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

Signed …… ………

Date …… 28/08/2017 ……..

COPYRIGHT STATEMENT

AUTHENTICITY STATEMENT

Modelling the Effect of the Number of Stop-&-gos on the Route Choice Behaviour of Car Drivers

Neeraj Saxena, Ph.D.

The University of New South Wales Sydney

Supervisors:

Prof. S. Travis Waller Assoc. Prof. Vinayak V. Dixit

ABSTRACT

Driving in stop-&-go (S&G) traffic can be a frustrating experience and can lead to an increase in fuel emissions, driver distress and safety risks. Past studies have modelled the time spent in S&G traffic as one of the factors for the disutility of a route. These studies did not account for the effect of the number of S&G waves, which has been found to bear a closer relationship with driver discomfort. This thesis tests the research hypothesis that an increase in the number of S&Gs on a route increases its disutility for a driver. As a proof of concept, a Stated Choice (SC) experiment was initially conducted on a sample of university staff and students. A D-efficient pivot design technique was used to generate the set of choice tasks. The competing routes were defined in terms of the number of S&Gs along with other travel attributes. The collected data was analysed using a Random Parameter Error Component Logit model to account for the taste heterogeneity and serial correlation among the choice tasks. The results showed that the number of S&Gs negatively impacted the utility of a route, thus validating the proposed hypothesis. Another follow-up SC study was then conducted on a sample of general commuters in Sydney. The data from this study was analysed using a Latent Class Choice Model which relaxed the distributional assumption associated with the previous statistical model. The obtained results showed a negative effect for the number of S&Gs, except for nearly one quarter of the sample which was indifferent towards this attribute. Lastly, a driving simulator experiment was conducted to further understand the role of S&G traffic characteristics on driver frustration. Participants were made to drive through virtual scenarios depicting varying S&G traffic conditions. A Structural Equation Model was estimated on the data which indicated that the number of S&Gs had a positive effect on the level of frustration and was also observed to influence route choice. The findings from this thesis not only extend the body of knowledge on intricate route choice behaviour of drivers, but also inform policies aimed at reducing traffic congestion and the resulting S&G traffic.

ACKNOWLEDGEMENTS

As the day approaches when I submit this once in a lifetime document, the Ph.D. thesis, I like to acknowledge the people whose continuous love, support and prayers have made me see this day. I first thank my thesis supervisors Professor S. Travis Waller and Associate Professor Vinayak V. Dixit for giving me the flexibility to work on something which I loved and the encouragement to keep on exploring my interests. I also thank them for their tremendous support and faith that they have kept on me. I also thank Dr. Taha Hossein Rashidi, Senior Lecturer, for igniting the passion towards statistics and econometrics in me. I would also like to thank my friend Subodh Kant Dubey for his mentoring in econometrics because of which I was able to code the different statistical models used in this thesis. I also thank Dr. Lauren Gardner and Dr. David Rey for giving valuable feedback during my candidature. I immensely thank the other staff and fellow Ph.D. colleagues at my research centre (rCITI, UNSW Sydney) for their support, especially during the driving simulator experiment. Right from Sai and Ted, who thoroughly tested the experiment, to Maria, Kasun and many more, who helped me get participants. I got this very special feeling seeing everyone trying equally hard like me to make this study a success. I thank Ted, my sisters (Divya and Nidhi), brother (Anuj) and uncle for proofreading this thesis.

I am grateful to my friends from my research lab, LG#09, who not only rejoiced with me in celebrations, but more importantly, not letting me surrender to situations whenever going became tough. I would first thank Divya for being a lovely sister and a wonderful friend to me. Ted and Sai for their valuable advice, both on professional and personal front. Hassan, Raed, Abdul, Shantanu along with others for always maintaining a cheerful and positive environment in our lab. I will always remember our outings, particularly to El Jannah, whose chicken kept me motivated to go through every week quickly. I am so thankful to Divya and Prasanna, and Alex and Deepa for treating me as their own family while I was miles away from my own. Of course, how can I forget to thank my little friends and packets of energy, Isha and Devvance.

I thank all my teachers, especially Sir R. K. Sharma, for instilling passion towards studies in me. He truly exemplifies that it takes just one good teacher to transform one’s life. I also thank my big family for always praying for my success. Particularly, my mummy and papa for their blessings and all the sacrifices they have made so that I can fulfil my dream. I thank all my school and college friends for giving me some unforgettable moments. I am grateful to my motherland, India, for giving me my identity. I also thank the people of Australia for respecting my background and ethnicity and giving me a healthy atmosphere to focus on my thesis. I would like to acknowledge the teachings from our holy scripture, the *Bhagavad Gita*, which gave me strength during hard times. Lastly, I thank the supreme power and the creator of this universe for listening to my prayers and making me see this day.

I, Neeraj Saxena, date 4th August, 2017, would like to thank each one of you from the bottom of my heart. This thesis would not have been possible without your love and support. I am so blessed to have you around.

(Neeraj Saxena)

LIST OF RELEVANT PUBLICATIONS AND AWARDS

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

Publications in Conference Proceedings

- **Saxena, N.**, Rashidi, T.H., Dixit, V.V. & Waller, S.T. (2017). Influence of Number of Stop-&-Gos on the Route Choice Behavior of Car Drivers in Urban Road Networks. *Proceedings of the 96th Transportation Research Board Annual Meeting*, January 2017, Washington D.C., USA.
- **Saxena, N.**, Dixit, V.V. & Waller, S.T. (2016). Techniques for Identifying the Occurrence of Stop-&-go Waves in Traffic: A Literature Review. *Proceedings of the 38th Australasian Transport Research Forum,* Melbourne, AU.

Papers under Review in Peer-reviewed Journals

- **Saxena, N.**, Rashidi, T.H., Dixit, V.V. & Waller, S.T. Modelling the Route Choice Behaviour under Stop-&-go Traffic for Different Car Driver Segments. In review with *Transportation Research Part A: Policy and Practice*.
- **Saxena, N.**, Rashidi, T.H., Dixit, V.V. & Waller, S.T. Effect of Number of Stop-&-gos Experienced on the Route Choice Behaviour of Car Drivers. In review with *Transportmetrica A: Transport Science*.

Awards

- Winner of the Best Paper Award at the Transportation Planning and Implementation Methodologies for Developing Countries conference held in Mumbai, India
- Recipient of the School of Civil and Environmental Engineering Prize at the Post Graduate Research Symposium held within the Faculty of Engineering at UNSW Sydney, Australia
- Honourable Mention for the David Willis Best Poster prize at the Australasian Transport Research Forum held in Sydney, Australia

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The paramount objective of this chapter is to introduce the research project and set the tone of the thesis. The first few sections in this chapter provide the context of the research problem and the motivation behind taking up this topic for research. The aims and objectives are defined next which is then followed by the potential contributions of this research work. The chapter concludes with the organisation of thesis where the contents of each contributing chapters are outlined.

1.1 Context to the Research Problem

Transportation infrastructure constitutes one of the key drivers of economic growth and prosperity of a region. Statistics from Infrastructure Australia (2015) show that the transportation sector, as a whole, contributed nearly 10 percent towards the GDP of Australia in the year 2011. Interestingly, the share of urban roadway infrastructure was more than half the overall contribution. The urban road network in Australia caters to nearly 70 percent of all passenger and a significant proportion of non-bulk domestic freight movement (BITRE, 2014). Due to growing demand on roads, major cities around the world are experiencing unprecedented traffic congestion problems. Traffic congestion impairs economic growth due to loss of person hours, fuel emissions, other environmental hazards, and safety risks. For example, the annual congestion cost, which includes fuel cost and monetary value of lost time, for Australia in the year 2015 was roughly 16.5 billion dollars (AU) (BITRE, 2015). The cost was much higher in the US, of the order of 120 billion dollars (US) per year (The White House, 2014).

Government agencies around the world are striving hard to alleviate the menace caused due to traffic congestion. One traditional and a widely used solution has been to add extra capacity to the existing roadway infrastructure. For example, the US government recently sanctioned 302 billion dollars (US), spread across four years, to build and maintain roadway infrastructure aiming to bring down congestion (The White House, 2014). The government of Australia has also proposed to invest 15 billion dollars (AU) towards key roadway development and upgradation projects like the WestConnex, Pacific highway upgrade, etc. in the state of New South Wales (NSW). Additionally, the Australian government also earmarked around 25 million dollars (AU) in 2011 to fund major studies to tackle the problem of congestion (BITRE, 2011).

In order to effectively introduce measures aimed at relieving traffic congestion, it is crucial to first understand its dynamics and impacts on driver behaviour. A vast literature exists on car driver behaviour under congested traffic and the resulting de-valuation towards such conditions. However, a peculiar aspect of congested traffic, namely stop-&-go traffic, has not been studied in depth to date. It is important to thoroughly study the impact of stop-&-go waves on drivers to gain better insights into the traffic congestion problem. This will aid transport agencies to allocate budget towards measures that are effective in reducing the ill-effects of traffic congestion and the consequent stop-&-go waves.

1.2 What is a Stop-&-go Wave

A stop-&-go (S&G) wave, also called traffic oscillation or phantom jam, is a traffic phenomenon that often exists in urban road networks during congestion. Shott (2011) defines S&G traffic as a condition where vehicles are forced to decelerate and travel at a lower speed, or even come to a halt, before resuming their original speeds. These waves were first observed inside the Lincoln tunnel in the US by Edie (1961). The study observed significant speed fluctuations on a one lane traffic stream caused due to small perturbations. Li et al. (2010) found S&G waves to occur in a cyclic pattern alternating between slow (stop) and fast (go) movements. Tanaka & Nakatsuji (2011) expressed S&G waves as a complete cycle of acceleration from a stop, travelling at different speeds in a short distance, and deceleration to a complete stop. Zheng et al. (2011) defined one cycle of S&G as deceleration followed by an acceleration of the vehicle. Empirical studies by Ahn et al. (2004), Laval et al. (2009), Li et al. (2010), Mauch & Cassidy (2004) found that S&G waves generally repeat in intervals of 2-15 minutes, last for up to 30 seconds and propagate backwards at a wave speed between 10 to 20 km/h on freeways.

Multiple reasons can lead to the initiation of S&G waves in congested traffic. A primary reason for that is asymmetric driver behaviour (Edie & Baverez, 1965). Yeo & Skabardonis (2009) analysed individual vehicle trajectory data to observe the variability in driver behaviour under different situations. They found heterogeneity in the measurement error across drivers with regard to judging the speed of the leader vehicle. Interestingly, their study revealed that drivers who overreact to braking situations often initiate S&G waves. Similar observations were also made by Laval & Leclercq (2010) who classified drivers into timid and aggressive on the basis of behaviour. Thus, the presence of few such drivers in a congested traffic stream could potentially instigate widespread S&G conditions. Apart from that, S&G waves can also be triggered due to reasons like:

- lane change manoeuvres (Ahn & Cassidy, 2007; Laval, 2007),
- any kind of moving bottleneck in a traffic stream (Koshi et al., 1992; Laval, 2007),
- a drop in the roadway capacity (Bertini & Leal, 2005; Cassidy & Bertini, 1999a,b),
- different roadway geometric features like curves and uphill segments (Jin & Zhang, 2005).

The aforementioned factors of initiation of S&G waves were mainly observed using the available vehicle trajectory data on freeways. Few other factors that can potentially cause the formation of these waves in urban road networks are:

- capacity drop due to the presence of on-street parking (Wijayaratna & Wijayaratna, 2016)
- presence of signalised intersections, roundabouts, etc.

Like traffic congestion, S&G waves also have an adverse effect on network performance and the surrounding environment. They not only cause an increase in fuel emissions (Helbing, 1997), but also lead to safety risks. Moreover, numerous studies have found that continuous exposure to S&G traffic can lead to detrimental effects on driver physiology, particularly the cardiovascular measurements like blood pressure and heart rate (Apparies et al. 1998; Yang et al. 2013). Levinson et al. (2004) showed in their study that S&G traffic results in heightened discomfort and frustration levels among drivers as they need to be more focussed while driving. Malta et al. (2011) also found drivers to experience elevated levels of frustration when subjected to recurring cycles of S&G. This built up frustration is often the precursor to aggressive driving behaviour which can pose a serious threat to on-road safety (Lee, 2010).

Given the demerits of S&G traffic, it can thus be inferred that drivers tend to associate a disutility towards a route where S&G traffic is prevalent. This intuition fuelled the motivation behind this research topic, which is described in the following section.

1.3 Motivation

The motivation behind this research idea can be understood through a hypothetical toy network shown in figure 1.1. The figure comprises an origin (O), destination (D) and two travel routes (route-I and route-II) between the OD pair. Consider the two routes to have a similar travel distance and cost, but different prevailing traffic conditions. Let the traffic condition be expressed in terms of two attributes: travel time and time spent in S&G. For simplicity, let us also assume that the travel time on both routes is the same, which is 20 minutes. Thus, in figure 1.1, the only distinguishing factor between route 1 and 2 is the time spent in S&G attribute.

A few studies have modelled the route choice decisions of car drivers as a function of time spent in S&G traffic. For example, Hensher (2001a) in his study found that car drivers, on average, perceived each minute under S&G traffic equivalent to 2.5 minutes of free flow time. Using this information, we can evaluate the disutility associated with each route given in figure 1.1. For the sake of brevity, let us express the disutility in terms of the total perceived time as shown in equations 1.1 and 1.2:

$$Route - I: (20 - 10) + 10 \times 2.5 = 35 \text{ minutes} \tag{1.1}$$

$$Route - II: (20 - 8) + 8 \times 2.5 = 32 \text{ minutes} \quad (1.2)$$

In the two equations, the first term represents the time spent in free flow traffic while the second term represents the perceived time under S&G traffic. Since route-II has a lower disutility, there should be more drivers on this route at network equilibrium. In other words, more car trips should be assigned to route-II at user equilibrium, where no driver can gain on travel time savings by unilaterally switching to another route (Wardrop, 1952). However, previous studies did not consider the effect of the number of S&Gs which is also expected to impact a route's disutility along with the time spent in such conditions, as indicated by Levinson et al. (2004) and Malta et al. (2011).

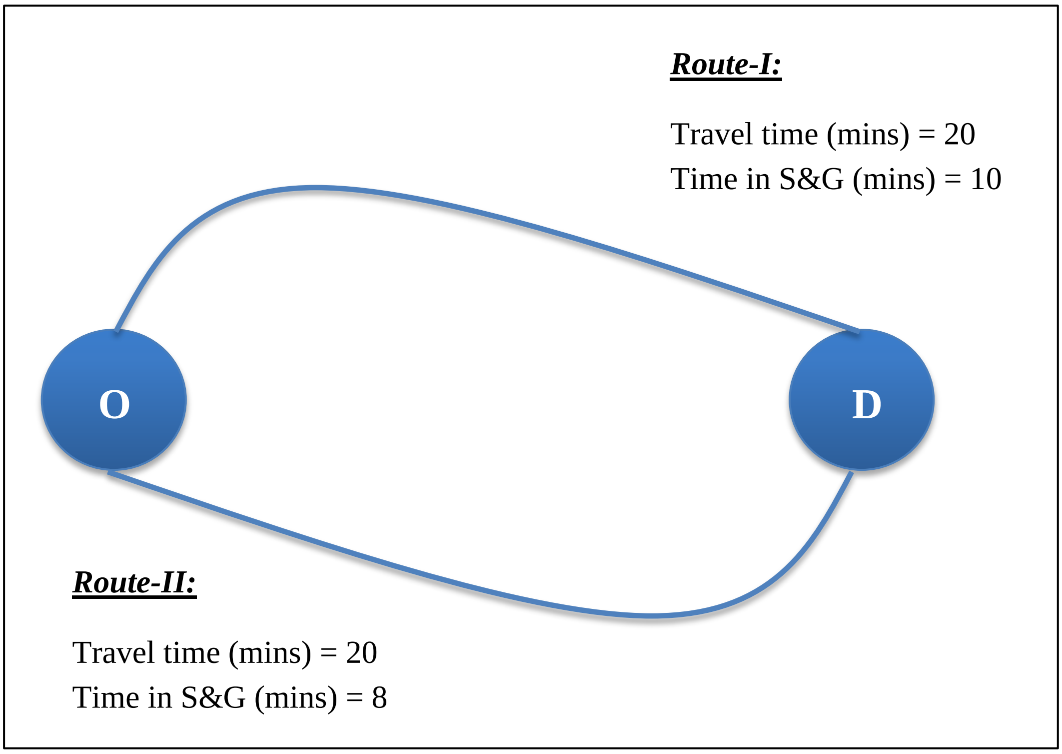

Figure 1.1: Hypothetical toy network with two route specific attributes

Figure 1.2 expands the previous network representation to include the additional attribute which is the number of S&Gs experienced. As shown in the figure, consider that routes I and II have 10 and 16 cycles of S&Gs prevailing on them. Using this extra piece of information, it can be calculated that the average time spent in each occurrence of S&G is 60 seconds and 30 seconds on routes I and II respectively.

Thus, it can be observed that route-II is more onerous for travel due to a frequent occurrence of S&G waves. Assuming a weighting parameter of 1 minute per S&G cycle, the new route disutility value can be evaluated using equations 1.3 and 1.4

$$Route - I: (20 - 10) + 10 \times 2.5 + 10 \times 1 = 45 \text{ minutes} \tag{1.3}$$

$$Route - II: (20 - 8) + 8 \times 2.5 + 16 \times 1 = 48 \text{ minutes} \tag{1.4}$$

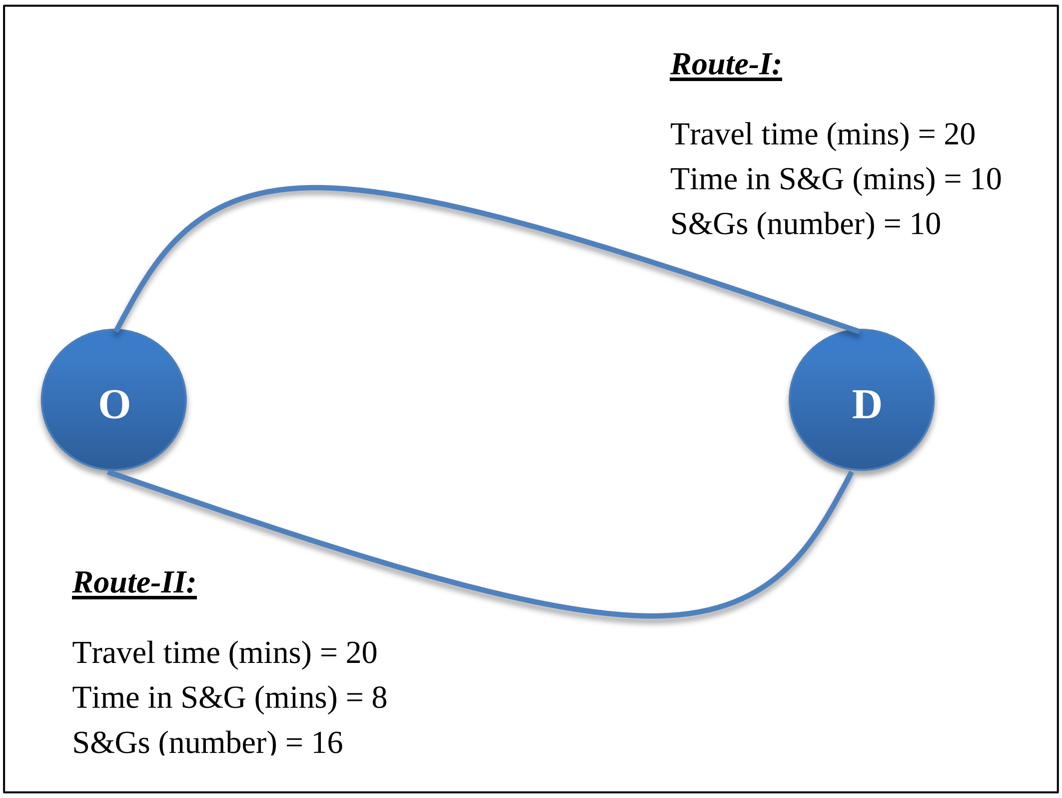

Figure 1.2: Modified toy network with the new attribute, number of S&Gs

Through equations 1.1 to 1.4, it can be seen that the route with the lowest disutility changed upon adding the new attribute. While route-II was the better option in figure 1.1, it got changed to route-II upon adding the number of S&Gs. Thus, figure 1.2 substantiates our initial intuition that drivers tend to have a disutility towards a route with more S&Gs, not accounting for which can lead to an inaccurate representation of vehicle assignment and the resulting traffic congestion. In reality, drivers generally try to jointly minimise both travel time and discomfort (expressed in terms of the number of S&Gs) instead of just minimising the former.

1.4 Aims and Objectives

The primary aim of this thesis is to test the research hypothesis: *an increase in the number of S&Gs on a route increases its disutility for a driver*. In other words, drivers are less likely to travel on a route which has a higher number of S&Gs *ceteris paribus*. Additionally, this thesis evaluates the willingness to pay (WTP) measures between the number of S&Gs with respect to travel time and cost. A series of experiments have been conducted on different samples from the car driver population in Sydney. The entire research project can be broken down into 2 main tasks which aim to test the validity of the proposed hypothesis. Figure 1.3 shows the funnel chart which illustrates the tasks that have been undertaken in this thesis. We now briefly discuss each of these objectives:

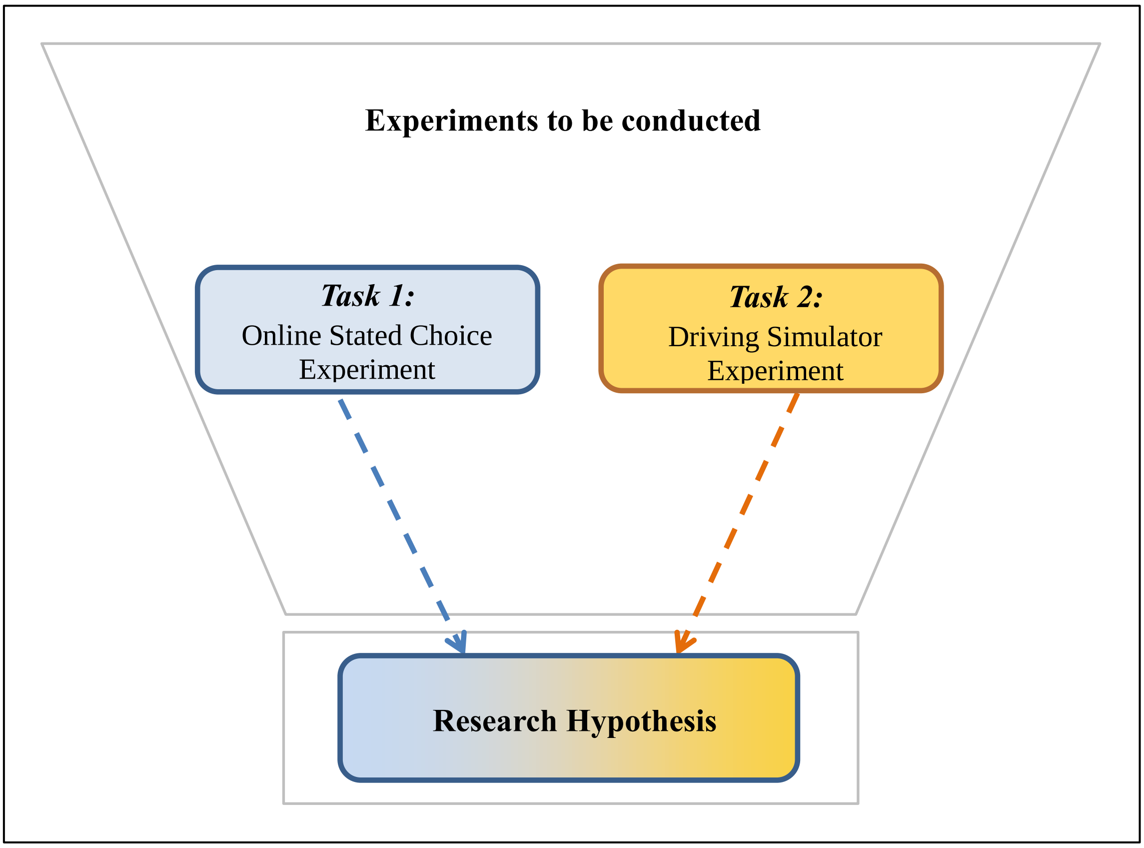

Figure 1.3: Funnel chart presenting the objectives defined for this thesis

<u>Task 1: Online Stated Choice Experiment</u>

Stated Choice (SC) experiments are a popular method of data collection in situations such as the addition of a non-existent alternative or a hypothetical policy decision (refer to Bliemer & Rose (2011) for an extensive review of the previous SC studies). SC experiments are not only quite effective, but are also quick, cheap and easy to conduct as they usually do not require any sophisticated setup for data collection. This task mainly involved the design of SC experiment which included the selection of appropriate:

- attributes and their corresponding levels,
- design strategy

The mode of data collection for the SC experiment was through an online web interface which further demanded

- preparation of the survey webpage
- systematically storing the survey responses in a database

The data collected from the online SC experiment was then analysed using discrete choice models which help in understanding the choice behaviour of individuals, the relative importance of attributes under consideration and the WTP measures to continue with the current choice. The candidate models used for analysis in this task were:

- *<u>Random Parameter Logit:</u>* It explains taste heterogeneity among individuals for a given attribute
- *<u>Latent Class Choice Model:</u>* It explains the choice preference of an individual and his/her association to a cluster or a population segment. All individuals belonging to a given segment depict similar choice preferences.

A detailed discussion of the specifications of experimental design, data collection, candidate model formulations and analysis of results is presented in chapter 3 and chapter 4 of this thesis.

Task 2: Driving Simulator Experiment

Driving simulator experiments present real-world scenarios to participants in a virtually controlled environment to closely study their driving behaviour. Additionally, these experiments help in minimising the measurement bias which could possibly arise during the online SC experiment. This task mainly involved:

- designing virtual driving scenarios that participants experienced
- preparing a survey questionnaire which asked the participants about their levels of frustration and route preferences in every driving scenario, socio-demographics and perception towards S&G traffic

The collected driving simulator data was then analysed using a Structural Equation Model (SEM). The SEM takes into consideration the effect of the number of S&Gs on the level of frustration of the participants. Chapter 5 discusses the design specifications of the driving simulator experiment along with the model formulation, quantitative data analysis and interpretation of results.

1.5 Contributions

The core contribution of this thesis is to closely understand the intricate route selection process of car drivers when subjected to S&G waves on a day-to-day basis. This research project is the first attempt towards including the number of S&Gs experienced as an additional attribute to explain a route’s disutility to a driver. The series of experiments conducted in this study not only test the proposed research hypothesis, but also explain the contribution of underlying socio-demographic and psychological factors that instigate a particular route choice behaviour in drivers. The findings from this study would potentially facilitate the development of transportation models that better represent the impacts of traffic congestion on network performance, fuel emissions, etc., by accounting for the ill-effects of S&G traffic. Additionally, the outcomes from this study will also aid transportation planners and decision-makers in framing policies, for example, toll pricing, aimed at mitigating both higher travel times and discomfort due to the increased number of S&Gs while evaluating the toll price. The main contributions of this thesis are summarised below along with the contributing chapter number in parenthesis:

- Understanding the impact of the number of S&Gs on the route choice behaviour of car drivers during work trips (Chapter 3)
- Determining the proportion of individuals that exhibit a specific taste towards the number of S&Gs attribute while making route choice (Chapter 4)
- Inferring the role of S&G traffic characteristics (including the number of S&Gs) along with socio-demographic characteristics on psychological factors, like the level of frustration of individuals. Also to check for the association between the inbuilt frustration and the resulting route choice (Chapter 5)

1.6 Organisation of Thesis

This thesis is divided into six chapters where each research contribution stated above is covered exclusively in a chapter. Figure 1.4 shows the schematic workflow diagram of this thesis. Chapters 3 and 4 have been put together under the same stream, *Online Stated Choice Experiments*, as they follow a similar experimental design methodology. A detailed discussion of each constituent chapter of this thesis is presented below:

Chapter 2: Background

This chapter conducts an extensive review of the literature to showcase the current state-of-the-art across three separate fields of knowledge where the phenomenon of S&G traffic has been studied. This review provides sufficient evidence for the motivation behind conducting this study, which has been discussed earlier in sections 1.2 and 1.3. This chapter first reviews the works which focussed on modelling the occurrence of S&G waves from the observed traffic data. It then discusses the studies which explained driver route choice in terms of the time spent in S&G traffic along with other travel related attributes. Lastly, this chapter reviews the works in the field of control theory relating to the design of S&G Adaptive Cruise Control (ACC) system in modern luxury cars using vehicle kinematic measures. This chapter identifies the research gaps associated with each of the three fields and highlights the contribution of the thesis with regard to addressing the identified gaps.

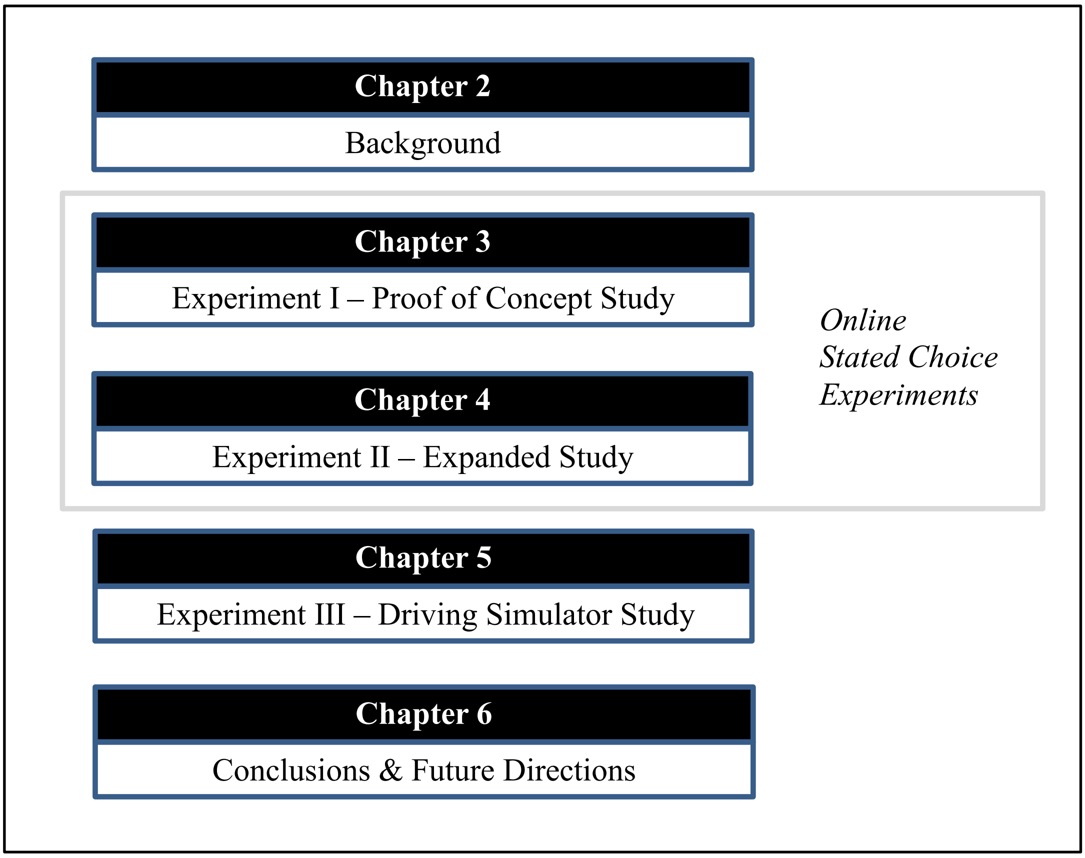

Figure 1.4: Workflow diagram illustrating the chapters of this thesis

Chapter 3: Experiment I – Proof of Concept Study

This chapter forms the first half of the stream *Online Stated Choice Experiments* and presents a route choice study that serves as a proof of concept. The study tested the research hypothesis on a sample of university staff and students who regularly drove to work by car during weekday mornings. A stated choice (SC) experiment design technique, which is a widely used method of experiment design, was selected in this study due to its advantages such as cost and time effectiveness and a simple design framework. This study utilised the SC design technique to come up with multiple choice tasks where each task comprised the current route and two other hypothetical routes that were pivoted around the first. This chapter first describes the design specifications like the selection of the number of alternatives, attributes and levels. It then discusses the layout of the online SC experiment that was presented to the participants. An empirical analysis of the collected data is presented next. Next, this chapter discusses the discrete choice model, the Random Parameter Error Component Logit (RPECL) model, and its formulation that was used for analysing the collected data. It is then followed by a discussion of the results from data analysis which shows that the participants do associate a non-zero disutility towards a route as the number of S&Gs increases, thus indicating the validity of the research hypothesis. Lastly, this chapter summarises the main findings, highlights the policy implications of the results and the limitations of the study.

Chapter 4: Experiment II – Expanded Study

This chapter constitutes the remaining half of the *Online Stated Choice Experiments* stream. It extends the scope of the research study by overcoming a few limitations associated with the proof of concept study discussed in chapter 3 of this thesis. This chapter first discusses the design procedure which was used to account for a different set of respondents. The data collection method is discussed next which is quite similar to the one discussed in the previous chapter. It is followed by an empirical analysis of the collected data. This chapter then presents a hybrid choice model, the Latent Class Choice Model (LCCM), and its formulation which was used for data analysis. Results show that drivers generally have a negative effect towards the number of S&Gs attribute; barring one-quarter of the sample which is indifferent towards this attribute. Finally, this chapter discusses the key findings and limitations of this study along with exploring few real world applications of the obtained results to better represent and counter the ill-effects of S&G traffic on the overall network performance.

Chapter 5: Experiment III – Driving Simulator Study

This chapter presents the driving simulator experiment that was conducted to further understand the role of S&G traffic characteristics on psychological factors like the level of frustration in drivers. It also investigates the dependency between the inbuilt frustration and the route choice of drivers. This chapter first presents the design specifications of the experiment which include developing the virtual scenarios, the parameters governing the formation of S&G waves and the layout of the experiment. An empirical analysis of the collected dataset is presented next. It then discusses the

Structural Equation Model (SEM) and its mathematical formulation that finds the relationship between the explanatory variables (socio demographic and route specific information) and the levels of frustration reported by the participants. Results show that attributes such as the time spent in S&G and the number of S&Gs have a positive effect on the built up frustration propensity (a latent variable). It also shows an association between frustration propensity and the observed route choice. This chapter highlights the potential applications of the results and limitations of the study at the last.

Chapter 6: Conclusions & Future Directions

This chapter summarises the thesis and discusses the main findings, policy implications and limitations of the research works that were conducted in this thesis. A few possible extensions of the research work are also proposed in this chapter.

CHAPTER 2

BACKGROUND

The occurrence of stop-&-go (S&G) waves, a phenomenon that usually exists in congested traffic, has motivated researchers to unravel its intricacies for more than five decades. Ever since their first detection inside the Lincoln tunnel in the US in the year 1958, there have been numerous works to learn more about the dynamics of S&G waves and how it affects driving behaviour. A wide spectrum of studies have been conducted till date which span across multiple domains such as driver psychology and behaviour, and different streams of engineering like transportation, electrical and automobile. These works can be broadly classified into three segments: 1) traffic flow studies, 2) route choice behaviour, and 3) adaptive cruise control theory. The works related to traffic flow study the life cycle of S&G waves such as their formation, propagation and dissipation. Studies associated with route choice modelling of drivers analyse the impacts of S&G conditions on the resulting route choice preferences of drivers. The adaptive cruise control theory, which is a relatively new field when compared to the other two, comprises works that design algorithms to enable modern luxury cars to traverse smoothly through the alternating cycles of S&G. These segments of research, even though quite extensive, looked at the S&G phenomenon from a unique and definite perspective. As a result, there exist few research questions which still need to be explored across these segments. Thus, the aim of this chapter is to discuss the state-of-the-art within each of the three segments of research and identify the potential research gaps which this thesis tries to address.

The organisation of this chapter is as follows: Section 2.1 provides an extensive literature review of the different traffic flow studies which looked at explaining the dynamics of S&G waves (also referred to as traffic oscillations (Li et al., 2010)) and quantifying their occurrence. Section 2.2 reviews the studies on understanding the route choice behaviour of drivers under S&G traffic. Section 2.3 delineates the literature in adaptive cruise control theory discussing the techniques used to quantify an occurrence of S&G. Finally, section 2.4 concludes the chapter with a discussion on the research gaps in the reviewed literature, and how this thesis aims to make a novel contribution to the existing knowledge base by addressing these identified gaps.

2.1 Traffic Flow Studies on S&G Waves

Stop-&-go (S&G) waves were first observed inside a tunnel, where lane change manoeuvres were prohibited, by Edie (1961) and Edie & Baverez (1967). The studies observed significant speed fluctuations in a one lane traffic stream caused due to small perturbations. Numerous works have been conducted since then to better understand the evolution and dissipation of S&G waves. These studies can be segregated into two categories: theoretical and empirical. The theoretical works modelled the dynamics of S&G waves by encompassing asymmetric driving behaviour (Yeo & Skabardonis, 2009) and driver heterogeneity (Laval et al., 2009; Laval, 2011) into the simplified car-following model proposed by Newell (2002). On the other hand, the empirical studies mainly analysed the observed traffic data using different techniques to uncover the formation, duration, dissipation and periodicity of S&G waves. However, despite the rich literature, a few complexities associated with the dynamics of S&G waves still remain unclear to researchers (Laval & Leclercq, 2010; Suh et al., 2012). We discuss the state-of-the-art for each of the two categories in the following subsections, along with highlighting the research gaps which this thesis aims to address.

2.1.1 Theoretical models to understand S&G waves

Earlier models which were used to represent traffic flow were based on the concepts of fluid mechanics (Pipes, 1950; Lighthill & Whitham, 1955; Richards, 1956). These were the first order macroscopic models that well represented the movement of vehicles in congested traffic. However, these models did not realistically represent the S&G phenomenon due to their inability to account for: 1) unstable flows, 2) spontaneous breakdowns and 3) capacity drops (Nagel & Nelson, 2005). One of the earliest work in the direction of modelling the S&G phenomenon can be traced back to Chandler et al. (1958) and Herman et al. (1959). For example, Chandler et al. (1958) proposed a linear car-following model which is illustrated in equation 2.1. The equation states that the acceleration of the follower vehicle at time t is directly proportional to the relative velocity difference between the leader and follower vehicles at time $t - \Delta$.

$$\ddot{x_f}\,(t) = \lambda\left(\dot{x_l}\,(t - \Delta) - \dot{x_f}\,(t - \Delta)\right) \qquad (2.1)$$

Where,

Δ represents the reaction time (time lag) which is about 1.5 seconds

λ is the proportionality constant which around 0.37 sec^{-1}

$\ddot{x_f}\,(t)$ is the acceleration of the follower vehicle at time t

$\dot{x_l}\,(t - \Delta)$ is the velocity of the leader vehicle at time $t - \Delta$

$\dot{x_f}\,(t - \Delta)$ is the velocity of the follower vehicle at time $t - \Delta$

The model explained the concepts such as the local and asymptotic stabilities which represent the response of the follower vehicle (in the case of local stability) and downstream vehicles (in the case of asymptotic stability) due to a fluctuation triggered by the leader vehicle. These local and asymptotic instabilities characterise the formation of S&G waves. Since then, a few research works have been carried out modelling the S&G phenomenon using the first order traffic flow models. For example, recently Shott (2011) applied a mesoscopic LWR model to observe the creation and propagation of S&G waves on a freeway with an on-ramp. The effect of oscillations on en-routing decisions was also studied, making available the travel time during oscillations on each of the two routes to the drivers. These linear models were easy to compute due to its simple formulation. However, these models did not represent the S&G evolution realistically, mainly due to the exclusion of physical constraints (for eg. speed range) and non-linear driving behaviour (Li & Ouyang, 2011).

Non-linear car-following models, unlike their linear counterparts, impose a speed bound which restricts the magnitude of S&G waves to approach infinity. For example, Gazis et al. (1959) highlighted the inability of the linear car-following models to represent the flow density relationship derived from the observed traffic flow data inside the Lincoln tunnel. Thus, the authors proposed a non-linear car-following model which states that the response of the follower vehicle is directly proportional to the velocity difference and inversely proportional to the relative spacing between itself and the leading vehicle. Equation 2.2 shows the proposed non-linear car-following model. Subsequent works by Edie (1961) and Gazis et al. (1961) proposed a general class of non-linear car-following models which is shown in equation 2.3.

$$\ddot{x_f}\,(t) = \frac{\lambda}{\left(x_l\,(t-\Delta) - x_f\,(t-\Delta)\right)}\left(\dot{x_l}\,(t-\Delta) - \dot{x_f}\,(t-\Delta)\right) \tag{2.2}$$

$$\ddot{x_f}\,(t) = \frac{\lambda\,.\,\dot{x_l}^{m}\,(t)}{\left(x_l\,(t-\Delta) - x_f\,(t-\Delta)\right)^{n}}\left(\dot{x_l}\,(t-\Delta) - \dot{x_f}\,(t-\Delta)\right) \tag{2.3}$$

In the two equations, $(x_l\,(t-\Delta) - x_f\,(t-\Delta))$ is the relative spacing between the follower and leader vehicles. m and n are the calibration parameters. The remaining terms have their usual meaning.

A wide range of non-linear car-following models have been proposed since then to realistically model the dynamics of S&G waves. Gipps (1981) proposed a safe distance car-following model where the follower vehicle adjusts its speed such that it can avert a collision in case of sudden braking by the leader vehicle. Bando et al. (1995; 1998) put forward a non-linear Optimal Velocity Model (OVM) to study S&G waves, which was later used and extended by several research works (Davis, 2003; Helbing, 1997; Jiang et al., 2001; Sawada, 2002; Zhao & Gao, 2005). However, the additional level of realism offered by these models was undermined by a few challenges. Some of the limitations of these models were: 1) these could not be solved using traditional analytical methods, due to the curse of non-linearity, but required simulation techniques instead, 2) the models required approximations to be made while representing local stability properties (Li et al., 2010), and 3) calibration of such models often became a daunting task. A recent work by Li & Ouyang (2011) proposed a solution technique to quantify the formation of S&G waves when using non-linear car-following models.

A wealth of works also exist in the literature which have associated the occurrence of S&G waves with heterogeneity in driver responses. Del Castillo (2001) proposed a car-following model that assumed a probabilistic distribution of headways during the deceleration state. Similarly, Kim & Zhang (2008) proposed a wave propagation model which considered stochastic driver reaction time. Kerner (2004) and Orosz et al. (2009) proposed another traffic state, in addition to the free flow and congested regimes. Traffic flow in this regime was considered to be stable, but sensitive to external perturbations, thus resulting in the formation of S&G waves. The studies discussed above followed the simplified car-following model proposed by Newell (2002) to represent vehicle dynamics. However, the car-following model assumed symmetric driving behaviour resulting in parallel vehicle trajectories. This behaviour was contradicted in a study by Laval & Leclercq (2010) who found the observed vehicle trajectories (on a freeway) around S&G locations to be deviating from the Newell trajectories. This brought out the limitation of such studies as they could not accurately model the dynamics of S&G waves, even though they assumed heterogeneity in driving response. Furthermore, these studies did not clearly identify what causes the difference in driver behaviour across traffic states.

Asymmetric driving behaviour was first observed by Edie (1965) and Foote (1965) who found a variation in driver’s reaction time and space headway during acceleration and deceleration states. Newell (1965) put forward two separate curves (in the speed-spacing diagram) for the acceleration and deceleration states. However, the actual shapes (functions) of such curves were not discussed. Yang & Koutsopoulos (1996) used a split fundamental diagram with a separate set of parameters for the acceleration and deceleration curves. Yeo & Skabardonis (2009) recently proposed the asymmetric traffic theory which states that drivers have different sensitivities towards gap acceptance and speed adjustments when exposed to acceleration and deceleration situations. The work analysed freeway vehicle trajectories and was able to identify the A-curve and the D-curve corresponding to the acceleration and deceleration states respectively. The study also found that the measurement and anticipation errors in human driving have an association with the growth and dissipation of S&G waves. For example, an S&G wave, triggered by a lane change, can grow in magnitude due to reasons such as: 1) proximity of the follower vehicle to the D-curve (reduced spacing), and 2) drivers overreacting to the anticipation error. Laval & Leclercq (2010) picked heterogeneity in driver aggressiveness to model the formation of S&G

waves. The study found the presence of both timid and aggressive drivers around the locations of S&G in the empirical vehicle trajectory dataset. Figure 2.1 shows the space-time trajectory of a few observed vehicles undergoing an S&G wave. The figure shows that while the aggressive drivers (shown as *) maintain a closer distance than the Newell trajectory, the timid drivers (shown as +) leave a wider spacing. It was also observed that the timid drivers caused a relatively larger impact on the evolution of S&G waves than the aggressive ones. The study proposed modifications to the car-following model proposed by Newell (2002) by taking into account driver heterogeneity in the modelling framework. Later, Chen et al. (2012) also proposed a car-following model that accounted for behavioural heterogeneity among drivers before and after undergoing S&G waves. The authors also found rubbernecking as a potential trigger behind the formation of S&G waves.

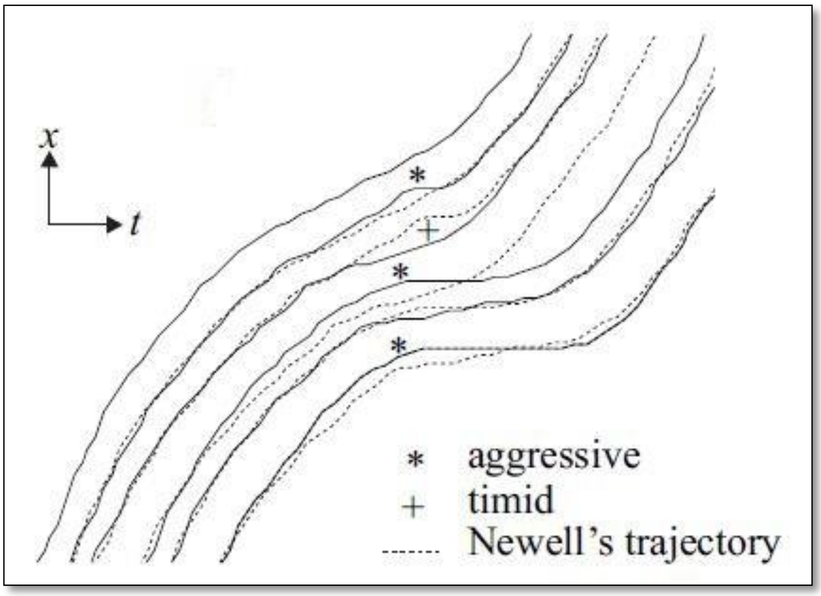

Figure 2.1: Contrasting driver aggressiveness against Newell's trajectory

(Source: Laval & Leclercq (2010))

The discussion above presented a snapshot of the theoretical works looking into the S&G phenomenon. Readers can find an extensive review of the state-of-the-art of such works in the articles by Laval & Leclercq (2010), Li & Ouyang (2011) and Suh et al. (2012).

2.1.2 Empirical studies to quantify S&G waves

Analogous to the theoretical studies, several research works have analysed real-world traffic data, like loop detector counts and vehicle trajectories, to learn more about the triggers, dynamics and safety hazards (eg. crash occurrence (Zheng et al., 2010)) of S&G waves. Table 2.1 lists out a few other triggers for S&G waves, apart from driver asymmetry and car-following behaviour that have been identified to date. Of all the triggers, lane changing manoeuvres have been identified as the primary reason behind the initiation of S&G waves. It leads to a sudden reduction in the available gap for the following vehicle, once the vehicle in front cuts in, thus triggering the formation of S&G waves (Yeo & Skabardonis 2009). Few studies also observed the cyclic nature of S&G waves which repeat in the interval of 2-15 minutes, last for up to 30 seconds and propagate backwards at a wave speed between 10 to 20 km/h (Ahn et al., 2004; Laval et al., 2009; Laval & Leclercq, 2010; Li et al., 2010; Mauch & Cassidy, 2004).

Table 2.1: Factors initiating S&G waves

Trigger	Identified by
Lane changing manoeuvres	Ahn & Cassidy (2007); Laval & Daganzo (2006); Laval (2007); Suh et al. (2012); Yeo (2008); Yeo & Skabardonis (2009); Zielke et al. (2008)
Moving bottlenecks	Koshi et al. (1992); Laval (2007)
Static bottlenecks (eg. merges and diverges)	Ahn & Cassidy (2007); Bertini & Leal (2005); Cassidy & Rudjanakanoknad (2005); Laval & Daganzo (2006); Laval et al. (2007); Mauch & Cassidy (2004); Menendez (2006)
Queue discharging side of capacity drops	Bertini & Leal (2005); Cassidy & Bertini (1999a,b)
Roadway geometric features (eg. curves and uphill sections)	Jin & Zhang (2005); Laval & Leclercq (2010)

A variety of empirical data processing methods have been used to extract S&G information from the observed traffic data. These can be classified into stationary and non-stationary signal processing techniques. An input (data) signal is considered as stationary if its frequency or periodicity remains constant with time. In other words, the analyst is only interested in knowing the underlying frequency component while working on a stationary input signal. A pure sinusoidal function is a perfect example of a stationary wave, which has got a uniform frequency of $1/2\pi$ cycles/radian. Non-stationary wave analysis techniques, on the other hand, are capable of reporting both frequency and time information of any fluctuation embedded in the signal. We review the two techniques in the following subsections highlighting their merits and limitations.

2.1.2.1 Stationary wave processing techniques

The earlier empirical studies analysed traffic oscillation properties using the time series of raw traffic data. For example, Kuhne (1987) fitted sinusoidal waves on the speed profile from a loop detector to determine the characteristics (amplitude and frequency) of S&G waves. Paolo (1988) also conducted a similar study using the traffic count information from different loop detectors. The two studies aggregated information over a given time period to smoothen the raw data. However, aggregation dampens the effect of traffic oscillations by smoothing it out along with other unwanted components (Zheng & Washington, 2012). Thus, these techniques were not reliable in estimating traffic oscillations, as they might have provided contradictory results from what was actually present in the original data. Neubert et al. (1999) conducted a cross-correlation analysis on traffic flow parameters (average speed, flow and density) which revealed that S&G waves were characterised by a strong correlation between flow and density ($\rho \sim 1$). The time period of oscillations was determined by measuring the separation between neighbouring peaks on a correlogram, which was found to be around 10 minutes. However, few limitations of this method, as highlighted by Li et al. (2010) were: 1) identification of distinct peaks becomes challenging in case of multiple comparable frequency components, and 2) amplitude of traffic oscillations cannot be determined from the correlogram plot, which is standardised between [-1,1]. Muñoz & Daganzo (2003) used another signal processing technique, called the oblique coordinate system, to plot cumulative traffic counts against time to reveal the underlying traffic oscillations. The analysed data comprised aggregated loop detector counts, occupancy and average speeds over a 20 second interval. The oblique coordinate system amplifies the signal pattern using a technique that is similar to the second order difference of cumulative vehicle counts with a moving time window (Mauch & Cassidy, 2004). The advantages of considering a moving time window were: 1) it helped reduce the local noise from traffic data, and 2) it provided frequency along with an approximate location of the signal fluctuation in time. The smoothed data signal is given by equation 2.4.

$$\widehat{x_m}\,(m_0) = f_m - \frac{1}{2}\left(f_{m+m_0} + f_{m-m_0}\right) = \frac{1}{2}\left[\sum_{i=0}^{m_0-1} x_{m-i} - \sum_{i=1}^{m_0} x_{m+i}\right] \qquad (2.4)$$

Where $\widehat{x_m}$ represents the cumulative traffic count at an instant m, and m_0 is the half window length on either side of m, which was taken as 7.5 minutes in the study. Wiggles in the oblique plot represented the oscillation pattern in the data. The technique became popular and was picked by other researchers due to its simple framework. However, later research works identified a few limitations inherent in the model formulation. The oblique coordinate system method requires a careful selection of the time window, a failure to do so might lead to biased traffic oscillation information from the smoothed data. Figure 2.2 has been taken from Li et al. (2010), which gives a good illustration of the impact of time window length on the resulting oscillation patterns. Figure 2.2(a) shows a pure sinusoidal signal used as an input and figures 2.2(b) and (c) show the resulting patterns using the window lengths as 10 and 30 units respectively. While figure 2.2(b) shows an amplification of the signal (meaning frequent traffic oscillation), figure 2.2(c) shows a considerable dampening of the same input signal (inferring negligible traffic oscillation). Another limitation of the method can be seen through figures 2.2(d-f) where different lengths of the time window might cause a noisy signal (figure 2.2(d)) to depict periodic oscillations of different magnitudes (figures 2.2(e and f)). Thus, an inappropriate window length might lead to an under or over representation of the underlying traffic oscillation patterns.

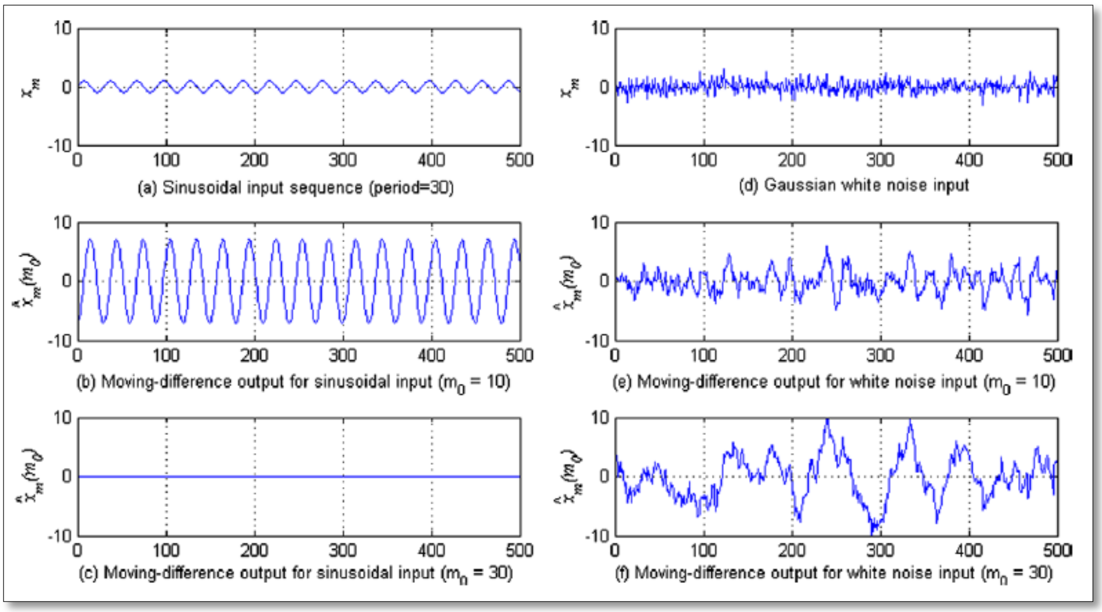

Figure 2.2: Effect of window length on signal resolution

(Source: Li et al. (2010))

Li et al. (2010) conducted a frequency spectrum analysis, a popular technique in signal processing, to reveal traffic oscillation patterns from the aggregated traffic data. The analysis comprises three steps: 1) de-trending the signal to remove traffic demand effects, 2) identifying stationary time intervals for analysis, and 3) detecting oscillations of interest in these time intervals. De-trending is generally carried out by fitting a lower order polynomial function. A Short Time Fourier Transformation (STFT) is then applied to identify the oscillation pattern within the de-trended, non-stationary time series data. STFT overcomes the limitations of the Fourier transform by using multiple smaller sized windows to capture irregularities in the non-stationary data. STFT plots are helpful in dividing the signal into smaller, same sized time intervals within which the oscillation pattern remains invariant. The oscillation pattern within this time interval is then analysed to determine its amplitude and frequency, which represents the magnitude and period of the oscillation. The authors also used a term called the cycle abundance index (CAI) to quantify the number of oscillations caused during a given time interval. The study found the average oscillation periodicity to be between 8 and 12 minutes with an average CAI around 6 across different study locations. However, this method also requires a subjective judgement while selecting the oscillation-invariant time period, making it difficult to reproduce the same result across analysts (Zheng & Washington, 2012). The study also proposed a model to relate oscillations observed at the loop detectors with the trajectory oscillations experienced by an individual driver. Equation 2.5 presents the proposed model which was developed using a simplified vehicle trajectory and principles of traffic theory

$$\frac{T_d}{T_t} = 1 + \frac{\bar{v}}{v_w} > 1 \tag{2.5}$$

In this equation, T_d is the oscillation period observed from the detector data (using the frequency spectrum analysis), T_t is the period of oscillation faced by a driver, $\bar{v}$ is the average speed of the vehicle and v_w is the traffic wave speed (~ 15 km/h). The model is a useful find as it provides a reasonable and cost-effective means of determining traffic oscillations at an individual vehicle resolution from the easily available loop detector data. However, the model made some simplifying assumptions, which limited the application of the model in the real world. Firstly, the model assumed a steady traffic state which does not hold true as the traffic flow rate approaches or exceeds the roadway capacity (Rouphail et al., 2005; Tanaka & Nakatsuji, 2011). Thus, the steady state assumption does not hold true in the presence of traffic oscillations, which are prevalent in congested traffic. Secondly, the model considered zigzag vehicle trajectories which do not reflect driver asymmetries (Laval & Leclercq, 2010; Laval, 2011; Yeo & Skabardonis, 2009).

2.1.2.2 Non-stationary wave processing techniques

The above discussion indicates that the observed traffic data, in general, shows a temporal variation in frequency which cannot be accurately picked by the previous techniques. Wavelet Transformation (WT) has evolved as a widely used technique over time for identifying transient locations in a non-stationary signal. WT is useful in discerning the location and frequency components of a pulse in a signal, which is not visible to a naked eye, thus making it useful in analysing local events. The technique is quite popular in the field of image processing, geophysics, finance, engineering and medicine (Addison, 2002; Kumar & Foufoula-Georgiou, 1997). In the last one decade, WT has found numerous applications in traffic engineering relating to automatic detection of freeway incidents (Adeli & Samant, 2000; Ghosh-Dastidar & Adeli, 2003), traffic features around freeway work zones (Adeli & Ghosh-Dastidar, 2004; Ghosh-Dastidar & Adeli, 2006), traffic flow forecasting (Boto-Giralda et al., 2010; Jiang & Adeli, 2005), and traffic pattern recognition (Jiang & Adeli, 2004; Vlahogianni et al., 2008). Recent studies by Zheng et al. (2011a,b) and Zheng & Washington (2012) applied WT to distil origins of S&G waves from the transient non-stationary vehicle trajectory data. Figure 2.3 has been taken from Zheng et al. (2011a) and shows how the spatio-temporal location of the start and end of an S&G wave can be extracted from a vehicle trajectory (top figure). The authors defined an S&G wave as one complete cycle of deceleration followed by acceleration. The transient points are characterised by sharp spikes in the wavelet based energy plot (bottom figure).

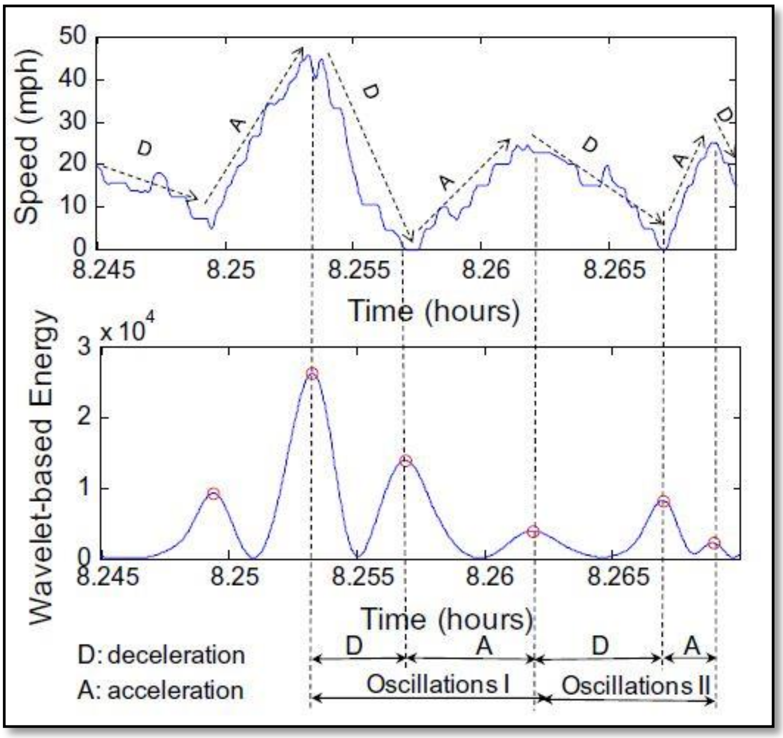

Figure 2.3: Detecting S&G formation using wavelet transformation

(Source: Zheng et al. (2011a))

Addison (2002) and Zheng & Washington (2012) conducted numerical simulation experiments to compare stationary wave processing techniques (in time and frequency domain) against WT. The studies suggested the superiority of WT over other popular techniques with regard to accuracy, robustness and consistency. Moreover, the WT technique requires no subjective judgement in selecting the size and shape of the time window Thus, the results can easily be replicated across analysts. A detailed discussion on the WT technique and its unique properties has been provided in Appendix A of this thesis.

In a quick recap to this section, research works in modelling the stop-&-go (S&G) phenomenon have evolved over decades, ever since their first detection in the early 1960s. The current state-of-the-art includes both theoretical models (such as the behavioural car-following model by Laval & Leclercq (2010)) and empirical techniques (such as wavelet transformation) that are capable of explaining and quantifying the dynamics of S&G waves. However, the field, even though deep, does not throw sufficient light on some research questions once the horizon is broadened. Firstly, the studies discussed above did not ascertain the effect of S&G traffic on the way people make their routing decisions. Secondly, a majority of the recent works analysed freeway vehicle trajectory data (NGSIM, 2010) to study S&G waves. Thus, their application gets constrained when trying to understand S&G traffic on urban roads, which exhibit different dynamics when compared to freeways. We discuss these limitations in detail in the last section of this chapter.

2.2 Studies on Route Choice in S&G Traffic

As discussed earlier in this chapter, stop-&-go (S&G) waves occur in cycles of deceleration followed by acceleration. It becomes onerous to traverse through such traffic conditions as drivers need to be more focussed while driving, resulting in elevated levels of discomfort and frustration levels (Hennessy & Wiesenthal, 1999; Levinson et al., 2004). Thus, a few studies have explored the impact of S&G traffic on the route choice behaviour of drivers. The first study in this direction was conducted by Small (1999). A stated choice (SC) experiment was conducted to understand the route preferences of individuals. Figure 2.4 shows one of the questions (referred to as choice task henceforth) that was presented to a participant. Each choice task comprised two hypothetical routes each of which was defined in terms of travel time, travel cost and the percentage of travel time spent in S&G traffic. Each of the three attributes was allowed to take 3 levels. S&G traffic was defined as the congested traffic state corresponding to the level of service (LoS) of E and F as per the highway capacity manual (TRB, 2000).

SC Experiment: Travel Time and Congestion Please circle either Choice A or Choice B	
Average total travel time: **11 minutes** Percentage of total time spent in S&G traffic: **36%** Your cost: **US $0.25** **Choice A**	Average total travel time: **8 minutes** Percentage of total time spent in S&G traffic: **38%** Your cost: **US $1.50** **Choice B**

Figure 2.4: Sample stated choice task used in Small (1999)

Further extensions of this study were conducted by Hensher (2001a) and Rose et al. (2009), who evaluated the impact of driving under different traffic conditions on a route’s disutility. For example, Hensher (2001a) also conducted an SC study where the participants were given three alternatives, namely, the status-quo (current) alternative and other two hypothetical alternatives that were pivoted around the status-quo alternative. Each alternative was defined in terms of six attributes: free flow travel time, slowed down travel time, S&G travel time, travel time uncertainty, car running cost and toll charge. Each attribute was set at 4 levels. The studies showed that drivers find it more onerous driving in S&G traffic than other traffic conditions. For instance, car commuters, on average, perceived each minute spent under S&G traffic equivalent to 2.5 minutes of free flow time (Hensher, 2001a). However, the attribute duration spent in S&G traffic is not the only indicator of discomfort experienced by a driver. Consider the case where the time spent in S&G on a route R is 4 minutes. This duration can either be due to the occurrence of 4 stops each of 1 minute duration, or by 8 stops of 30 seconds duration each. In reality, the latter case is expected to be more cumbersome causing a significant increase in the level of frustration and discomfort among drivers (Levinson et al., 2004; Malta et al., 2011). Thus, it is more intuitive to also study the effect of the number of S&Gs on a route’s disutility.

Reviewing the SC experiment design, multiple choice tasks were presented to the participants during the survey. The total number of choice tasks generated ranged between 11 (Small, 1999) and 64 (Hensher, 2001a). The selection of choice tasks was made using the fractional factorial method which picks the desirable set from the full factorial (3^3 and 4^{12} respectively for the above two studies). Further, a blocked design style was adopted where a subset of the selected choice tasks was presented to a participant (2 blocks of 5 and 6 choice tasks each in Small (1999) and 4 blocks of 16 each in Hensher (2001a)). Although the fractional factorial technique is computationally easy, it has been found to suffer from a few limitations like the dominant alternative case, non-orthogonality of alternatives and attribute level imbalance (Hensher et al., 2009). An inappropriate SC design could lead to inaccurate estimation of the model parameters and WTP measures. Thus, a more principled SC design strategy would be more accurate in unravelling the route choice behaviour of drivers.

We now review the data analysis techniques that were used in the aforementioned studies. Small (1999) used a logistic regression on the collected SC experiment data to evaluate the trade-off among the different attributes. Although the logit model has a simple closed form framework, it makes rather restrictive assumptions like the IIA property and a uniform or systematic taste preference across individuals (Train, 2009). This inadequacy led to a widespread application of the mixed logit model in this context. The mixed logit model not only estimates the mean parameter effect, but also its standard deviation which explains the preference heterogeneity observed across individuals. Hensher (2001a) and Rose et al. (2009) used mixed logit models to quantify the impact of driving in S&G traffic on the route choice of car commuters. The model had better goodness-of-fit statistics than the conventional multinomial logit (MNL) model for the SC dataset under consideration. However, the mixed logit data analysis technique (particularly the random parameter logit model) was later challenged by researchers due to a few reasons such as: 1) presence of a status-quo alternative which remains time-invariant for an individual, 2) a priori assumptions regarding the mixing distribution of the random coefficients, and 3) tricky to include socio-demographic variables directly in the utility function. We elaborate on these shortcomings towards the end of this chapter.

2.3 Modelling S&G Conditions in Control Theory

A rich literature exists in the field of control theory on the Adaptive Cruise Control (ACC) (also known as Adaptive Driver Assistance Systems, ADAS) feature that is available in modern vehicles today. Starting in the early 1990s within the luxury car segment, ACC systems are now available in vehicles across different car and truck segments. While a Cruise Control (CC) feature controls just the throttle of a vehicle, ACC is a semi-autonomous system that operates both the throttle and brake systems of the car by maintaining a pre-set value of time headway from the target (leader) vehicle in a traffic stream (Naus et al., 2008). As a more recent development, these advanced systems now come with a stop-&-go functionality that enables low-speed adaptive cruise control in congested traffic. The feature is specifically designed for vehicles experiencing recurring cycles of stop-&-go (S&G) waves in urban road networks. The advantages of stop-&-go ACC (S&G-ACC) are as follows. Firstly, it relieves the driver from the additional stress caused due to frequent cycles of deceleration followed by acceleration (Marsden et al., 2001; Venhovens et al., 2000). Secondly, a simulation study by Benz et al. (2003) showed that the S&G-ACC feature enables the vehicle to accelerate and decelerate smoothly, which significantly helps in bringing down fuel emissions. Moreover, the study also observed improved traffic efficiency as vehicles cruised at a smaller and consistent time headway between 1 – 1.8 seconds, when compared to vehicles without this functionality, thus increasing traffic throughput. Figure 2.5 shows the working mechanism of the S&G-ACC equipped vehicle. A radar is mounted on the host vehicle (the white coloured vehicle in figure 2.5), which tracks the relative speed and distance between itself and the target vehicle (the immediate leader in the same lane). A suitable time headway value is set upfront by the user (driver of the host vehicle), and the desirable relative distance is calculated using the selected time headway, the host vehicle speed and the spatial separation between the two vehicles at standstill. The S&G feature is activated once the actual distance goes below the desirable value, which then decelerates the host vehicle by applying brakes. Alternatively, Stanton et al. (2011) defined the activation of S&G-ACC once the target vehicle was detected and the speed of the host vehicle went below 26 km/h.

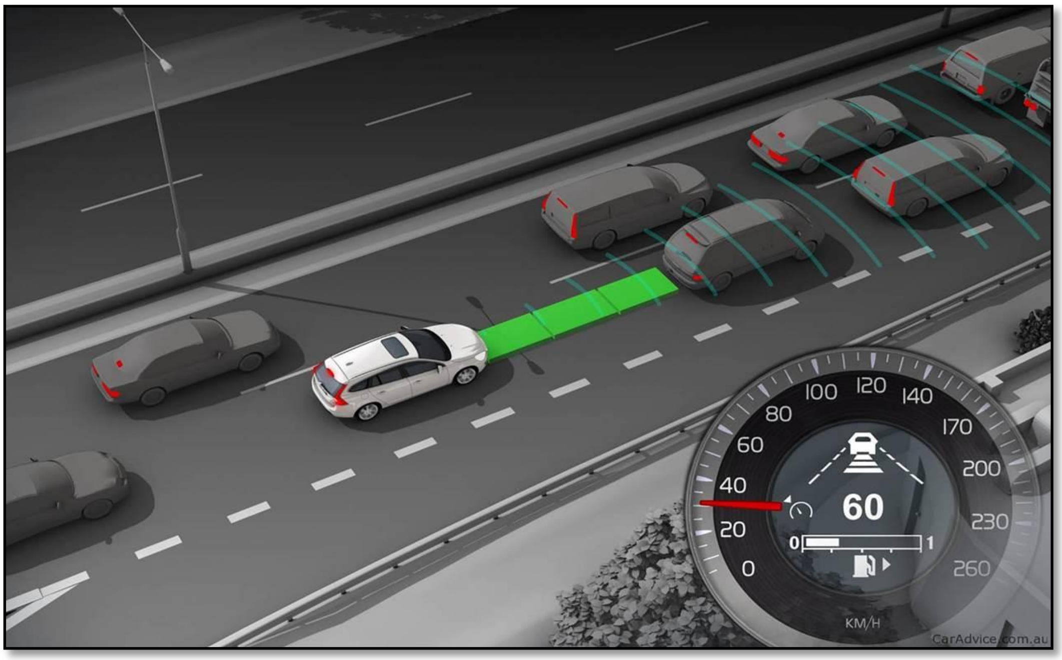

Figure 2.5: Working of the S&G-ACC function in modern cars

(inset: Odometer of the host vehicle) (Source: Gritzinger (2015))

Most ACC algorithms in practice are formulated as a linear (Naus et al., 2008; van Driel et al., 2007) or a non-linear (Martinez and Canudas-de-Wit, 2007) programming problem. The main control objective is to maintain a desirable relative distance between the host and leader vehicles. The objective function is bounded by constraints or criteria for a better performance evaluation of the ACC system. These criteria can be divided into two main categories: 1) comfort and 2) driving behaviour characteristics. As S&G-ACC is a semi-autonomous system, it should resemble the non-linearity in driving behaviour to enhance its user acceptability (Stanton et al., 2011). Time headway and time to collision (TTC) are generally used as proxies to analyse driving behaviour when following a preceding vehicle (Han & Yi, 2006; Yamamura et al., 2001). While time headway is defined as the time difference between the fronts of the leader and host vehicles, TTC denotes the time before the two vehicles collide, if neither of the vehicles take an evasive action. The comfort criterion, on the other hand, can be expressed in terms of the longitudinal motion or fluctuations a driver experiences while travelling in S&G traffic. Typically, peak acceleration (deceleration) and jerk (rate of change in acceleration or deceleration) values are taken as comfort metrics. Bounded values of the two metrics can ensure a certain degree of comfort in the longitudinal control of the vehicle (Martinez & Canudas-de-Wit, 2007; Naus et al., 2008; Yi & Moon, 2004).

The algorithms undergo a rigorous testing and calibration exercise which involves simulation experiments. The experiments use wider bounds for the comfort and driving behaviour constraints to accommodate heterogeneity in driving behaviour, which enhances the safety aspect of the ACC system. For example, one of the design constraints, the peak deceleration, was set a value as high as -7.8 m/s^2 (Sieler et al., 1998), -9 m/s^2 (van Driel et al., 2007) and -10 m/s^2 (Martinez & Canudas-de-Wit, 2007) to evaluate the performance of the ACC system being studied. However, the selected thresholds are quite high which is not generally witnessed in a real-life context. Moreover, likewise the traffic flow literature, these studies also did not explore the effect of S&G traffic on the routing behaviour of drivers.

2.4 Discussion

This chapter reviewed the three domains of knowledge where the characteristics of S&G waves have been extensively studied. First, section 2.1 presented the state-of-the-art in the field of traffic flow studies. The current models and techniques are capable of explaining the perplexities associated with the S&G phenomenon. However, these methods have a few shortcomings which were briefly discussed towards the end of this section. The studies provided an in-depth analysis of the life-cycle of S&G waves, but could not explain the consequences of this phenomenon on the route choice behaviour of drivers. It is equally important to know how drivers react to such conditions, as ignoring this can lead to inaccurate modelling of vehicle assignment and the resulting traffic congestion in a road network. Furthermore, the works were mainly built upon analysing freeway data (traffic counts and vehicle trajectory information). This restricts the transferability of the results to urban roads which exhibit contrasting traffic dynamics to freeways. In other words, there exist a few additional factors, specific to urban roads, such as parking manoeuvres and presence of roundabouts and signalised intersections that can also trigger an S&G wave (Fosgerau et al., 2013; Wijayaratna & Wijayaratna, 2016).

Section 2.2 presented the literature on the effect of S&G traffic on the route preferences of drivers. However, these studies took into account the time spent in S&G traffic and did not consider the number of S&Gs experienced which also represents an indicator to the discomfort caused. Thus, the paramount objective of this thesis is to determine the impact of the number of S&Gs experienced on a route towards its disutility. Secondly, the fractional factorial technique used to design the SC experiment has been widely criticised due to a random selection of choice tasks. A more recent design strategy, also called the D-efficient design, provides a more principled way of picking choice tasks from the full factorial. The strategy facilitates significant parameter estimation utilising a smaller sample size than other design techniques (Bliemer & Rose, 2011). Due to these advantages, we propose to use the D-efficient design technique to generate the SC experiments in this thesis. Lastly, section 2.2 also identified a few limitations of the previous studies discussed with regard to the quantitative data analysis techniques used. The conventional MNL specification is not a right tool to analyse SC datasets. The mixed logit model is able to capture the unobserved correlation structure in SC datasets, but is prone to some deficiencies. Firstly, the model is not well-suited for SC datasets which comprise the status-quo alternative. The status-quo alternative induces unobserved serial correlation among other alternatives and choice tasks due to its time-invariant nature (for an individual) (Hess & Rose, 2008; Train & Wilson, 2008). This correlation cannot be captured by the simple mixed logit (random parameter model to be precise) specification. Thus, we estimate a modified version of the mixed logit model, called the Random Parameter Error Component Logit (RPECL), proposed by Hess & Rose (2007) to analyse the SC dataset in this thesis (chapter 3). Secondly, the mixed logit model requires a priori specification of the mixing distribution for each random coefficient, where incorrect selection can have deleterious effects on the parameter estimates and model interpretation (Hensher & Greene, 2003; Hess et al., 2005). Furthermore, simply knowing that a parameter is randomly distributed across individuals is of lesser interest to policy-makers (Hess et al., 2009). Thirdly, the works discussed in this section did not take into consideration the role of demographics, socio-economic characteristics, and attitudes towards an individual’s route choice. There exists a vast literature which discusses the differences in attitudes and behaviours of drivers towards congested traffic (see Hennessy & Wiesenthal (1997) for example). These differences arise due to factors like age and gender (Blanchard et al., 2000; Wells-Parker et al., 2002; Wiesenthal et al., 2000). Thus, to overcome the last two shortcomings, we use hybrid choice models like the latent class choice model (chapter 4) and the structural equation model (chapter 5) in the later chapters of this thesis. These models are advanced econometric frameworks which have the following advantages: 1) they do not require any a priori assumption on the mixing distribution, 2) these can model latent constructs (quantities) which typically cannot be observed by the analyst, and 3) they furnish enriched predictive power of the model along with its interpretation which can be useful information to policy-makers.

With technological advancements in the past decade, a few studies have utilised GPS data from cars to identify the travelled path along with other observed attributes (Fosgerau et al., 2013; Frejinger & Bierlaire, 2007; Oshyani et al., 2012; Ramos et al., 2012). For example, Fosgerau et al. (2013) and Frejinger & Bierlaire (2007) expressed the utility of a route in terms of the presence of traffic signals (for making a left turn manoeuvre). Fosgerau et al. (2013) found that a traffic signal resulted in a disutility that was equivalent to 20 seconds of travel time. Frejinger & Bierlaire (2007) also found this disutility to worth around 40 seconds. It can be argued that drivers mainly experience S&G waves around signalised intersections in urban networks. Therefore, the time spent at a signal can be used to represent disutility towards the number of S&Gs. However, a limitation behind this idea is that it would consider S&G waves to occur around the locations of traffic signals only. This assumption is rather restrictive as it ignores other possible locations like roundabouts, merge and diverge sections on freeways, presence of uphill sections, and other roadway geometric features where an S&G wave can occur (refer to table 2.1). This thesis, in contrast, explains disutility of a route in relation to the number of S&Gs experienced along with other route specific attributes, irrespective of the trigger and the location that leads to the formation of these waves.

Finally, section 2.3 reviewed the works on modelling S&G waves in the adaptive cruise control algorithm of modern luxury cars. The S&G-ACC function controls both the throttle and brake pedals to smoothly traverse through congested traffic conditions. The algorithm is optimised in a way such that it closely represents the actual human driving reaction, making it more safe and comfortable. The algorithm undergoes extensive testing which generally assigns higher threshold values to the constraints like maximum deceleration, etc. making the ACC system robust against driver variability. However, such high cut-off points cannot be used for quantifying the occurrences of S&G waves in a real-world scenario as they focus more on avoiding S&Gs rather than accounting for it. Moreover, these works also focussed primarily on accounting for the occurrence of the S&G phenomenon without exploring its effect on the route choice behaviour of drivers.

This thesis adds to the body of knowledge in the following way. This thesis evaluates the impact of the number of S&G waves experienced, alongside other travel related attributes, on the route choice behaviour of drivers in urban road networks. In other words, it tests the research hypothesis: *an increase in the number of S&Gs on a route increases its disutility for the driver*. Although the idea sounds intuitive, the study makes a novel contribution by not only examining the research hypothesis, but also quantifying the relationship between discomfort due to the number of S&Gs and the resulting route choice. One of the outcomes of the experiments will be the trade-off value between travel time and the number of S&Gs experienced that drivers are willing to make. The obtained trade-off value would potentially be beneficial in modifying the existing urban transportation models to better reflect the evolution of congestion in a road network by assigning some weightage on the number of S&Gs experienced on a route. This trade-off has not been estimated so far in the literature to the best of our knowledge where the closest discussion is about the duration of being in an S&G situation.

CHAPTER 3

EXPERIMENT I – PROOF OF CONCEPT STUDY

One of the first tasks was to test the credibility of the research hypothesis, which has been discussed earlier in chapter 1. A survey was conducted on a sample of university staff and students which served as a proof of concept. The survey was formulated using the Stated Choice (SC) experimental design technique. The survey was circulated through an online interface which was not only quick, but also required lesser manpower when compared to the traditional paper based method. The collected data was then quantitatively analysed using a discrete choice framework called the Random Parameter Error Component Logit (RPECL) model.

This chapter comprises five main sections which are: 1) a background to SC experiments, 2) design of the SC questionnaire for this study, 3) survey research, 4) empirical analysis of the collected data, and 5) quantitative analysis of the data. The first section gives an introduction to the SC technique highlighting its advantages and appropriateness to this survey. The second section delineates the SC design methodology that was adopted in this study, specifying the inputs (design specifications), the design procedure (D-efficient pivot design method), and the output (blocked SC tasks). The third section discusses the survey administration using the online survey instrument along with data collection. The fourth section presents the descriptive statistics of the collected data. The fifth section presents the mathematical formulation and the estimation routine of the RPECL model which is followed by results from the analysis. Finally, the chapter concludes with a discussion of key findings and limitations of the study.

3.1 Stated Choice Experiments: A Background

One of the limitations for using the Revealed Preference (RP) data, as highlighted by Hensher (2001a), is the inability to capture a non-existing or a non-perceived alternative. This limitation led to the introduction of Stated Choice (SC) experiments, which were first introduced by Louviere & Woodworth (1983) and Louviere & Hensher (1983). SC methods have become a popular mode of data collection in the fields such as marketing, health economics, environmental economics, etc. (Louviere et al., 2000). Furthermore, these methods are increasingly being used in transportation research for forecasting the impacts of a proposed alternative or a hypothetical policy decision (Hensher, 2001c). The paper by Bliemer & Rose (2011) provides an excellent review of the applications of SC methods in the field of transportation engineering. Collecting data through this method is generally time and cost effective since it does not require sophisticated instruments for the experiment. An SC experiment involves presenting a series of questions (which will be addressed as choice tasks henceforth) to a participant, where each choice task comprises the current (optional) and hypothetical alternatives. Each alternative is defined using a bundle of attributes, where each attribute can take up different levels of magnitude across the multiple choice tasks. The participant chooses the best among the presented alternatives in each choice task. The objective of conducting an SC experiment is to estimate the trade-offs among the attributes which define an alternative. The trade-off information thus obtained is used in the relative ranking of attributes on the basis of their importance. Thus, we decided to conduct an SC experiment in this study based on the merits discussed above.

Designing SC experiments, over the years, has transformed more into a science than as an art. The aim of a good SC design is to extract meaningful and significant trade-off information by presenting a minimum number of choice tasks to participants. One of the earliest design methods, the full factorial design, presented every possible combination of the attributes to participants. For example, consider that a choice task comprises 2 options, namely, the iPhone smartphone and the Samsung smartphone. Each alternative is defined in terms of 2 attributes, the phone cost and battery life. Also, assume that the phone cost and battery life can take 2 levels each. Thus, the full factorial for this experiment becomes 8 (2 X 2 X 2) choice tasks. As a general rule of thumb, if there are J alternatives, each having K_j attributes, where the attribute $k \in K_j$ has L_{jk} levels, the total number of choice tasks (T^{ff}) in the full factorial can be calculated using equation 3.1.

$$T^{ff} = \prod_{j=1}^{J} \prod_{k=1}^{K_j} L_{jk} \tag{3.1}$$

A major limitation of the full factorial design is that T^{ff} explodes as J, K_j and L_{jk} increase. For example, Hensher, (2001a) conducted an SC experiment involving the current (referred to as the status-quo alternative henceforth) along with two other hypothetical unlabelled alternatives. Every alternative was defined in terms of 6 attributes, each of which had 4 levels. This made the full factorial design of the order of 4^{12} which was impossible to present all the choice tasks to a participant in the experiment. Moreover, the full factorial design consists of several choice tasks with a dominant alternative which is the case where one of the alternatives is better in every aspect (attribute) than the other alternatives. Assuming that the participants are rational in choice making (which means selecting the best alternative), such situations do not reveal any useful trade-off information to the analyst. Another design technique, the fractional factorial, presents a set of choice tasks from the full factorial rather than the full factorial itself, thus making it less burdensome for the participants. For example, in the same study by Hensher (2001a), 64 choice tasks were selected using the fractional factorial method and further divided into 4 blocks of 16 choice tasks each. The method became quite popular until it was recently shown to suffer from a serious limitation. The fractional factorial technique picks up choice tasks at random which leads to shortcomings like the dominant alternative case, non-orthogonality among alternatives and an attribute level imbalance (Hensher et al., 2009). These shortcomings could lead to an inaccurate estimation of the model parameters and WTP measures, as found by Rose et al. (2008).

Another method, the orthogonal design technique (also known as the D-optimal design), proposed by Street et al. (2001) presents a structured way of selecting the choice tasks. The technique suggested that SC tasks should be selected such that the attributes common across alternatives should never take the same level over the experiment. In other words, the D-optimal design aims to maximise the attribute level difference for the common attributes among the alternatives (Burgess & Street, 2003; Street & Burgess, 2004; Street et al., 2005). This way, a respondent is forced to make a trade-off among all the attributes in a choice task, while orthogonality ensures an independent effect for each attribute in a choice task. However, this method has been found to suffer from a few limitations as well. Some of these are: 1) possibility of getting a dominant alternative case, dropping which would lead to a loss of orthogonality, 2) unsuitability for labelled SC experiments, 3) it assumes zero prior information for the attributes, and 4) the design may promote lexicographic choice behaviour where only a few key attributes influence the overall choice behaviour of the participant (ChoiceMetrics, 2012).

A recently introduced design technique, referred to as the D-efficient design, follows another principled approach for selecting the SC tasks, which overcomes the shortcomings of the D-optimal strategy. A D-efficient design selects choice tasks such that it generates parameter estimates with the minimum possible standard errors. Thus, it provides a design which has the highest t-statistic for the attributes defining the alternatives. The standard errors are obtained by taking the square root of the diagonal elements of the Asymptotic Variance Covariance (AVC) matrix. The AVC matrix can be determined if there exists some prior information on the parameters. This prior information can be obtained either from the literature or by conducting a pilot study. A few advantages of using the D-efficient design are: 1) it requires a relatively small sample size than the D-optimal method for determining significant parameter estimates (Bliemer & Rose, 2011), 2) it utilises prior information on the parameters for the design which makes it better than the D-optimal design that assumes no prior information (Bliemer & Rose, 2005), and 3) it provides the analyst with an additional flexibility to construct the SC experiment in accordance with the econometric model that will be later applied during data analysis.

The efficiency of the D-efficient design can be determined from the AVC matrix. Instead of evaluating the entire AVC matrix, it is computationally convenient to determine the same for a single participant. Thus, a commonly used efficiency indicator, the D-error, uses the AVC matrix for a single respondent. Equation 3.2 shows the formula to evaluate the D-error.

$$D - error = \det(\Omega_1(X, \tilde{\beta}))^{1/K} \qquad (3.2)$$

Where Ω_1 is the AVC matrix for the respondent using the observed attributes X and prior parameter values $\tilde{\beta}$. The AVC is the negative inverse of the expected Fisher information matrix, where the latter is equal to the second derivative of the likelihood function (Train, 2009). K represents the number of parameters to be estimated. Smaller the D-error value more efficient is the design.

Both the D-optimal and D-efficient designs present the same (fixed) attribute levels to all the participants. This creates a problem where the participants are overly optimistic or pessimistic about the attribute levels being presented to them. For example, consider that an individual has a travel time of 20 minutes while driving to work by car. The person could depict inaccurate perception, and the resulting selection, if he/she is presented 80 minutes as the travel time in a choice task. On the other hand, the person would have a lower perception error had the attribute level been 30 minutes. This led to the introduction of the pivot or reference alternative design which uses the respondent's knowledge base to derive attribute levels of the hypothetical alternatives in the SC experiment (Hensher, 2004; Rose et al., 2008). In other words, the attribute levels of the hypothetical alternatives are pivoted around the revealed (status-quo) attributes of the participants. This helps in minimising the perception bias towards the levels which are much higher/lower than what the participants usually experience. Each segment of individuals comprises its own reference alternative, which represents the base case for all the individuals belonging to that segment. Rose et al. (2008) listed different approaches for developing an efficient pivot design which included: a single design using the population average (homogeneous); segmenting population based on the reference alternatives (heterogeneous); determine an efficient design on the fly, and a two-stage design process. Of these approaches, the single design using the population average performed well; close enough to the heterogeneous design. Also, the four methods were found to outperform the orthogonal design which assumes no or zero prior information.

We used the D-efficient homogeneous pivot design technique in our experiment because of two main reasons. First, previous studies by Hensher (2001a) and Rose et al. (2009), which expressed disutility of a route in terms of the time spent in stop-&-go (S&G), presented the SC tasks comprising the status-quo and two other hypothetical routes. However, the studies used the fractional factorial design, which is now widely accepted to suffer from some serious limitations. Thus, we maintained a similar experiment layout by using a more structured design strategy. This would not only lead to better parameter estimates, but also give us an opportunity to compare our results with these studies. Secondly, the sampled pool of the participants was expected to show different travel characteristics. Thus, maintaining fixed attribute levels across the participants could potentially cause a magnification of the perception error (towards an attribute level that is much higher/lower than what the participants usually experience) which could seriously undermine the obtained results.

3.2 Methodology Adopted for the SC Design

Upon zeroing in on the D-efficient homogeneous pivot design as the suitable design technique for this study, the next task was to generate the SC tasks following this approach. As with any other design process in general, it is imperative to first accumulate the inputs, or the design specifications, required for the design of the experiment. These design specifications are then synthesised using a set procedure to get the final output, i.e. the set of SC tasks. Figure 3.1 shows the schematic representation of the SC experiment design process which was adopted in this study. The rest of this section elucidates each of the three steps shown in the figure.

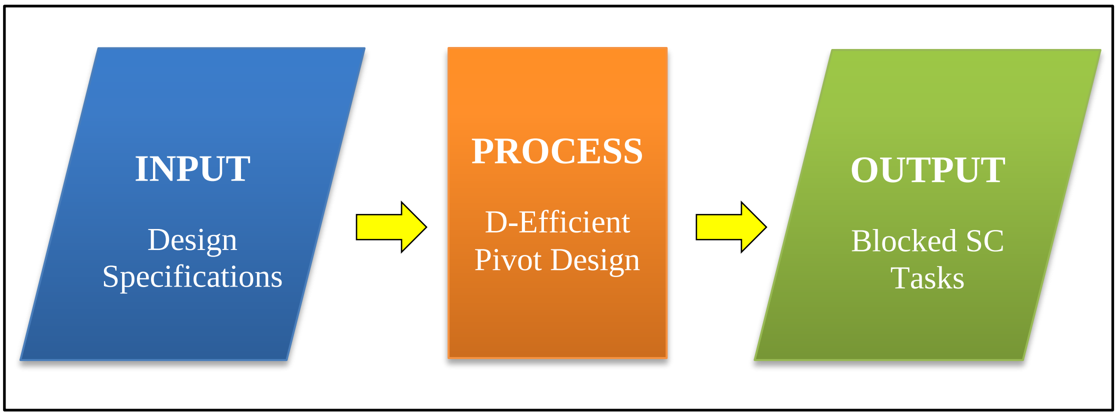

Figure 3.1: Schematic diagram of the SC experiment design methodology

3.2.1 Design specifications

This subsection describes the necessary inputs for the design of the experiment and how these were acquired.

3.2.1.1 Number of alternatives

We decided to present 3 alternatives in every choice task, namely, the status-quo and two unlabelled hypothetical routes. The reasons behind specifically selecting the 3 alternatives were: 1) we wanted to be consistent with the previous studies (on route choice as a function of the time spent in S&G (Hensher (2001a) and Rose et al. (2009))) for a comparison of the results at the end, and 2) it would give additional flexibility to the participants to continue with their currently travelled route instead of forcing them to select between the two hypothetical routes.

3.2.1.2 Attributes and their levels

We expressed each of the three alternatives in terms of 4 attributes, namely, 1) the total travel time, 2) the time spent in S&G conditions, 3) the number of S&Gs experienced, and 4) the vehicle running cost. Each attribute was further defined using 5 levels. Table 3.1 presents the attribute levels that were selected for the design in this experiment. The attributes and levels for the time spent in S&G and the vehicle running cost were taken from the previous study by Hensher (2001a). Maintaining the same levels for these two attributes would facilitate comparison of the results at the end. The number of S&G experienced was the newly added attribute which is also the novel contribution of this thesis.

Table 3.1: Attributes and their levels adopted for the SC design

Attribute name	**Levels**
Travel time (minutes)	-20%, -10%, 0, 10%, 20%
Time spent in stop-&-go traffic (minutes)	-50%, -25%, 0, 25%, 50%
Number of stop-&-gos experienced	-50%, -25%, 0, 25%, 50%
Running cost of vehicle (AU $)	-25%, -12.5%, 0, 12.5%, 25%

3.2.1.3 Number of choice tasks and blocks

It was decided to keep the number of choice tasks to 10 per participant. The reasons behind specifically selecting this number were: 1) since our each choice task would consist of 3 alternatives with 4 attributes each, presenting 10 choice tasks to each participant would potentially provide us with unbiased responses without exposing them to a higher cognitive load, and 2) the SC design presented to each participant in the study by Hensher (2001a) comprised 16 SC tasks using 6 attributes in all. As we selected 4 attributes for our study, presenting 10 choice tasks per respondent ($16/6 \times 4$) was considered an adequate number. Furthermore, we selected a block design approach where two sets of 10 choice tasks each would be generated. Each of the two blocks would later be used in the actual survey such that they almost have an equal representation in the collected data.

3.2.1.4 Prior attribute information

We referred to the previous studies by Hensher (2001a) and Rose et al. (2009) to determine the estimated coefficients for the 4 attributes we used in this study. From the survey of the literature, we identified that the parameters for travel time, vehicle running cost and time spent in S&G traffic were highly negative and statistically significant, which indicated disutility towards travel. Since, none of the past studies looked at the impact of the number of S&Gs, we considered it to bear a negative sign as well (indicating disutility). Thus, we set the initial prior values that were negative and close to zero for all the 4 attributes under consideration. An initial SC experiment was generated using these prior values and was then used to conduct a pilot survey.

We received responses from 25 participants in the pilot survey, which equated to 250 rows of observations (as each person responded to 10 choice tasks). The collected sample was of a reasonable size and was slightly bigger than the one used by Collins et al. (2007) in their pilot survey. The dataset was quantitatively analysed using the random parameter logit model (Train, 2009). Table 3.2 shows the parameters that were obtained from data analysis of the pilot survey data. These parameters were further used to re-design the SC experiment for the full (final) survey. For the full survey design, we specified the prior parameters to follow a mixing distribution. A normal distribution was assumed for the parameters total travel time, time in S&G and number of S&Gs. Vehicle running cost was treated as a non-random parameter in the design. Further, a Bayesian normal distribution was assumed for the mean and standard deviation of all the prior parameters. Bayesian distributions are useful as they consider the mean and standard deviation values to be in turn derived from an underlying distribution. This improves the accuracy of the design as the prior values are not treated as fixed, thus accounting for unobserved variation (Sándor & Wedel 2002).

Table 3.2: Prior parameter values and distributions assumed for the full design

Attribute	Mean	Std. dev.	Distribution for full design
Total travel time (minutes)	-0.2746 **	0.1923 **	[n,(n,-0.2746,0.0544),(n,0.1923,0.0526)]
Time in S&G (minutes)	-0.1228 **	0.1843 **	[n,(n,-0.1228,0.0379),(n,0.1843,0.0449)]
No. of S&Gs	-0.1382 **	0.2310 **	[n,(n,-0.1382,0.0654),(n,0.2310,0.0770)]
Vehicle running cost (AU $)	-1.5784 **	--	[(n,-1.5784,0.3706)]

** significant at 95%; (n,μ,σ) where n is normal distribution, μ and σ are the mean and standard deviation respectively

Table 3.2 also shows the Bayesian distribution which was assumed for the priors. For example, the prior distribution of travel time is represented by the expression [n,(n,-0.2746,0.0544),(n,0.1923,0.0526)]. The expression implies that the travel time parameter follows a normal distribution (the first instance of n in the expression) with some mean and standard deviation. The mean of this distribution is in turn derived from another normal distribution (the second instance of n in the expression) having a mean of -0.2746 and a standard deviation of 0.0544. Similarly, the standard deviation of the travel time distribution is also normally distributed (the third instance of n in the expression) with a mean of 0.1923 and standard deviation of 0.0526. Looking at another example, the prior distribution of vehicle running cost is given by the expression [(n,-1.5784,0.3706)]. It implies that the prior is normally distributed having a mean and standard deviation of -1.5784 and 0.3706 respectively. In this expression, the mean and standard deviation are considered fixed and do not follow any distribution, as was the case with the previous example.

3.2.1.5 Population segments and their reference alternatives

The study was planned to be conducted on UNSW staff and students who drove to the university campus by car. Thus, to come up with a homogeneous pivot design, it was necessary to first segregate the survey population into well-defined segments. We referred to the UNSW travel survey report for the year 2014, an annual document released by the university, which gave the travel pattern (information like travel time, distance, mode of commute) of staff and students to UNSW Kensington campus (UNSW 2014). Table 3.3 classifies the population of car users who drive to UNSW into six segments on the basis of travel time.

Table 3.3: Segment-wise weightage and reference alternative used for the design

Segment	Travel time range (minutes)	Weightage (%)	Reference alternative taken for the design			
			Travel time (minutes)	Time in S&G (minutes)	Number of S&Gs	Running cost (AU $)
1	0-15	19.05	7.5	2	5	0.85
2	16-30	16.28	22.5	6	12	2.45
3	31-60	20.67	45	11	20	5.3
4	61-90	28.92	75	19	30	12.0
5	91-120	13.43	105	26	42	19.3
6	> 120	1.65	135	34	50	21

The table shows that close to 65 percent of the population has a travel time of greater than 30 minutes which indicates that commuters do not stay in the vicinity of the campus and have longer driving periods. Thus, there is a higher likelihood that the survey sample experiences S&G traffic while driving to UNSW on a regular basis. The reference alternative for each segment was defined following a certain set of rules which are given below:

- The travel time attribute was set as the mean of a segment's travel time range
- Time spent in stop-&-go (S&G) was kept between 20 to 25 percent of the travel time attribute (as found by Hensher, (2001b))
- Number of S&Gs experienced was set between 30 to 40 percent of the upper limit of the travel time range for a segment
- Vehicle running cost was calculated as the product of cost per kilometre (AU $0.15) and the travelled distance for each segment (RACQ, 2014).

The segment-wise weight and the reference alternatives were used as the design specification. The D-efficient homogeneous pivot design technique would yield a single SC experiment for the entire survey sample, i.e. the choice tasks (a combination of attribute levels, or proportions in this case) remain the same across the participants.

3.2.1.6 Model specification for the design

The D-efficient SC design technique also required the econometric model according to which the SC tasks would be generated. We selected a Random Parameter Error Component Logit (RPECL) model specification to generate the SC experiment for this study. Given the panel nature of the dataset, where the participant provides multiple responses, the reasons behind selecting this model were: 1) to express the unobserved errors (in the overall utility) in terms of the preference heterogeneity across individuals (Train, 2009), and 2) to capture the correlations across choice tasks arising due to the presence of a time-invariant status-quo alternative (Hess & Rose, 2008; Train & Wilson, 2008).

3.2.2 Design of the SC experiment

Once the design specifications were decided, we used these inputs to generate the D-efficient homogeneous pivot design. The entire design procedure was carried out using the stated choice (SC) experimental design software package Ngene (ChoiceMetrics, 2012).

As our plan was to analyse the dataset using a mixed logit model (which has been used in the previous studies by Hensher (2001a) and Rose et al. (2009)), the objective of the design was to generate a homogeneous pivot design on the lines of the mixed logit model specification (which we addressed as the RPECL in subsection 3.2.1.6). Thus, the proposed method resulted in a complex design, the solution space of which was constrained by multiple conditions (design specifications). The binding constraints that we specified in our design were:

- No dominant alternative required
- Using the RPECL specification for the panel dataset
- Evaluating the Fisher information matrix across the six segments
- Three parameters following a Bayesian normal distribution
- Including the error component part to the utility equations of the three alternatives
- Using a simulated sample of 200 respondents

- Generating random parameter and Bayesian draws using Gaussian quadrature with an abscissa of 2
- Generate 2 blocks of 10 choice tasks each

All these constraints made the choice task selection and evaluation of its efficiency more complicated and time-consuming (ChoiceMetrics, 2012). We found a solution to overcome this computational hurdle in the Ngene software manual and its user's forum (Bliemer (2014); ChoiceMetrics (2012)). The solution steps were as follows:

- Estimate the MNL model on the pilot survey dataset. Call it as model-MNL
- Input the estimated parameters from model-MNL (presented in table 3.4) as prior values in the experiment design code also specifying the model as MNL. Let's call the design code as code-MNL
- Save the resulting SC tasks from code-MNL in a file. Let's call it design-MNL
- Estimate the mixed logit model on the same pilot survey dataset. Call it as model-RPECL
- Prepare another design code, code-RPECL, using the prior parameter values obtained from model-RPECL (presented in table 3.2) and specify the model as RPECL
- Evaluate the efficiency of design-MNL using the specifications given in code-RPECL

This process greatly reduced the level of complexity and computational time as the MNL specified design was much easier to extract when compared to the highly constrained design. The efficiency of the resulting design was evaluated next by assuming it to be generated through the complex RPECL specified design. In other words, the proposed steps indirectly provided a design as per the RPECL specification consuming much lower time and resources.

As the design procedure involved multiple constraints (discussed above), it was quite computationally challenging and time consuming to generate a single design of 20 choice tasks split into 2 blocks. For example, a single evaluation of the complex design took around 40 minutes on a 2.4 GHz processor with 8GB RAM. In other words, executing the script for all possible combinations would have become time infeasible. Thus, the above steps were repeated twice with the evaluated designs (two separate sets of design-MNL) having 10 choice tasks in each design (Bliemer, 2015b). Therefore, we also estimated an MNL model on the pilot survey dataset. Table 3.4 shows the parameter estimates that were obtained from the model. These estimates were used as the priors in the design procedure code-MNL.

Table 3.4: Prior parameter values for model-MNL on pilot study data

Parameter	**Estimated value**
Total travel time (minutes)	-0.136 **
Time in S&G (minutes)	-0.0410 **
No. of S&Gs	-0.0236 **
Vehicle running cost (AU $)	-0.694 **

** significant at 95%

Thus, we followed the steps discussed above and prepared the two design scripts, code_RPECL and code_MNL, using the prior information given in tables 3.2 and 3.4 respectively along with the other design specifications. The two scripts are available at the weblink given in appendix B of this thesis. We do not discuss the technical (coding) details of the script in this thesis and would instead encourage the readers to go through the software's user manual (ChoiceMetrics, 2012).

3.2.3 The resulting SC tasks

We generated two separate designs of 10 stated choice (SC) tasks each instead of a single blocked design of 20 choice tasks for the reasons discussed above. Table 3.5 shows the two sets of choice tasks that were finally obtained from this design exercise. The two blocks are identified by the field BlockID in the table. Each row in table 3.5 corresponds to the attribute levels (as proportions) for the two hypothetical alternatives. These proportions would be multiplied by the corresponding attributes of the status-quo alternative to relatively pivot the hypothetical alternatives around it. The three alternatives would thus make the choice tasks.

Table 3.5: Two blocks of choice tasks from the homogeneous pivot design

BlockID	TaskID	RI_TT	RI_TTS	RI_S&G	RI_VR	RI_TT	RI_TTS	RI_S&G	RI_VR
0	1	1.1	0.50	1.00	1.000	0.9	1.50	1.00	1.000
0	2	1.2	1.50	0.75	1.000	0.8	0.50	1.25	1.000
0	3	1.1	1.25	0.75	1.250	0.9	0.75	1.25	0.750
0	4	0.8	1.00	1.25	0.750	1.2	1.00	0.75	1.250
0	5	0.9	1.00	1.50	1.125	1.1	1.00	0.50	0.875
0	6	0.9	1.25	0.50	0.875	1.1	0.75	1.50	1.125
0	7	1.0	0.75	1.25	1.250	1.0	1.25	0.75	0.750
0	8	0.8	0.75	0.50	1.125	1.2	1.25	1.50	0.875
0	9	1.2	0.50	1.00	0.750	0.8	1.50	1.00	1.250
0	10	1.0	1.50	1.50	0.875	1.0	0.50	0.50	1.125
1	1	0.9	1.25	0.50	0.750	1.1	0.75	1.50	1.250
1	2	1.2	1.00	0.50	1.000	0.8	1.00	1.50	1.000
1	3	0.9	0.50	1.25	0.750	1.1	1.50	0.75	1.250
1	4	1.2	0.75	1.25	1.000	0.8	1.25	0.75	1.000
1	5	1.1	1.00	0.75	1.250	0.9	1.00	1.25	0.750
1	6	1.0	1.50	1.50	0.875	1.0	0.50	0.50	1.125
1	7	1.1	1.25	1.00	0.875	0.9	0.75	1.00	1.125
1	8	0.8	1.50	1.00	1.250	1.2	0.50	1.00	0.750
1	9	0.8	0.50	0.75	1.125	1.2	1.50	1.25	0.875
1	10	1.0	0.75	1.50	1.125	1.0	1.25	0.50	0.875

The D-error was used as the performance indicator to evaluate the relative efficiency of the generated block designs. The software usually keeps on exploring new designs as per the modified Federov algorithm (Cook & Nachtsheim, 1980) until all possible combinations of the attributes have been tried which satisfy the specifications defined above and then evaluates their D-error. Since the number of possible combinations in our survey was of the order $5^{4\times 2} = 390625$ (5 levels; 4 attributes and 2 alternatives), which along with longer processing time (around 40 minutes per evaluation) would have become time infeasible to converge upon a design with the least D-error (ChoiceMetrics, 2012). Thus, a general practice is to run the design code for a few hours (around 4 to 5 hours) until the D-error stabilises (Bliemer, 2015a). We also followed the same approach and the resulting blocks of choice tasks were chosen when the D-error value changed by less than one-thousandth (0.001) between the two consecutive iterations (evaluations). The D-error value upon meeting the stopping criterion for the two blocks was found to be 0.736 and 0.824 respectively (using equation 3.2). As discussed earlier in subsection 3.2.1.5, the two blocks would later be used to generate the SC tasks such that they have almost an equal representation in the collected dataset.

3.3 Survey Research

Once the Stated Choice (SC) tasks were ready, the next tasks were to: 1) decide the survey administration method, 2) design the survey instrument for the study, and 3) data collection. This section provides a discourse on the three tasks mentioned above.

3.3.1 Selecting the survey administration method

There exist a wide range of survey administration methods in practice which mainly include the traditional paper-based surveys, personal interviews, telephone surveys, and online surveys. Each survey method has its pros and cons (Research Lifeline, 2012). Until very recently, paper-based surveys generally had among the highest response rates, much more than online surveys. Nulty (2008) reviewed works across disciplines like education, health, etc which compared paper-based against online surveys and found the former to have a higher response rate by an average margin of around 25 percent. However, with the recent boom in internet patronage along with the smartphone revolution, online surveys are fast catching up on this statistic. For example, Hohwü et al. (2013) found the response rate for paper-based and online surveys as 56.2 and 53.4 percent respectively in their study on children health and welfare. Apart from that, online surveys have other advantages like: 1) these are quite cost effective as they do not require additional manpower (in the form of survey staff) during data collection, 2) a well programmed online survey, which can be accessed especially through smartphones and tablets, can considerably increase the number of responses, and 3) data assembly and storage is quite convenient in contrast to the manual data entry required for paper-based surveys. Due to these advantages along with a higher response rate as observed by studies discussed above, we decided to circulate our survey instrument through an online medium.

3.3.2 Layout of the survey instrument

The online survey was thus programmed for this study. The survey instrument was developed in-house using the programming languages like HTML (primarily for the web page design), Jquery and Java Script (mainly for the back end logics and validations) and PHP (to store the responses in an SQL database). The webpage was developed using the software tool Wamp Server (WampServer, 2017). The URL for the survey webpage can be found in appendix B of this thesis. The online survey, with the title *Route choice in stop-&-go conditions*, comprised the following layout:

- Introduction to the survey
- Revealed travel characteristics
- Route choice tasks
- Socio-demographic Information
- Advertisement for the driving simulator study

3.3.2.1 Introduction to the survey

This section first welcomed the participants and informed them about the aims of the study, contents of the survey, expected completion duration, and a chance of winning gift vouchers worth AU $20 through a lottery. The amount was decided based on the following: 1) The available funding which was AU $100 (given that this was a proof of concept study), and 2) we anticipated the survey duration (based on initial testing of the survey) to be around 15 minutes and this compensation was well above the average existing wage rate of the sample population. More details on the observed survey duration and average wage rate will be presented in section 3.4 of this chapter.

The survey then moved on to the part which defined the stop-&-go (S&G) wave, time spent in S&G and number of S&Gs. The description of S&G waves shown was: "*Stop-&-Go waves are characterised by the sudden braking, followed by acceleration, of vehicles. These waves are often prevalent in congested traffic conditions on urban road networks. Under stop-&-go conditions, vehicles are forced to decelerate and travel at slower speeds or even come to a halt, before accelerating again, many times over the duration of the trip.*" The number of S&Gs was described as follows: "*The number of times one experiences the situation of decelerating to a halt and then accelerating again while driving.*" Similarly, the time spent in S&G traffic was defined as: "*The travel time that you spend in stop-&-go traffic. This component of travel time is all included in the total travel time.*" A focus group was conducted to finalise the wording of these definitions to ensure that respondents have a proper idea of S&G waves in the survey. An animated video (refer to appendix B for the weblink to the survey page) was also shown to the participants to increase their awareness about this phenomenon and help them better associate with such conditions, if they experienced any. From the modelling (data analysis) standpoint, the reason behind showing the video was to minimise, if not eliminating, the measurement or the perception bias that might have been introduced if the participants found it hard to precisely recollect the number of S&Gs experienced during the travel. The implications of this bias will be discussed in section 3.6 of this thesis.

3.3.2.2 Revealed travel characteristics

In this section, the participants were asked to recollect their most recent trip to work on a weekday morning. Based on their travel experience, they were supposed to provide details on their total travel time, time spent in S&G, number of S&Gs and the distance travelled. The vehicle running cost was indirectly calculated using the distance and multiplying it with AU $0.15 per kilometre (RACQ, 2014). There were two things that the participants were made aware of: 1) the total travel time was inclusive of the time spent in S&G, and 2) the car running cost represented the average fuel cost incurred during the trip. Validation checks were applied on these fields to ensure that the participants did not give null or infeasible values. Apart from the non-negativity check, the upper thresholds of the attributes were constrained as given below:

- Travel time should be less than or equal to 250 minutes
- Time spent in S&G should be less than or equal to 250 minutes
- Number of S&Gs should be less than or equal to 250
- Distance travelled should be less than or equal to 300 kilometres

3.3.2.3 Route choice tasks

The route choice section first presented a scenario to the participants where they had two alternate hypothetical routes along with the current route in their choice set. All three routes had a similar distance but varying traffic conditions prevailing on them. The participants were then informed about the 10 choice scenarios (tasks) where each scenario represented varying traffic conditions on the three routes. Based on their judgement, the participants were asked to select the most preferred route for travelling to work among the three candidate routes. An example was also shown which explained the contents of the choice task and how the participants should record their responses. Figure 3.2 shows the example that was presented to the participants. The example was then followed by the set of 10 Stated Choice (SC) tasks. The response was mandatory for every choice task and the back button of the web browser, which lets one go back to the previous question, was disabled in this part of the survey. This was done to prohibit the participants to alter their response to the previous set of questions based on the later ones. The participants were informed about this restricted browsing feature before commencing with the actual 10 choice tasks. Additionally, the order of the choice tasks presented was randomised across the participants to counter the learning effect.

3.3.2.4 Socio-demographic information

The socio-demographic section followed the SC tasks where the participant information was collected. A set of 11 questions were asked in this section, which included gender, age, annual gross income, occupation, household size, number of cars owned, etc. A question on their frequency of commute to work by car was also asked in this section. The response to each of these questions was mandatory.

Section 3: Route Choice Section

Example:

Consider a respondent gave the following details for a recent trip to work:

Total travel time (minutes) = **25**

Time spent in stop-&-go traffic (minutes) = **8** *(included in the total travel time)*

Number of stop-&-gos experienced = **12**

Average car running cost (AU $) = **1.2**

The following table is then shown to the person as one of the choice scenarios.

SCENARIO 1 of 10

	Current Route	Route-1	Route-2
Total travel time (minutes)	25	28	23
Time spent in stop-&-go (minutes)	12	18	6
Number of stop-&-go experienced	8	6	10
Average vehicle running cost (AU $)	1.2	1.05	1.35
* **I would choose:**	☐	☐	☐

Next

NOTE:
1. For the 10 scenarios, the backspace key and back button of your browser will be ***disabled***
2. You ***must*** (*) select the most preferred route in each choice scenario, based on your judgement.

Please select from the alternatives carefully as it is important for a sound analysis.

Figure 3.2: An example scenario shown to the participants

3.3.2.5 Advertisement for the driving simulator study

Towards the end of the survey, the participants were informed about a follow-up driving simulator study in the future. The section asked for their willingness to participate in the study once it starts. Interested participants were asked to provide their email address which would be used in the future to invite them for the study. The section showed a few images of the driving simulator setup and subjects driving in a virtual scenario to increase the participant awareness and to potentially increase the response rate for the future study. It also warned the participants prone to motion sickness and pregnant women to avoid participating in the study.

3.3.3 Data collection

All necessary clearances from the university human ethics committee, Human Research Ethics Advisory (HREA) Panel H: Science and Engineering, were taken prior to data collection. The copy of the approval letter (HC No. 15007) is available in appendix C of the thesis. We ensured that the survey instrument does not pose any health and safety hazards and maintained the confidentiality of the participant information.

Individuals who drove to UNSW, Kensington campus for work by car at least thrice a week were selected for the analysis. The exclusion criteria to the survey included people who: 1) did not drive to UNSW for work, 2) did not possess a driver's license, and 3) were undergraduate or postgraduate students. A respondent satisfying any of these criteria was dropped from the analysis. The reason behind using these exclusion criteria was to obtain a sample of car drivers who even had obligatory duties like teaching, administrative works, etc. Postgraduate research (Ph.D.) students were also considered as staff in this study as they too have research and teaching obligations to be met. The selected sample represented work trips to UNSW where a person is required to be on time at work to perform their duties. Studying non-work trip purposes, which characterise a different travel behaviour (for example, drivers won't mind travelling longer if on an outing), were beyond the scope of this thesis and have been discussed in the future research directions in chapter 6 of this thesis.

A set of rules were coded in the SC tasks to avoid situations arising due to unreasonable combinations of the attribute levels for the hypothetical alternatives. The used logics were as follows: 1) setting the minimum and maximum values on the revealed travel attributes (10 and 150 minutes for the travel time; 2 and 120 minutes for the time spent in S&G; 5 and 150 for the number of S&Gs, and AU $0.75 and AU $18 for the vehicle running cost), and 2) bounding the ratio of the time spent in S&G and the travel time between 0.2 and 0.6. In order to avoid confusion during the choice tasks, respondents were informed that the total travel time was inclusive of the time spent in S&G traffic.

The survey was circulated between Tuesdays and Fridays for around 20 weeks. The link to the survey was emailed to the administrative managers of roughly 15 schools within the university, asking them to forward the link among their staff. Additionally, the link was also circulated via the university's explode email, a group mainly comprising university staff. Strategies like sending repeated reminder emails along with incentivising the survey in the form of prizes to respondents, awarded through a lottery, were used to further improve the response rate (Nulty, 2008). At last, a total of 200 responses were received from the main survey. The survey completion time was found to be between 15 and 18 minutes across all the participants.

3.4 Empirical Analysis

The collected data was first cleaned and a few incomplete or invalid responses were dropped from further analysis. The responses that were dropped included: 1) *incomplete responses:* where the participants could not complete the survey due to reasons like losing interest during the survey and using other web browsers (like internet explorer) instead of the recommended ones (Google Chrome, Mozilla Firefox and Safari) to fill the survey, and 2) *left biased responses:* where the participant always selected the left most option, the status-quo alternative, across all the 10 choice tasks. We analysed these responses closely and found that the participants still selected the status-quo alternative when other alternatives were better on few attributes. That is, the participants never traded-off on any attribute and had a zero willingness to shift away from the currently travelled route. The left biased responses constituted roughly 8 percent of the total collected data, which being a small proportion, were discarded from data analysis.

The resulting dataset (will be addressed as the effective dataset henceforth) comprised 145 participant responses, which equated to 1450 stated choice (SC) observations. We discuss the descriptive statistics of the effective dataset in this section. A thorough empirical analysis gives us useful insights on selecting an appropriate quantitative data analysis technique.

Figure 3.3 shows the distribution of the four revealed travel related characteristics in the effective dataset. Nearly three-quarters of the participants incur a vehicle running cost up to AU $3 which equates to roughly 20 kilometres of travel to UNSW from their homes. It indicates that most of the participants stay within a 20 kilometre radius of the UNSW campus.

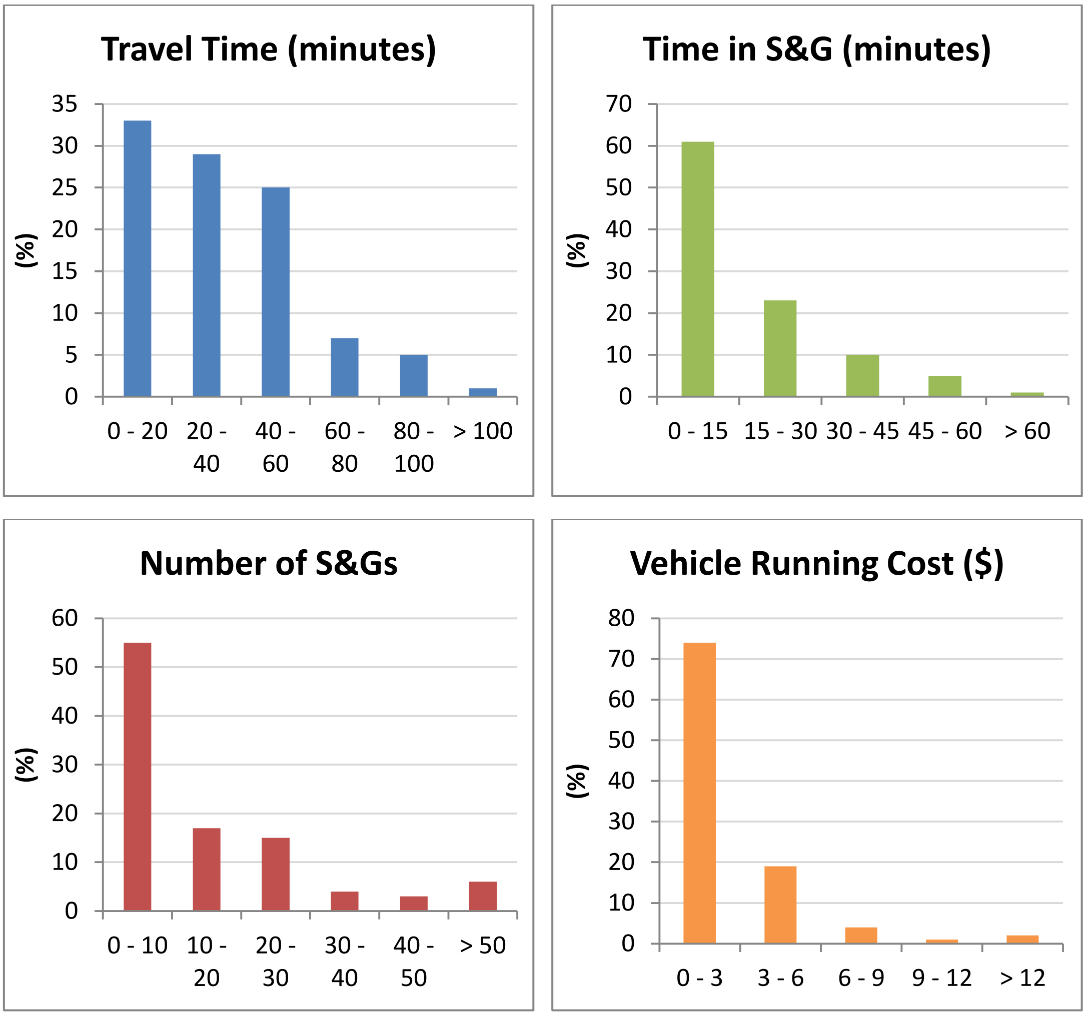

Figure 3.3: Distribution of travel characteristics of the participants

For this distance, the other three graphs reveal interesting information. The travel time for almost half the sample is between 20 to 60 minutes which makes it more likely for the drivers to experience traffic congestion on the way to the campus. The conjecture is further corroborated from the time spent in stop-&-go (S&G) graph which shows that nearly 80 percent of the sample spends up to half an hour in S&G traffic. They also undergo multiple cycles of S&G, with a majority of the population experiencing up to 30 S&Gs during the travel to work. Thus, we can infer two key points from this figure: 1) the sampled respondents have to travel through congested traffic on a regular basis. Thus, they are expected to have a better association with S&G traffic and can approximately recollect the number of S&Gs experienced, and 2) the graphs show significant variability in the revealed travel related attributes which potentially indicates heterogeneity in the travel behaviour of the participants. That is, the participants travelling for 60 minutes are expected to show different preferences towards travel characteristics in comparison to the ones driving for 20 minutes.

Table 3.6 gives some more statistics on the trends observed in figure 3.3. The table shows the mean of the travel related information of the effective dataset to be on a higher side. For example, the average driving time and the vehicle running cost are around 40 minutes and AU $2.6 respectively. It implies that a majority of the sampled respondents have a travel time of more than 30 minutes. Nearly 48 percent of the total travel time is observed to be spent driving in stop-&-go (S&G) traffic with an average of 18 stops experienced. It shows that sampled drivers have to experience congested driving coupled with S&G conditions while travelling to UNSW during morning peak hours. The minimum, maximum and percentile columns convey that each travel specific attribute has a wide range and the distribution shown in figure 3.3.

Table 3.6: Summary of descriptive statistics of the effective dataset

Data	**Mean**	**Std. dev.**	**Min.**	**Max.**	**20^{th} percentile**	**80^{th} percentile**
Travel Time (minutes)	38.13	23.33	5	120	18	55
Time in S&G (minutes)	18.24	15.81	1	75	5	30
No. of S&Gs	17.61	27.06	1	100	4	30
Running cost (AU $)	2.61	2.55	0.3	15.75	0.9	3.75

Figure 3.4 shows the descriptive statistics of the socio-demographic characteristics of the effective dataset. The survey received a good number of responses from females when compared to males. The age distribution shows that roughly 68 percent of the population is above 30 years which generally corresponds to the lower age limit of university staff.

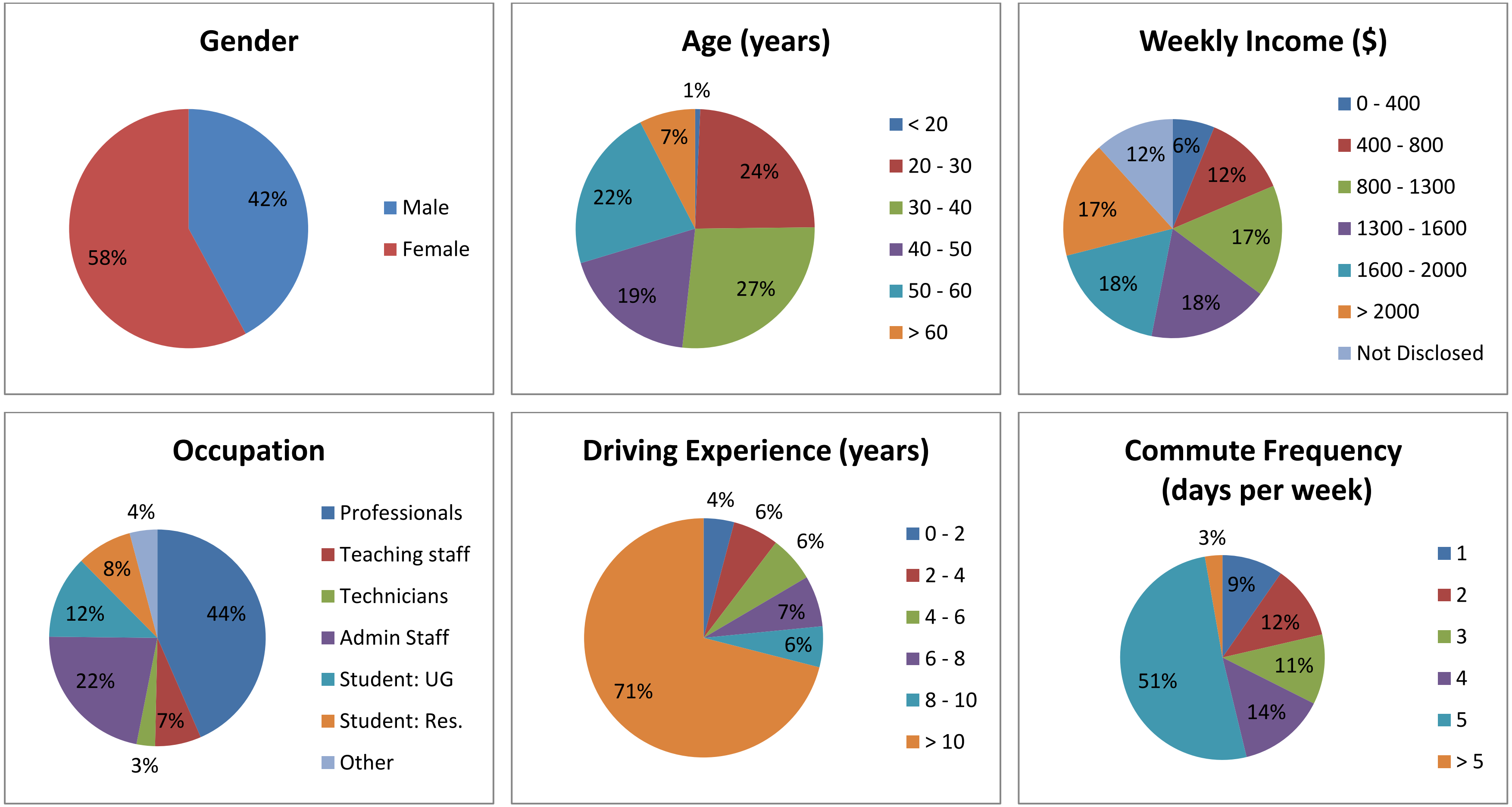

Figure 3.4: Descriptive statistics of socio-demographic variables of the collected data

The remaining participants are between 20 to 30 years which represents the age limit of the majority of the post-graduate research students in the university. The weekly income graph has a good representation (most of the segments are more than 10 percent) across different segments. It shows that roughly 30 percent of the sample has an income of up to AU $1,000 per week which corresponds to the regular stipend of research students. The remaining sample belongs to a higher income bracket which includes skilled labour, technicians, administrative and teaching staff (refer to the pie chart on the occupation status for more details). The average hourly wage rate for the effective dataset is calculated as a minimum of AU $25.50, assuming 52 weeks and 45 hours of work per week. Nearly half of the participants drove to the UNSW campus on all 5 days of the week which has two implications: 1) they generally belonged to a high-income group which could afford the daily fuel, toll and parking costs, and 2) the regular drive to work would have exposed them to the phenomenon of S&G waves and would have helped them to recollect the number of S&Gs experienced in the travel. Thus, in summary, we concluded that the selected sample was suitable to test the relationship between route choice and the number of S&Gs experienced. As a sidenote, it is worth mentioning that we could not validate whether the collected sample was a representative of the sampling frame of interest, i.e. participants who drove at least thrice a week to UNSW for work in the morning, due to lack of available data.

3.5 Discrete Choice Analysis

The stated choice dataset thus obtained was further analysed to understand the route choice behaviour of the sampled respondents and to test the validity of the proposed hypothesis. Additionally, we were also interested in determining the trade-off, also known as the willingness to pay (WTP), between the attributes that the participants made while assessing the different routes presented to them across the choice tasks. Thus, we used a discrete choice modelling technique for the quantitative analysis of the dataset. In this section, we review the various discrete choice models in practice and select the most suitable model for this study.

One of the earliest models, the logistic regression, proposed by McFadden (1973) provided forecasts on the mode choice of individuals based on the alternate specific and person specific attributes. The model has salient features like: 1) a closed form solution which is easy to compute, and 2) it provides the relative impact of attributes (on the choice) along with the WTP measures. This seminal model is widely used in practice to date and has served as the pivot in the quest for other econometric frameworks. The nice closed form of the logit kernel is a result of a few underlying assumptions like the unobserved errors follow an EV-1 (Gumbel) distribution and are Independent and Identically Distributed (IID). The consequence of these, especially the IID assumption, is the well-known Independence of Irrelevant Alternatives (IIA) property which imposes a proportional substitution pattern among the alternatives (Ben-Akiva & Lerman, 1985). Moreover, the IID assumption has been found to be quite restrictive in a few situations. For example, the logit model is ill suited to analyse stated choice (SC) datasets where few unobserved errors such as the individual taste remain the same across the multiple choice tasks (Hensher, 2001a). As a result, numerous models have been proposed by researchers over the years which relax (partly of fully) the assumptions made in the logistic regression. Table 3.7 tabulates a few studies which compared the logit model against more flexible frameworks and found the logit model to under estimate the value of time (VoT) and other WTP measures on most of the occasions.

Over the years, the mixed logit model has evolved as a better tool to analyse the choice data. Hensher and Greene (2003) discussed the current state-of-the-art and application of the Mixed Logit model (MXL) in the field of transportation engineering. The model is particularly useful in SC experiments where an individual is subjected to a series of choice tasks. The MXL relaxes the IIA condition by splitting the unobserved component of the overall utility into two components, a user defined part and an idiosyncratic part. The user defined part can be reshaped to form a Random Parameter Logit (RPL) or an Error Component Logit (ECL) model (Train, 2009). While the RPL formulation explains preference heterogeneity among individuals, ECL captures the correlation among the unobserved component of utility across alternatives and multiple choice tasks (Hess & Rose, 2008; Train, 2009).

Table 3.7: Comparison of WTP using logit and other models

Author(s) (year)	**Data type**	**Response variable type**	**Models used**	**VoT MNL**	**VoT other model**
Bhat (1995)	RP	Multinomial	MNL & HEV	14.7^{US}	20.8^{US}
Hensher (2001a)	SP	Multinomial	MNL & MXL	14.8 ^{AU}	20.7 ^{AU}
Hensher (2001b)	SP	Multinomial	MNL, MNP & H-MNP	4.6 ^{AU}	MNP: 5.3 ^{AU} H-MNP: 7.6 ^{AU}
Hensher (2004)	SP	Multinomial	MNL & MXL	10.8 ^{AU}	6.1 ^{AU}
Hess, Bierlaire, & Polak (2005)	Simulated	Binary	MNL & MXL	17.2 ^{AU}	62.8 ^{AU}
Phanikumar & Maitra (2006)	SP	Multinomial	MNL & MXL	32.1^{IN}	46^{IN}

^{AU} value in AU \$/hr ^{IN} value in INR/hr ^{US} value in US \$/hr

MXL: Mixed logit *HEV:* Heteroscedastic extreme value

MNP: Multinomial probit *H-MNP:* Heteroscedastic multinomial probit

RP: Revealed preference data *SP:* Stated preference data

The ECL model is particularly useful in analysing the pivot design SC experiment data where the hypothetical alternatives are pivoted around the Status Quo (SQ) alternative. The presence of the SQ alternative gives rise to two kinds of correlations: 1) between the SQ and hypothetical alternatives as the former remains unchanged while the latter are forced to vary in every choice task, and 2) between the non-existing hypothetical alternatives which might be highly correlated with each other than the reference alternative (Train & Wilson, 2008). These correlations cause systematic substitution patterns which lead to an inaccurate estimation of the WTP measures. Thus, adding normally distributed (ND) error components to the deterministic component of the utility of alternatives can better account for the correlations (in the unobserved part of the utility) arising due to the presence of the SQ alternative (Hess & Rose, 2008; Scarpa et al., 2005).

This study used a Random Parameter Error Component Logit (RPECL) model, which was a combination of both the RPL and ECL specifications. The model not only accounted for the taste heterogeneity across individuals, but also captured the correlation (in the unobserved part of the utility) across the multiple choice tasks that arose due to the presence of the time invariant SQ (reference) alternative in the pivot design SC experiment (Hensher, 2008; Hess & Rose, 2008). In other words, the model is particularly useful in the case of pivot design SC experiments. We now discuss the model formulation of RPECL in the following subsection. We adopt the following formatting styles while discussing matrix algebraic notations used in the model formulation: 1) *scalar* quantities are written in italics, 2) ***vectors*** in italics and bold face, and 3) **matrices** in bold face.

3.5.1 Model formulation

Consider individual $n \in N$ where N is the total number of participants surveyed. The individual n faces T choice tasks within the experiment. Each choice task $t \in T$ comprises three alternatives, the SQ and two other hypothetical alternatives that are pivoted around the first. According to Hess & Rose (2008), the general utility specification is given by equations 3.3 – 3.5:

$$U_{n1t} = \boldsymbol{\beta'_n X_{n1}} + \sigma\xi_{n1} + \varepsilon_{n1t} \tag{3.3}$$

$$U_{n2t} = \boldsymbol{\beta'_n X_{n2t}} + \sigma\xi_{n2t} + \varepsilon_{n2t} \tag{3.4}$$

$$U_{n3t} = \boldsymbol{\beta'_n X_{n3t}} + \sigma\xi_{n3t} + \varepsilon_{n3t} \tag{3.5}$$

In these equations, $\boldsymbol{X_{njt}}$ is an $[i \times 1]$ vector of i alternate specific attributes for individual n and alternative j $(j \epsilon J \; where \; J \equiv \{1,2,3\})$ in choice task t. $\boldsymbol{\beta_n}$ is an $[i \times 1]$ vector of parameter weights (fixed and random) for that individual. The idiosyncratic term ε_{njt} follows an EV-I (Gumbel) distribution. The terms ξ_{n1}, ξ_{n2t} and ξ_{n3t} represent the individual specific error components and are drawn from a standard normal distribution. The notations $\boldsymbol{X_{n1}}$ and ξ_{n1} for the SQ alternative do not have the subscript t because they remain invariant for individual n. The parameter σ ensures homoscedasticity among the error components across the three alternatives. The term ξ mainly focusses on capturing the correlation (in the unobserved error term of the utility) across the multiple choice tasks which arise due to the invariant SQ alternative. It does not accommodate the panel impact where some unobserved individual specific effects remain the same across the multiple choice tasks for an individual.

The distribution of the random parameter vector is specified by the analyst. Some commonly used distributions are the normal, lognormal, uniform and triangular distributions. Each distribution is applied in a specific case and has its own merits and demerits. For example, a normal distribution can produce both negative and positive parameter estimates for travel time due to its symmetrical shape. On the other hand, a lognormal distribution considers the same parametric sign, but yields a much higher mean and variance because of its long tail, which makes it unsuitable for WTP calculations (Hensher, 2001a). Similarly, Hess et al., (2005) have found that the lognormal distribution gives the best model fit, but the worst mean and standard deviation during the WTP estimation. Conversely, the normal distribution has the poorest model fit, but the best estimate for the WTP. Moreover, Hensher & Greene (2003) found that VTTS estimates using the normal and triangular distributions are generally similar in magnitude. Since the aim of the study was to test the impact of the number of stop-&-gos (S&Gs) on route choice and evaluating the WTP between the travel time and the number of S&Gs, we decided to assign a normal distribution to the random parameters.

Hess and Rose (2007) provide the likelihood function of the RPECL model. Bhat & Castelar (2002) and Hensher (2008) have also used a similar likelihood function, but in a different context. Equation 3.6 gives the log-likelihood function $LL\ (\boldsymbol{W})$ for this model.

$$LL\ (\boldsymbol{W}) = \sum_{n=1}^{N} \ln\left[\int_{\boldsymbol{\beta_k}}\left(\prod_{t=1}^{T}\left(\int_{\xi} P_{nt}\ (j_{nt}|\ \boldsymbol{X_{njt}}, \boldsymbol{\beta_k}\ , \boldsymbol{\beta_f}, \sigma)\ h(\boldsymbol{\xi})\ d\boldsymbol{\xi}\right)\right)\ g(\boldsymbol{\beta_k})\ d\boldsymbol{\beta_k}\right] \tag{3.6}$$

$$P_{nt}\ (j_{nt}|\ \boldsymbol{X_{njt}}, \boldsymbol{\beta_k}, \boldsymbol{\beta_f}, \sigma) = \frac{\exp\ (V_{njt})}{\sum_{k=1}^{J} \exp\ (V_{nkt})} \tag{3.7}$$

In this equation, $\boldsymbol{\beta_k}$ and $\boldsymbol{\beta_f}$ are the random and fixed parameter vectors respectively from the parent coefficient vector $\boldsymbol{\beta_n}$. The expression in equation 3.7, $P_{nt}\ (j_{nt}|\ \boldsymbol{X_{njt}}, \boldsymbol{\beta_k}, \boldsymbol{\beta_f}, \sigma)$, represents the probability of individual n to select alternative j, which is the chosen alternative, in choice task t. The expression corresponds to a logit kernel where the observed utility, V_{njt}, is expressed in terms of the observed attributes, fixed and random parameters (expressed as $\boldsymbol{\beta'_n X_{njt}}$ in the equations 3.3 – 3.5). $h(\boldsymbol{\xi})$ corresponds to the probability density function (pdf) of the standard normal distribution and $g(\boldsymbol{\beta_k})$ is the pdf of the mixing distribution (which is again a normal distribution in this study).

3.5.2 Model estimation

The estimated set of parameters in equation 3.6, $\boldsymbol{W}$, comprises the mean of $\boldsymbol{\beta_k}$, $\boldsymbol{\beta_f}$, elements of a variance-covariance matrix of the attributes that are randomly distributed ($\boldsymbol{\Omega}$), and elements of a variance-covariance matrix of the error components ($\boldsymbol{\sum}$). Only the diagonal elements of $\boldsymbol{\Omega}$ are estimated, the square root of which represent the standard deviation of the random parameters $\boldsymbol{\beta_k}$. In other words, all the off-diagonal elements in $\boldsymbol{\Omega}$ are constrained to zero. The error component matrix $\boldsymbol{\sum}$ is a 1 X 1 matrix which corresponds to a univariate dependent variable, the route choice. $\boldsymbol{\sum}$ is normalised to 1 for identification and is multiplied to the error variance parameter σ. Equation 3.6 is generally estimated using the Maximum Simulated Log-likelihood (MSL) technique discussed in Bhat (2001); Bhat & Gossen (2004); Revelt & Train (2000) and Train (2009). The MSL form for equation 3.6 is given in equation 3.8, where R represents the number of simulated draws for the random and error component parameters used in the estimation.

$$MSL\ (\boldsymbol{W}) = \sum_{n=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \left(P_{ntr}\ (j_{nt}|\ \boldsymbol{X_{njt}}, \boldsymbol{\beta_{kr}}, \boldsymbol{\beta_f}, \sigma) \right) \right] \tag{3.8}$$

The estimation procedure of the RPECL model was coded in Matlab which can be accessed from the URL given in appendix B of the thesis. The simulated likelihood function given in equation 3.8 was constructed using 1,000 standard Halton draws (R) for each of the random parameters and the error components (Train, 2009). We used the non-linear unconstrained numerical optimisation scheme proposed by Broyden-Fletcher-Goldfarb-Shanno (BFGS) to maximise equation 3.8. The method is particularly useful in situations where the analytical form of the Hessian matrix is unavailable or is too expensive to compute at every iteration. The numerical Hessian matrix obtained upon convergence ($\mathbf{H}$) was used to evaluate the vector of standard errors of the parameters (Std_{error}) using equation 3.9 where diag is the diagonal operator.

$$Std_{error} = \sqrt[2]{\text{diag}\,(\mathbf{H}^{-1})} \tag{3.9}$$

3.5.3 Deriving the WTP measures between the attributes

Once the parameter set ($\boldsymbol{W}$) was obtained upon solving the likelihood function until convergence, a variety of Willingness To Pay (WTP) measures were evaluated using these parameters. The willingness to pay (WTP) is defined as the price which makes a consumer trade-off between buying and not buying a product (Schlereth et al., 2012). Equation 3.10 expands the observed part ($\boldsymbol{\beta_n' X_{njt}}$) of the utility specification for one of the alternatives. The notations used in the equations 3.3. – 3.5 still apply, but have been suppressed for brevity.

$$\boldsymbol{\beta' X} = \beta_{TT}.X_{TT} + \beta_{TTS}.X_{TTS} + \beta_{SnGo}.X_{SnGo} + \beta_{VRC}.X_{VRC} \tag{3.10}$$

$$\boldsymbol{\beta' X} = \beta_{TT}.X_{Other} + (\beta_{TTS} + \beta_{TT}).X_{TTS} + \beta_{SnGo}.X_{SnGo} + \beta_{VRC}.X_{VRC} \tag{3.11}$$

The four attributes used here are: the total travel time (X_{TT}), the time spent in stop-&-go(X_{TTS}), the number of stop-&-gos (X_{SnGo}), the vehicle running cost (X_{VRC}). The attribute X_{TT} represents the total travel time for an alternative, which is inclusive of the time spent in stop-&-go (S&G) traffic, X_{TTS}. Equation 3.11 represents an adjusted utility specification of equation 3.10 to get a revised estimate for X_{TTS}. X_{Other} denotes the time spent in other traffic conditions in this equation. The WTP estimates as obtained from equation 3.11 are given below.

Running cost – Travel time (AU \$/hr) $= (\beta_{TT}/\beta_{VRC}) \times 60$ (3.12)

Running cost – Time in S&G (AU \$/hr) $= \{(\beta_{TT} + \beta_{TTS})/\beta_{VRC}\} \times 60$ (3.13)

Running cost – No. of S&Gs (AU \$/stop) $= \beta_{SnGo}/\beta_{VRC}$ (3.14)

Travel time – No. of stop-&-go (min/stop) $= \beta_{SnGo}/\beta_{TT}$ (3.15)

Time in S&G – No. of S&Gs (min/stop) $= \beta_{SnGo}/(\beta_{TT} + \beta_{TTS})$ (3.16)

Equations 3.12 and 3.13 represent the Value of Travel Time Savings (VTTS) under overall and S&G traffic conditions respectively. Equation 3.14 denotes the WTP in dollars associated with the occurrence of an S&G. Equations 3.15 and 3.16 represent the trade-off between the number of S&Gs against the time spent under overall and S&G conditions. As discussed earlier in subsection 3.2.7.5, the travel time was inclusive of the time spent in S&G in the experiment. Therefore, equations 3.13 and 3.16 have the coefficient for the time spent in S&G as $\beta_{TT} + \beta_{TTS}$. Equation 3.15, which represents the willingness to pay in minutes of an individual to reduce the occurrence of S&G by one, is one of the novel contributions that this study aims to make.

3.5.4 Model results

Two formulations of the mixed logit model were tested and compared for the effective dataset to determine the trade-off between the four attributes used in the SC experiment. Model 1 represents the Random Parameter Logit (RPL) model to capture preference heterogeneity towards the four attributes across the participants. Model 2 is the Random Parameter Error Component Logit (RPECL) model to account for the correlation (in the unobserved component of the utility) across the multiple SC tasks along with the preference heterogeneity. All the parameters for the observed attributes except the vehicle running cost were set as random. The random parameters were assigned a normal distribution for both the models because of the following reasons: 1) the merits of using a normal distribution which have been discussed earlier in subsection 3.5.1, 2) the model with a triangular distribution did not give better goodness of fit. The lognormal distribution was not tested due to the reasons discussed earlier in subsection 3.5.1. The results for the triangular distribution are provided in appendix D (table D.1) of this thesis. The estimation routines for models 1 and 2 were coded in Matlab. Alternatively, model 1 was also estimated in STATA using the mixlogit package for verification (Hole, 2007; StataCorp., 2013).

Table 3.8: Estimation results from the selected model specifications

Attribute	Model 1 (RPL)	Model 2 (RPECL)
Mean of random parameters		
Travel time	-0.3011 ***	-0.3657 ***
Time spent in stop-&-go	-0.1270 ***	-0.1581 ***
Number of stop-&-go	-0.0821 ***	-0.0973 ***
Standard deviation of random parameters		
Travel time	0.1897 ***	0.2493 ***
Time spent in stop-&-go	0.1064 ***	0.1136 ***
Number of stop-&-go	0.0571 ***	0.0736 ***
Non-random parameters		
Running cost	-1.1854 ***	-1.3994 ***
Sigma (σ)	--	1.0982 ***
Log-likelihood at convergence	-1120.37	-1070.43
Adjusted Rho-squared	0.1797	0.2141
LR test versus model 1 (H_0: Model 1 is true)	--	99.88 ***

*** significant at 99%

Table 3.8 shows the estimation results for both the models. Both the models report highly significant and negative parameter estimates which represent an increase in the disutility of a route as the level of the attribute under consideration increases. A negative sign on the number of stop-&-gos (S&Gs) implies that the attribute also contributes towards disutility of a route, thus indicating the validity of the proposed hypothesis. A pairwise correlation coefficient between the parameters of travel time and number of S&Gs was observed to be 0.29, and that of the time spent in S&G and the number of S&Gs as 0.20. These low correlation coefficients indicated that the estimated parameters were nearly independent of one another.

Table 3.8 also shows a significant standard deviation for the number of S&Gs experienced which signifies a different perception across individuals towards this attribute. The parameter σ in model 2 is significant and indicates a correlation in the unobserved errors across the multiple choice tasks due to the presence of the status quo alternative. These error components are normally distributed ($\xi \sim N[0, 1.0982]$), which is not accounted for in model 1 ($\sigma = 0$). Model 2 has a superior final log-likelihood value than model 1 which indicates a better fitting model for the given dataset. A Likelihood Ratio (LR) test was also conducted to check the null hypothesis that model 2 was statistically no different to model 1 ($H_0: \sigma = 0$). The calculated LR value (99.88) was higher when compared to the critical chi-squared value (3.84) for 1 degree of freedom at 5 percent significance. Thus, the null hypothesis was rejected and it could be concluded that the RPECL (model 2) outperformed the RPL (model 1) model for the dataset considered in this study.

The RPECL model also performed better than a few alternate model specifications which were tested in this analysis. A comparison of the overall goodness of fit of the different models is provided in appendix D (tables D.2 – D.4) of this thesis. Furthermore, the impact of socio-demographic variables was not included in model 2 due to three reasons. Firstly, due to the panel nature of the dataset, these variables were introduced after interacting (multiplying) them with the given route specific attributes. Several combinations of these interaction variables were tested, but found to be statistically insignificant. Secondly, as we wanted to compare the findings from our model with the one used by (Hensher, 2001a), which also did not include socio-demographic information, we decided not to include them for a consistent comparison. Thirdly, the alternate model specification (Table D.2 in appendix D), the 2 segment latent class model, also showed a negative and significant effect of the number of S&Gs on the disutility of a route. The model also showed that being a female makes one more likely to have a high WTP for travel time reduction. However, this observation was found to be inconsistent with the previous literature which found females to have a lower value of time (Srinivasan 2005). Thus, we concluded that the current sample size could not provide interesting socio-demographic characteristics, and did not include them in the final results. A detailed discussion on the latent class model will be presented in chapter 4 of this thesis.

Table 3.9 summarises the mean of the WTP estimates obtained from model 2 which were calculated using equations 3.14 – 3.18. The WTP measures corresponding to equations 3.17 and 3.18 were simulated through the Krinsky and Robb parametric bootstrapping technique, using 1,000 draws, because both the parameters in each equation were normally distributed (Krinsky & Robb, 1990). The other WTP measures given in equations 3.14, 3.15 and 3.16 were evaluated as point estimates of the mean.

Table 3.9: Mean WTP estimates from the RPECL model

Willingness To Pay measure	Estimated value
Running cost – Travel time (AU $/hr)	15.68
Running cost – Time in stop-&-go (AU $/hr)	22.45
Running cost – No. of stop-&-go (AU $/stop)	0.070
Travel time – No. of stop-&-go (min/stop)	0.267
Time in stop-&-go – No. of stop-&-go (min/stop)	0.186

The ratio of the Value of Travel Time Savings (VTTS) in S&G to other driving conditions was observed to be 1.4. If the ratio is greater than one, it would validate the fact that drivers find travelling in S&G conditions more onerous due to an increased focus and discomfort due to these conditions (Levinson et al., 2004). This ratio was calculated as 1.3 by Hensher (2001a) and 1.2 by Rose et al. (2009) (ratio of VTTS in S&G to the summation of VTTS in free flow and slowed conditions). Additionally, the ratio between travel time and time spent in S&G was found to be 1:2.37 in this study, which was again close to the one (1:2.5) reported in Hensher (2001a). Thus, the results from this study were found to be consistent with the previous literature. Furthermore, both the VTTS estimates (under S&G and other driving conditions) were less than the minimum hourly wage rate of the effective sample (AU $25.50) which indicated the obtained estimates to be meaningful. Table 3.9 also reports the trade-off estimate between the overall travel time and the number of S&Gs. The estimated value implies that drivers do not mind travelling for an additional 16 seconds (0.267*60), on average, to reduce the occurrence of stop-&-go (S&G) by one on their travelled route before shifting to another route. This value reduces to 11 seconds while driving in S&G conditions.

3.6 Discussion

This study tested the hypothesis of an increase in the disutility on a travelled route with an increase in the number of S&Gs experienced by a driver. The collected Stated Choice (SC) data on the university car commuters (drivers) was analysed using the Random Parameter Error Component Logit (RPECL) model which explains both the correlation among the choice tasks and the preference heterogeneity. Results from the SC experiment data showed an increase in disutility as the number of S&Gs increased on a route. In other words, drivers experience elevated discomfort levels when undergoing alternating cycles of S&G traffic. Thus, the outcome of this proof of concept study looks promising as our intuition (the research hypothesis) was proved right. This finding adds to the existing literature which, until now, focussed on expressing driver discomfort and the resulting disutility as a function of the time spent in S&G traffic. Results also showed that drivers, on average, were willing to spend up to 16 seconds extra on travel time to reduce an additional occurrence of S&G on their current route.

This estimate of discomfort can have interesting policy implications. One of the policy implications could be a need to re-assess some existing policies like the toll pricing. Another implication of the obtained results would be towards the development of transportation models that better represent the route choice of drivers and how it affects the traffic congestion pattern in a road network. We discuss these policy implications later in chapter 6 of the thesis.

This study brought out some interesting findings that added to the existing literature on driver route choice behaviour. However, the study made a few considerations (assumptions) which limited the scope of this work. Firstly, the study considered the number of S&Gs that was revealed by the respondents during the SC experiment as the true measure. Generally, it is challenging to recollect the instances of S&Gs when compared to other quantities like travel time and cost. This might bring in a measurement bias in the observed attribute (number of S&Gs) which might not have been greatly reduced by showing video animations in the survey to the participants. Secondly, the study was conducted on a sample of university staff and students which do not represent the demographics and travel characteristics of the general population in Sydney. That is, the WTP measures obtained from this study might be skewed which limits its applicability to the regional transportation models of Sydney. Moreover, the collected sample was relatively small to derive reliable and meaningful interpretations using the advanced and parameter intensive models, like the hybrid choice models (refer to appendix D for more information). Thirdly, the effective sample in this study comprised individuals whose commute frequency to work ranged between 1 and more than 5 days per week. We initially wanted to restrict the minimum commute frequency to at least 3 days a week. The reason behind it was that these individuals were expected to show a better association towards S&G traffic than the occasional car drivers. As a result, they could better recollect the time spent in S&G and number of S&Gs experienced with minimum measurement bias in their response. However, due to a low response rate, we did not receive sufficient responses in the desirable range. Moreover, removing the participants with commute frequency less than 3 days considerably reduced our effective sample size. Thus, we analysed the data for all the individuals who drove to work by car. Fourthly, the RPECL model used in this study did not yield significant effects for the socio-demographic variables, which were included as interaction variables. Including them can further help in disentangling the complex route choice behaviour of drivers in terms of their socio-economic status. We address a majority of these shortcomings in the extension of this study, which is discussed in the next chapter.

CHAPTER 4

EXPERIMENT II – EXPANDED STUDY

The motivation behind this chapter came from the positive results that were obtained from the proof of concept study. In this chapter, we further enrich the analysis by relaxing a few assumptions that were made during experiment I, which we discussed towards the end of chapter 3. A similar online survey was conducted on a sample of car commuters residing in Sydney or its neighbouring regions who regularly drove to work. The collected data was analysed using a hybrid choice modelling framework, called the Latent Class Choice Model (LCCM), which segregates individuals into subgroups where each subgroup has a unique set of characteristics and preferences. The results obtained from this study would provide a better mapping to the larger (regional) population in Sydney. It would also offer a richer set of information to decision-makers in proposing new and innovative policies aimed at reducing traffic congestion and the resulting stop-&-go (S&G) waves.

This chapter mainly covers: 1) design of the SC questionnaire, 2) survey research, 3) descriptive statistics of the collected data, and 4) quantitative analysis of the data. We do not repeat the discussion on the background to SC experiments and encourage the readers to refer back to section 3.1 for a detailed review. The first section presents the inputs, or the design specifications, a slightly modified design procedure and the resulting output from the Stated Choice (SC) design exercise. The second section discusses the survey administration and data collection using the similar online survey instrument that was used in the previous experiment. The empirical analysis of the collected data constitutes the third section of the chapter. The fourth section presents the econometric framework and the estimation routine of the LCCM which is followed by the results from the analysis. Finally, the chapter concludes with a discussion of the results and limitations of the study.

4.1 Methodology Adopted for the SC Design

We used the same approach, i.e. the D-efficient homogeneous pivot design SC technique, which was adopted in experiment I – a proof of concept study. For the current experiment, the set of inputs parameters (design specifications) to the SC design was revised to accommodate the characteristics of the new survey sample, which comprised the people residing in Sydney or its neighbouring regions who also drove regularly to work. This section primarily elaborates on the modifications that were made to the SC design process, previously discussed in section 3.2, to accommodate the new design requirements. More details on the unchanged components can be found in section 3.2 of this thesis.

4.1.1 Design specifications

This subsection describes the set of inputs for the SC survey design and how these were acquired. While some of the inputs remained unchanged from the previous experiment, few design specifications required additional exploration.

4.1.1.1 Number of alternatives

We retained the same number of alternatives (three) from the previous study. The presented alternatives included the status-quo along with two unlabelled hypothetical routes which were pivoted around the status-quo alternative. We discussed the reasons behind selecting three alternatives in subsection 3.2.1.1.

4.1.1.2 Attributes and their levels

We maintained the same number of attributes (four) and the attribute levels (five) that were employed previously. A detailed discussion on the attributes along with their attribute levels can be found in subsection 3.2.1.2.

4.1.1.3 Number of choice tasks and blocks

It was decided to present 10 choice tasks per participant in the current study, which was still manageable with regard to the cognitive load on participants. We followed the blocked design method where two distinct blocks of 10 choice tasks each were formed. More details on the number of choice tasks and blocks is presented in subsection 3.2.1.3.

4.1.1.4 Prior attribute information

The results from the previous chapter, obtained upon estimating the RPECL model on the effective dataset, were used as priors to the new full survey design. Table 4.1 lists the mean, standard deviation of the prior estimates along with their assumed distributions used for the design exercise.

Table 4.1: Prior parameters and distributions used in the expanded study

Attribute	Mean	Std. dev.	Distribution for full design
Total travel time (minutes)	-0.3657 ***	0.2493 ***	[n,(n,-0.3657,0.0238),(n,0.2493,0.0207)]
Time in S&G (minutes)	-0.1581 ***	0.1136 ***	[n,(n,-0.1581,0.0149),(n,0.1136,0.0134)]
No. of S&Gs	-0.0973 ***	0.0736 ***	[n,(n,-0.0973,0.0110),(n,0.0736,0.0107)]
Vehicle running cost (AU $)	-1.3994 ***	--	[(n,-1.3994,0.1178)]

*** significant at 99%; (n,μ,σ) where n is normal distribution, μ and σ are the mean and standard deviation respectively

Similar to the priors reported in table 3.2, a normal distribution was assumed for the parameters total travel time, time in S&G and number of S&Gs. Vehicle running cost was treated as a non-random parameter in the design. Further, a Bayesian normal distribution was assumed for the mean and standard deviation of all the prior parameters, which is shown in the rightmost column of table 4.1. For example, the prior distribution for the number of S&Gs is represented by the expression $[n,(n,-0.0973{,}0.0110),(n,0.0736{,}0.0107)]$. The expression implies that the parameter follows a normal distribution (the first instance of n in the expression) with some mean and standard deviation. The mean of this distribution is in turn derived from another normal distribution (the second instance of n in the expression) having a mean of -0.0973 and a standard deviation of 0.0110. Similarly, the standard deviation of the prior distribution is also normally distributed (the third instance of n in the expression) with a mean of 0.0736 and standard deviation of 0.0107.

4.1.1.5 Population segments and their reference alternatives

The target audience for the study consisted of the individuals residing in Sydney or its neighbouring regions who regularly drove to work by car. Thus, to come up with a homogeneous pivot design, it was necessary to first classify the survey population into well-defined segments. We found some key statistics on the car driving (not passengers) population in Sydney, which constituted 63.5 percent of the total average commute trips, from the Household Travel Survey (HTS) report of Sydney for the waves 2012-13 and 2014-15 (BTS, 2014; 2015). The statistics of our interest like the travel time and travel distance distribution of vehicle (car) drivers were extracted from the HTS database and presented in tables 4.2 and 4.3 respectively.

Table 4.2: Travel time distribution of car drivers in Sydney. Source: (BTS, 2015)

Travel time range (in minutes)	**Trips (in ,000)**	**Representation (%)**
0 – 10	3657	42.378
10 – 20	2348	27.214
20 – 30	1242	14.387
30 – 40	452	5.239
40 – 60	644	7.464
Above 60	286	3.319
Total	8630	100

Table 4.3: Travel distance distribution of car drivers in Sydney. Source: (BTS, 2015)

Travel distance range (in km)	**Trips (in ,000)**	**Representation (%)**
0 – 2	1953	22.360
2 – 5	2468	28.597
5 – 10	1813	21.008
10 – 20	1385	16.048
Above 20	1011	11.714
Total	8630	100

The tables show that nearly half the car driving population has a travel time and travel distance of more than 10 minutes and 5 kilometres respectively, making them more susceptible to experience congestion and S&G traffic. We used these inputs along with travel time and distance information from a route choice study conducted by

Shakeel et al. (2016) to generate a two way classification table between travel distance and time. An Iterative Proportional Fitting (IPF) algorithm was applied to render the desired classification table for the statistics reported in the above two tables. The IPF is a widely used technique to synthesise population level information using aggregate and some sampled information (Norman, 1999). The resulting two way cross classification table thus obtained is shown in table 4.4. The table shows that around 60 percent of the population covers more than 5 kilometres with a travel time greater than 10 minutes. Table 4.4 was further used to calculate the weighted average distance for each travel time segment (by multiplying the percentage with the mean distance of each segment) which is presented in table 4.5.

We classified the population into 6 segments based on travel time. The reference alternative for each population segment was synthesised using the inputs presented in tables 4.4 and 4.5. Table 4.6 shows the reference alternatives that were used as an input in the SC design exercise. The set of rules that were followed to derive the reference attributes are listed below. These are the same rules that were used in chapter 3.

- The travel time attribute was set as the mean of a segment’s travel time range
- Time spent in stop-&-go (S&G) was kept between 20 to 25 percent of the travel time attribute (as found by Hensher, (2001b))
- Number of S&Gs experienced was set between 30 to 40 percent of the upper limit of the travel time range in a segment
- Vehicle running cost was calculated as the product of cost per kilometre (AU $0.15) and the average distance (from table 4.5) in each segment (RACQ, 2014).

These segment specific reference alternatives were used to generate a D-efficient homogeneous pivot design which yielded a single SC experiment for the entire survey sample, i.e. the choice tasks remained the same across the participants.

Table 4.4: Two way classification table between travel time and distance. Cell values in ,000 trips

		Travel time (minutes)							
		0-10	**11-20**	**21-30**	**31-40**	**41-60**	**60+**	**Total**	**BTS Total**
	0-2	1881.479	0	0	0	0	0	1881.479	1953
	2-5	1752.991	624.628	0	0	0	0	2337.619	2468
Trip	**5-10**	22.679	1723.927	0	0	0	0	1746.606	1813
distance	**10-20**	0	0	1057.966	293.546	165.454	0	1516.966	1385
(km)	**20 +**	0	0	183.639	158.520	478.652	286.519	1107.330	1011
	Total	3657.149	2348.555	1241.605	452.066	644.106	286.519	**8630**	
	BTS Total	3657	2348	1242	452	644	286		**8630**

Table 4.5: Weighted average of travel distance for each travel time segment

	Travel time (minutes)					
	0-10	11-20	21-30	31-40	41-60	60+
Wt. avg. distance (km)	2.24	6.44	17.22	20.26	26.15	30.00

Table 4.6: Reference alternatives selected for the SC design on Sydney sample

Segment	Travel time range (minutes)	Weightage (%)	Reference alternative taken for the design			
			Travel time (minutes)	Time in S&G (minutes)	Number of S&Gs	Running cost (AU $)
1	0-10	42.378	5	2	4	0.35
2	10-20	27.214	15	5	7	1.00
3	21-30	14.387	25	8	10	2.60
4	31-40	5.239	35	10	14	3.05
5	41-60	7.464	50	15	20	3.95
6	> 60	3.319	75	20	30	4.50

4.1.1.6 Model specification for the design

We used the same Random Parameter Error Component Logit (RPECL) model specification to generate the SC tasks in the current experiment. The reasons behind selecting the model have already been discussed in the subsection 3.2.1.6.

4.1.2 Design of the SC experiment

With all the known design specifications, the next step was to generate the D-efficient homogeneous pivot SC design. We followed a slightly modified approach to the SC design when compared to the one previously discussed in subsection 3.2.2. The key difference was the way we specified segment weights in the design. We allowed the population segment proportions to be flexible during this design exercise. This was unlike the previous design style (discussed in subsection 3.2.2) which assumed these weights to be fixed. Thus, the modification offered more robustness to the SC design as it was difficult to know the true weight of a population segment a priori.

Two separate design procedures were written for the model specifications MNL and RPECL respectively. Let's name the scripts as code_full_MNL and code_full_RPECL, which can be accessed through the weblink provided in appendix B of this thesis. The parameters for the MNL model were estimated using the effective dataset of 145 participants from the previous experiment. The estimates have been reported in table 4.7. These estimates were used as prior values in the design script code_full_MNL. Additionally, the segment weights were specified in ranges for the reasons discussed above. Table 4.8 shows the original segment weights that are reported in table 4.6 along with their corresponding ranges that were used in code_full_MNL. As discussed earlier in subsection 3.2.2, we executed the code_full_MNL script first to obtain the set of SC tasks, which we refer to as design_full_MNL. The design also provided the optimum proportion weights that further minimised the overall efficiency of the design. The optimum weights, which are shown in the third column of table 4.8, were used as new weight inputs in the design script code_full_RPECL. Moreover, code_full_RPECL used the prior parameter information given in table 4.1.

Table 4.7: Prior parameter values from the MNL model

Parameter	Estimated value
Total travel time (minutes)	-0.1750 **
Time in S&G (minutes)	-0.0679 **
No. of S&Gs	-0.0256 **
Vehicle running cost (AU $)	-0.7790 **

** significant at 95%

Table 4.8: Segment weights used in the design procedure

S. No.	Travel time segment	Original weight	Range specified in code_full_MNL	Optimum weight for code_full_RPECL
1	0-10	0.424	0.30 - 0.45	0.384
2	10-20	0.272	0.20 – 0.35	0.284
3	21-30	0.144	0.10 – 0.20	0.157
4	31-40	0.052	0.02 – 0.10	0.065
5	41-60	0.075	0.05 – 0.10	0.078
6	> 60	0.033	0.01 – 0.05	0.032

The steps for evaluating the efficiency of design_full_MNL using code_full_RPECL remain the same. Readers can find a detailed explanation of these steps in subsection 3.2.2 of this thesis. The entire design procedure was carried out using the Ngene software package (ChoiceMetrics, 2012).

4.1.3 The resulting SC tasks

Two separate blocks of 10 choice tasks each were generated from the design exercise. Table 4.9 shows the choice tasks within each block where the field BlockID is the identifier for blocks. Each choice task in the table corresponds to proportions which are multiplied with the attributes of the status-quo alternative to obtain the attributes for the two hypothetical alternatives. The D-error for the two blocks, upon stabilising, was found to be 0.745 and 0.746 respectively. More details on the adopted blocked design approach and stability of the D-error have been discussed in subsection 3.2.3.

4.2 Survey Research

A similar methodology was adopted for the survey administration and data collection as discussed previously in section 3.3 of this thesis. In this section, we briefly discuss the survey administration method, layout of the survey instrument and data collection.

4.2.1 Survey administration method

We decided to use the same online survey instrument which was used to survey the UNSW sample in the previous chapter. The reasons behind selecting an online method have been previously discussed in subsection 3.3.1. For this study, we contracted a data collection and management organisation, Qualtrics, to administer the survey (Qualtrics, 2016). They circulated the online survey within their own participant pool in Sydney, which has a decent representation of individuals with diverse demographic and socio-economic characteristics. Additionally, it was also possible to control the composition of the survey sample by applying a certain set of quotas during data collection. This level of control during data collection was not witnessed during experiment I, where extra filtering criteria had to be dropped due to a low response rate to the survey. Hence, we could find a sample of the participants in this study which well represented the socio-demographics of Sydney's population.

4.2.2 Layout of the online survey

The online survey instrument mainly comprised four sections: 1) introduction to the survey, 2) socio-demographic, 3) revealed travel characteristics and 4) route choice tasks. For this experiment (survey), the socio-demographic section was put ahead of the route choice part due to the reasons discussed in the next subsection. Subsection 3.3.2 of this thesis elucidates each of these sections. The weblink to the survey page is available in appendix B of this thesis.

Table 4.9: Two blocks of homogeneous pivot design for the full survey

BlockID	TaskID	RI_TT	RI_TTS	RI_S&G	RI_VR	RI_TT	RI_TTS	RI_S&G	RI_VR
0	1	1.1	1.50	0.75	0.875	0.9	0.50	1.25	1.125
0	2	1.0	0.75	1.00	1.125	1.0	1.25	1.25	0.875
0	3	0.8	0.75	0.50	1.125	1.1	1.00	1.50	0.875
0	4	1.1	1.00	1.25	0.750	0.9	1.00	1.00	1.250
0	5	0.9	1.25	1.50	1.250	1.2	0.75	0.50	0.750
0	6	0.8	0.50	1.25	0.750	1.2	1.50	0.75	1.250
0	7	1.0	1.50	1.50	0.875	1.0	0.50	0.50	1.125
0	8	1.2	0.50	1.00	1.000	0.8	1.50	0.75	1.000
0	9	1.2	1.00	0.75	1.250	0.8	1.25	1.00	0.750
0	10	0.9	1.25	0.50	1.000	1.1	0.75	1.50	1.000
1	1	0.8	1.25	1.00	1.250	1.2	0.75	1.00	0.750
1	2	0.9	0.50	0.50	1.125	1.1	1.50	1.50	0.875
1	3	1.0	1.25	1.50	0.875	0.9	1.00	0.50	1.125
1	4	0.9	0.75	1.25	1.125	1.1	1.25	0.75	0.875
1	5	1.2	1.50	1.00	0.875	0.8	0.50	1.25	1.000
1	6	1.1	0.50	0.75	0.750	0.9	1.50	1.25	1.250
1	7	1.2	0.75	1.50	1.250	0.8	1.25	0.50	0.750
1	8	0.8	1.00	1.25	0.750	1.2	1.00	0.75	1.250
1	9	1.1	1.00	0.75	1.000	1.0	0.75	1.00	1.125
1	10	1.0	1.50	0.50	1.000	1.0	0.50	1.50	1.000

4.2.3 Data collection

Individuals residing in Sydney who regularly drove to work by car were selected as the target population for this study. The criteria used to define the population of interest were: 1) residing in Sydney or its neighbouring regions, 2) possessing a driver's license, and 3) driving at least thrice a week to work. A respondent not satisfying any of these criteria was dropped from the survey. The data collected thus represented a sample of car drivers who regularly drove by car, thus having a better perception towards stop-&-go (S&G) traffic than the people who drove occasionally. In order to maintain representativeness of the collected sample with regard to the target population, a set of quotas were also defined on the socio-demographic attributes like gender, age and income. The quotas represented the summary statistics of the target population and were obtained from the Sydney Household Travel Survey (HTS) report for the year 2012-13 (BTS, 2015). The statistics represented the characteristics of individuals who drove by private vehicle for their commute, which constituted 63.5 percent of the total weekday trips. The socio-demographic questions were asked in the starting section of the survey and participants were discontinued from the survey once any of the specified quotas got fulfilled. This ensured that we maintained a sufficient representation and richness of demographics, with regard to the target population, in the collected data.

The survey was circulated between Tuesday and Friday every week for around 5 weeks. We received a total of 288 responses from the survey. The survey response duration was found to be around 10 minutes, on average, across all the respondents, which was lower than what was observed in the previous experiment. This could be due to two reasons: 1) the surveyed participants regularly respond to other surveys (as they get paid each time), and thus can quickly respond to some sections, particularly the socio-demographic section, and 2) the survey did not have the section on the advertisement to the driving simulator study which further reduced the response time. All necessary clearances from the university's human ethics committee were taken before commencing the survey.

4.3 Empirical Analysis

The collected data was first cleaned and a few incomplete or invalid responses were excluded from further analysis. The responses that were dropped included: 1) *incomplete responses:* where the participants could not complete the survey due to reasons like losing interest or loss of internet connection while responding to the survey, 2) *left biased responses:* where the participants always selected the left most alternative, the status-quo, across the 10 choice tasks, and 3) *unusual response time:* where the participants took less than 5 minutes or more than 20 minutes to complete the survey. The minimum and maximum response time was found to be 2 and 829 minutes respectively, which was infeasible given the length of the survey and the level of ease of participants. The justification behind dropping the left biased responses has been discussed earlier in section 3.4 of this thesis. All these responses constituted roughly 13 percent of the total collected data which, not being a sizeable proportion, were discarded from further data analysis.

The effective dataset comprised 249 participant responses, which equated to 2490 stated choice (SC) observations in all. We discuss the descriptive statistics of the effective dataset in this section. A thorough empirical analysis will provide us useful insights on selecting an appropriate quantitative data analysis technique.

Figure 4.1 shows the descriptive statistics of the socio-demographic characteristics of the effective dataset. The survey received nearly equal responses from both females (47 percent) and males (53 percent). The age is evenly distributed among different age brackets except the youngest and the eldest age group, which seldom drive due to legal and physical constraints respectively. The annual income plot shows that nearly 60 percent of the sample has an income of AU $50,000 per annum which is well above the low-income group category as defined by ABS (2015). The pie chart on occupation shows that the sample primarily comprises participants with white-collar (office going) jobs who use a car to drive to work. Exactly three-quarters of the sample have a driving experience of 10 years and above which makes them more likely to be aware of S&G traffic. Thus, the each of these socio-demographic attributes shows decent representation of segments (most of them are more than 10 percent) in the collected data.

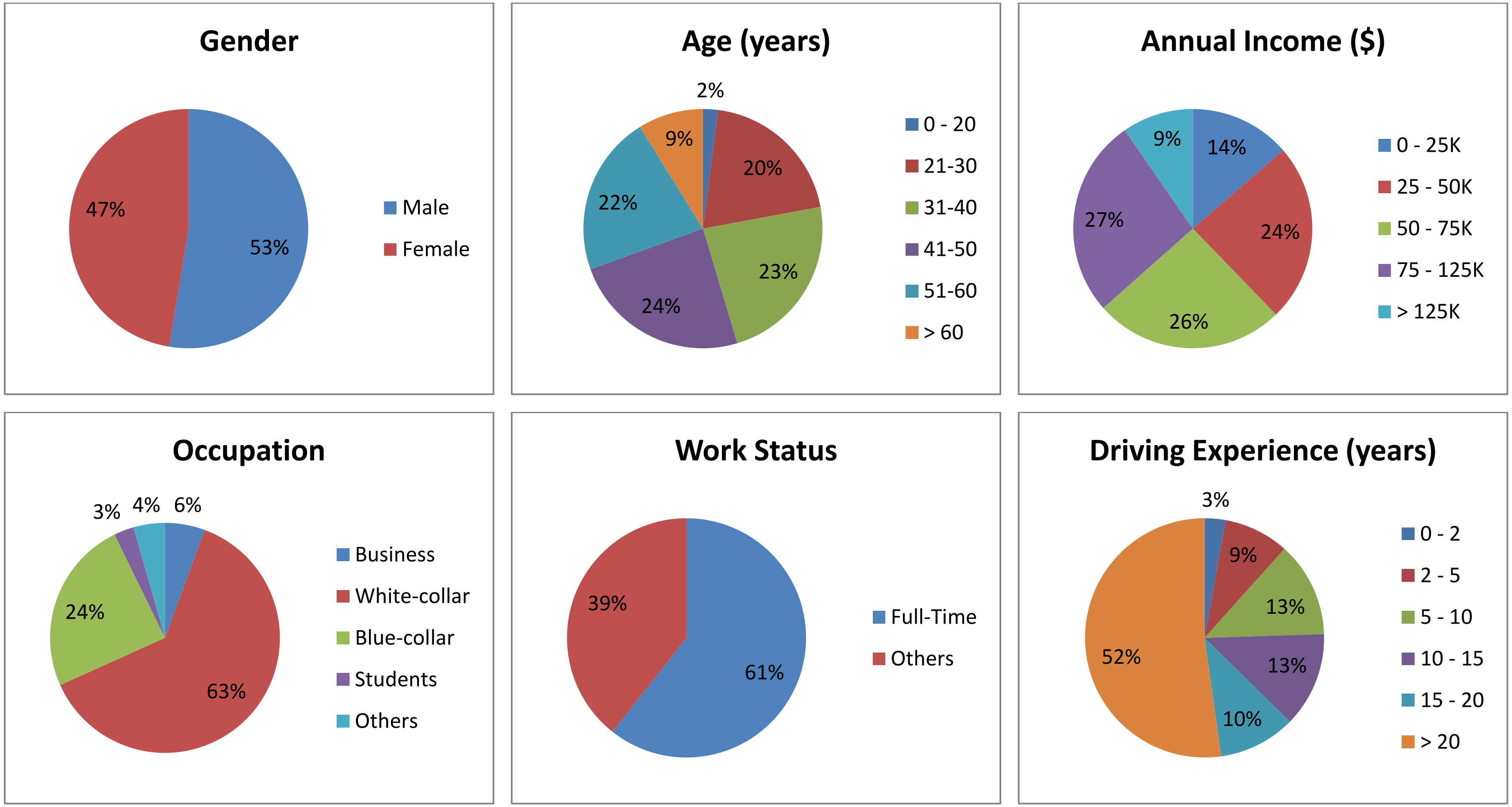

Figure 4.1: Summary statistics of socio-demographic attributes of the available data

Table 4.10 compares the sample statistics (from the survey) with the summary statistics of the target population obtained from the Sydney Household Travel Survey (HTS) report . The latter set of statistics was used as quotas to control the quality of collected data (discussed earlier in subsection 4.2.3). The survey sample maps well with the target population statistics, with up to 5 percent difference between the two proportions for most of the socio-demographic segments. A deviation of more than 10 percent (twice when compared to the other attributes) is observed in the elderly (60 years and above) and high-income (AU $125K and above) groups due to their low response rate. We tried to boost their participation, but were not successful despite sending several waves of reminder emails within the target population. Nevertheless, the collected dataset gives a good representation of the general car driving population in Sydney. Thus, the results obtained from the collected sample can also be mapped to a regional level.

Table 4.10: Comparison between the sample and population statistics

Category	Segment	Sample (%)	Population (%)
Gender	Male	52.61	53
	Female	47.39	47
Age	20 years and less	2.01	3.59
	21 to 30 years	20.08	12.98
	31 to 40 years	23.29	16.57
	41 to 50 years	24.10	19.34
	51 to 60 years	21.69	18.23
	60 years and above	8.84	29.28
Income (AU $)	25K and less	13.65	18.48
	25.1K to 50K	24.10	19.94
	50.1K to 75K	25.70	20.82
	75.1K to 125K	26.91	20.82
	125K and above	9.64	19.94

Figure 4.2 and table 4.11 show the distribution and summary statistics of the four revealed travel related characteristics respectively. The figure shows that the distributions are skewed towards the left for a majority of the sample. Table 4.11 reports a mean travel time of 36 minutes and a vehicle running cost of AU $3.57 for car commuters (which equates to roughly 25 kilometres of travelled distance), with 80 percent of the sample having a drive time and running cost up to 45 minutes and AU

$5.25 respectively. Nearly half of this time is spent driving in stop-&-go (S&G) traffic with around 12 occurrences of S&G experienced on average.

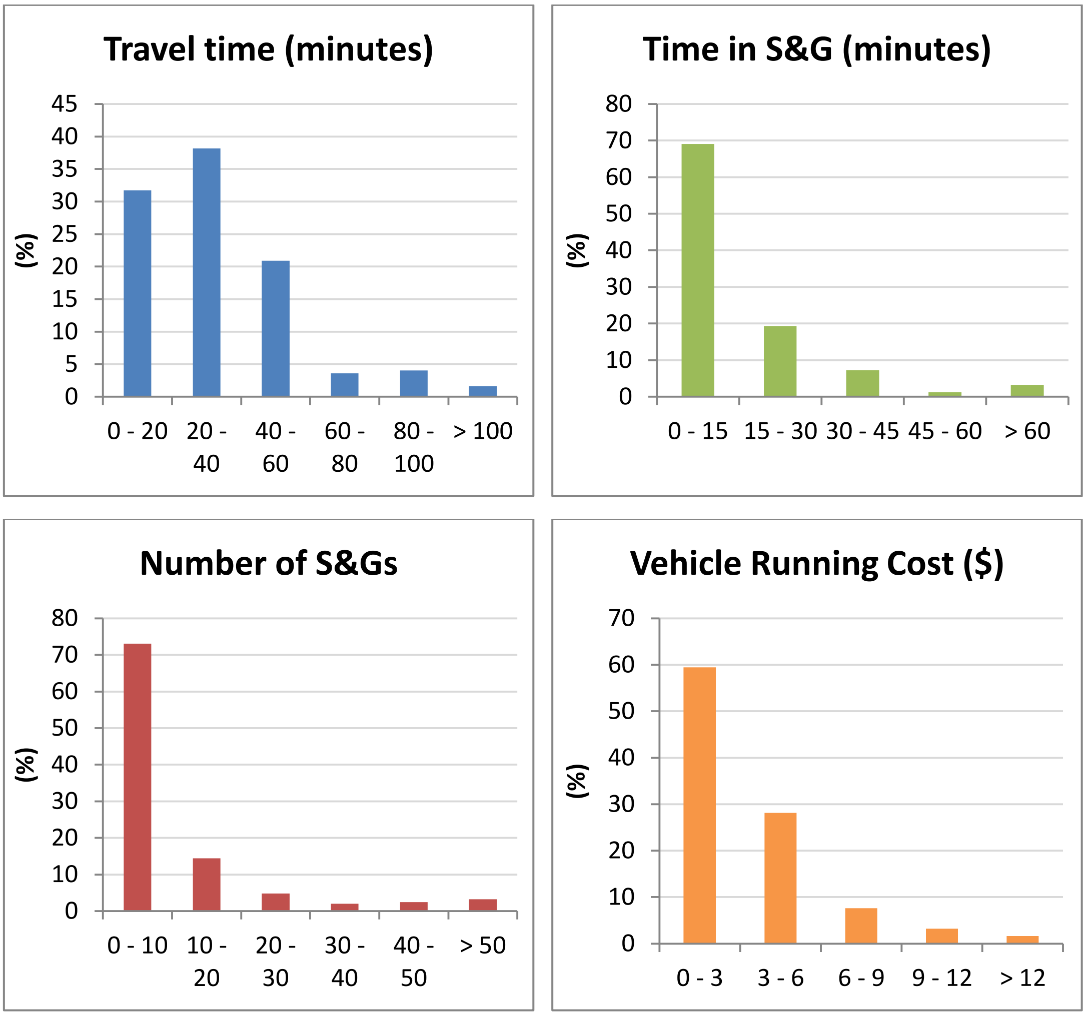

Figure 4.2: Distribution of travel specific attributes of the participants

Table 4.11: Summary of descriptive statistics of travel specific attributes

Data	Mean	Std. dev.	Min.	Max.	20th percentile	80th percentile
Travel time (minutes)	36.06	22.89	10	140	20	45
Time in S&G (minutes)	16.47	18.75	0	130	5	25
No. of S&Gs	12.06	17.25	0	100	3	15
Running cost (AU $)	3.57	2.86	0.15	17.25	1.5	5.25

In other words, the sampled drivers have to experience congested driving coupled with S&G conditions while travelling to work during morning peaks which can be used to test the relationship between route choice and the number of S&Gs experienced. The data for the time spent in S&G traffic and the number of S&Gs is found to have a greater standard deviation value than the mean. These measures are generally hard to perceive when compared to travel time and cost, thus causing few respondents to report an over-estimated value. Nevertheless, the mean of the estimate is still reasonable and can be used for further analysis.

We also closely analysed the Stated Choice (SC) responses of the participants in this experiment to possibly discern different behaviours (strategies) adopted by the participants while making choices. Figure 4.3 shows the choice tasks 3, 5, 6, 8, 9 and 10 belonging to the same block (block-0 shown in table 4.9) of SC tasks used in the study. Unlike the other choice tasks, these selected choice tasks demonstrate a clear preference towards an alternative. Within each choice task plot, the four attributes are represented by histograms where the height denotes their levels across the three alternatives. The rectangular boxes at the feet of the histograms give the percentage of respondents selecting the route in a given choice task. The three main types of respondent behaviour that can be identified through visual inspection of the plots are:

1. Individuals who are more inclined towards reducing their travel time and the number of S&Gs while making route selection. These people do not have a high disutility towards the running cost and are willing to pay a little extra to minimise the travel time and the number of S&Gs. Choice tasks 3, 8 and 10 represent this behaviour.
2. Drivers who are more likely to select the route with the lower running cost and the number of S&Gs. This kind of behaviour is only visible in choice task 5.
3. Individuals who primarily try to minimise their travel time and running cost. They generally do not consider the number of S&Gs while making their route choice. Choice tasks 6 and 9 depict this kind of behaviour.

Similar observations were also made for the other block (block-1 shown in table 4.9) of SC tasks. Identification of these patterns would give us a justification to classify the available dataset on the basis of behavioural interpretation. It would also be a handy input during data analysis, using the Latent Class Choice Model (LCCM), to classify and describe different segments of individuals depicting a particular route choice behaviour.

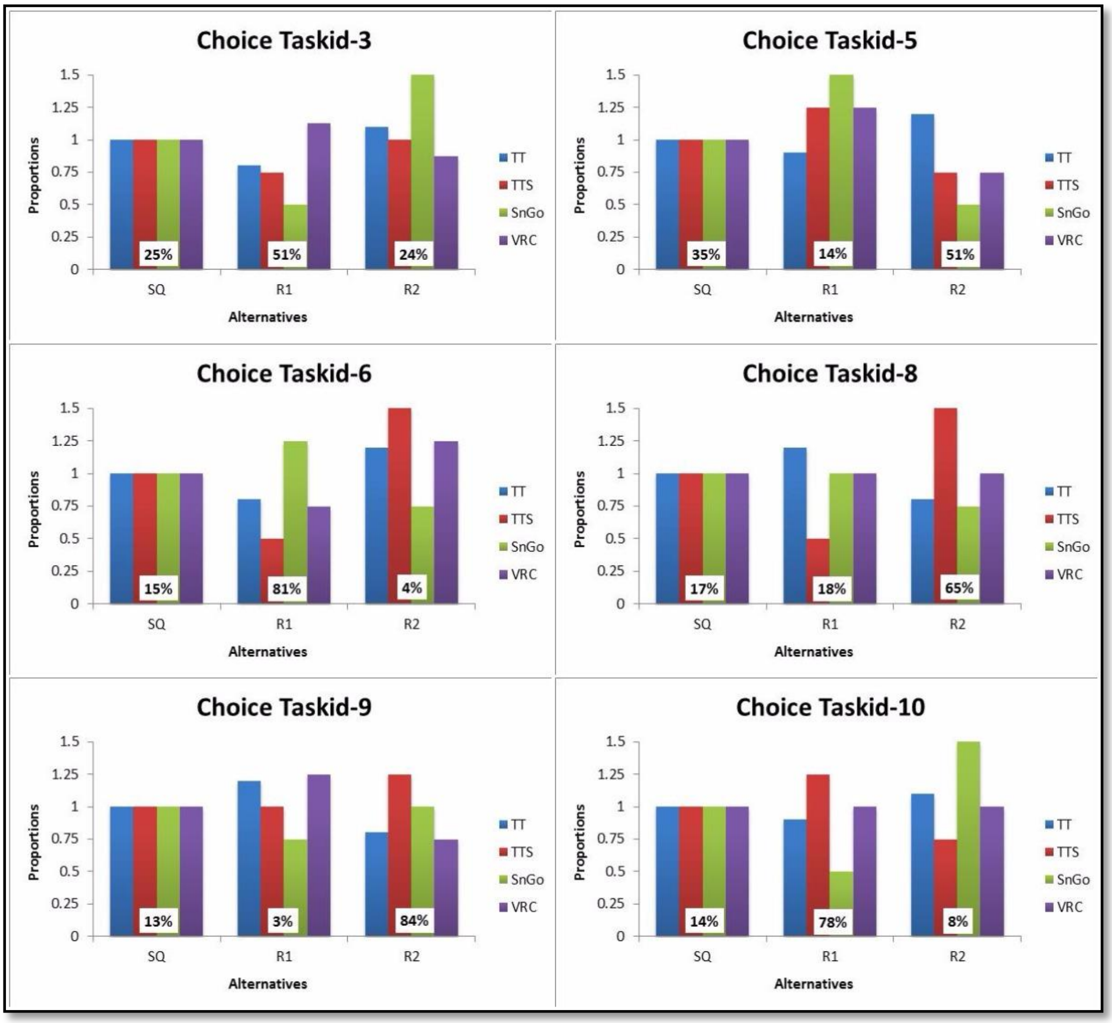

Figure 4.3: Selected SP choice tasks from block 0 presented to the respondents

(Percentage of route selection is given in the rectangular boxes)

4.4 Discrete Choice Analysis

We used discrete choice models again to quantitatively analyse the collected SC dataset and discern the underlying route choice behaviour and WTP measures of the participants. Mixed logit models in general, an adaptation of which was used in chapter 3 of this thesis, are able to explain the unobserved correlations among the alternatives and choice tasks by specifying a mixing distribution on the parameters of interest. The continuous mixing distributions often provide a good fit to the data, but the resulting correlation structure (i.e. the asymptotic variance covariance matrix) among parameters is a black box which implies that the cause of the distribution cannot be readily explained (Walker & Ben-Akiva, 2011). Moreover, mixed logit models have also found criticism as they require the analyst to make an a priori assumption on the type of the mixing distribution to be used. Fosgerau (2006) and Hess et al. (2005) discuss some of the unfavourable consequences of specifying an incorrect distribution on the model interpretation, parameter and WTP estimates. In fact, simply knowing that a parameter is randomly distributed across individuals is of lesser interest to policy-makers (Hess et al., 2009). These challenges prompted us to study more about advanced econometric frameworks like the hybrid choice models. These models not only overcome the limitations of the mixed logit model, but also provide a more comprehensive and coherent interpretation of the results. We used a hybrid choice model called the Latent Class Choice Model (LCCM) in this study to analyse the SC dataset.

A Latent Class Choice Model (LCCM) is a statistical tool that can reveal the underlying subgroups of individuals from the observed multivariate data based on the frequency of these variables and response patterns (Hagenaars & McCutcheon, 2009). The tool, which was first developed in the field of marketing sciences (Kamakura & Russell, 1989), is a parsimonious technique of clustering the observed choice patterns of individuals into mutually exclusive latent segments. Unlike parametric discrete choice models like the mixed logit, LCCM does not require any mixing distribution to be assumed upfront. It, in turn, identifies latent segments in the population from the observed data using information, such as the socio-demographics of individuals. Unlike the cluster analysis approach which congregates points in a sample to form separate groups (segments), the LCCM is a statistical model which even tells the probability of a point belonging to a particular segment (Antoine & Molenaar, 2016). It also does not make assumptions such as linearity, normality and homogeneity that are required in cluster analysis (Vermunt & Magidson, 2002). The LCCM is similar to a factor analysis with the key difference being a discrete latent variable (number of segments) in the LCCM against a generally continuous latent construct in factor analysis (with recent computational developments researchers have also started using factor analysis on ordinal data). LCCMs have found numerous applications in transportation planning to study the heterogeneity in the mode choice behaviour (Hess et al., 2009) and modality styles of individuals (Krueger et al., 2016; Vij et al., 2013).

The LCCM comprises two components, namely, a class membership model and a discrete choice model. The class membership model expresses the unobserved latent class segments in terms of the available data, like the socio-demographic information of individuals. The model can be specified as a multinomial logit (MNL) which estimates a set of coefficients that are the same for the individuals within a segment. The choice model, on the other hand, evaluates the probability of observing the response pattern of the individual, conditioned that the individual belongs to a specific latent segment. The response pattern can be a set of choices made by the individual in the SC experiment. The choice model can be specified using different formulations depending upon the nature of the available response data. The integrated framework is then run multiple times by progressively increasing the number of latent classes at each run. The optimum number of latent segments is determined based on the three criteria: 1) overall goodness of fit, 2) model parsimony, and 3) behavioural interpretation of the latent segments (Vij et al., 2013). We now discuss the model formulation that will be used in this study. We use the following formatting styles to represent matrix algebraic notations in the following section: *scalar* quantities are written in italics, ***vectors*** in italics and bold face, and **matrices** in bold face.

4.4.1 Model formulation

Consider that the collected dataset for N individuals contains two parts: cross sectional data on the socio-demographic information and panel data of the choice patterns for every individual. We first discuss the class membership model specification. Assuming that the sample comprises C latent class segments, the utility (or the membership propensity) (U_{nc}) for individual n belonging to latent class c is given by equation 4.1.

$$U_{nc} = \boldsymbol{\alpha}_{\boldsymbol{c}}' \boldsymbol{W}_{\boldsymbol{n}} + \varepsilon_{nc} \tag{4.1}$$

In this equation, $\boldsymbol{\alpha_c}$ is a vector (of size $p \times 1$) of parameters that is exclusive to class c. $\boldsymbol{W_n}$ denotes a vector of observed socio-demographic characteristics of n of size $p \times 1$ where p is the number of observed attributes. ε_{nc} represents the idiosyncratic error term and is considered to follow a Gumbel distribution with a variance of $\pi^2/6$. This forms the logit kernel for the class membership model which is given in equation 4.2.

$$\gamma_{nc} = \frac{\exp(\boldsymbol{\alpha_c'W_n})}{\sum_{k=1}^{C}\exp(\boldsymbol{\alpha_k'W_n})} \tag{4.2}$$

In equation 4.2, γ_{nc} is the latent class prevalence for individual n being in class c. In order to maintain model identification, one of the latent segments is set as the base category. It means that parameters for only $C - 1$ segments can be estimated from a class membership model, with $\boldsymbol{\alpha_c}$ vector for the base category being normalised to zero.

For the choice model, an error component logit (ECL) specification is used to capture the correlation across the multiple choice tasks for individual n. Assume that an individual is presented with T choice tasks, each of which comprises J alternatives. The utility ($U_{njt|c}$) that individual n, belonging to class c, derives from alternative j in choice task t is given by equation 4.3 where $\boldsymbol{X_{njt}}$ is a vector (of size $r \times 1$) of r route specific attributes presented in that choice task for the alternative. $\boldsymbol{\beta_c}$ is a vector of generic parameters of size $r \times 1$ and σ_c is the estimated variance of the error component for every latent segment c. The error component $\xi_{nj|c}$, which is considered to capture the impact of multiple responses by one individual, is assumed to be normally distributed with a mean and variance of 0 and 1 respectively. ε_{njt} is again the idiosyncratic term (like ε_{nc} in equation 4.1) that follows a Gumbel distribution. Equation 4.4 gives the logit kernel for evaluating the probability of choosing the alternative in a single choice task.

$$U_{njt|c} = \boldsymbol{\beta_c'X_{njt}} + \sigma_c\xi_{nj|c} + \varepsilon_{njt} \tag{4.3}$$

$$P_{njt|c} = \int \frac{\exp\left(\boldsymbol{\beta_c'X_{njt}} + \sigma_c \xi_{nj|c}\right)}{\sum_{l=1}^{J} \exp\left(\boldsymbol{\beta_c'X_{nlt}} + \sigma_c \xi_{nl|c}\right)} f\left(\xi_{nj|c}\right) d\xi_{nj|c} \tag{4.4}$$

Let $\boldsymbol{Y_n}$ be the vector (of size $1 \times T$) of observed response pattern across T choice tasks for the individual n. Then the probability of observing $\boldsymbol{Y_n}$ conditional on the latent class c is given by equation 4.5. In this equation, $\mathrm{y_{njt}}$ is an indicator which is equal to 1 if individual n selects alternative j in task t and 0 otherwise.

$$P(\boldsymbol{Y_n}|c) = \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \tag{4.5}$$

Equation 4.6 gives the total probability of observing $\boldsymbol{Y_n}$ across C latent segments which is calculated as the expected value of latent class prevalence and its corresponding conditional choice probability.

$$P(\boldsymbol{Y_n}) = \sum_{c=1}^{C} \gamma_{nc} \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \tag{4.6}$$

Equation 4.6 is repeated over all individuals N to give the likelihood function. Equation 4.7 gives the likelihood function for the LCCM model. In this equation, $\boldsymbol{\alpha}$ is a matrix of size $(p \times (C-1))$ formed by horizontally concatenating estimable membership coefficients for each of the $C - 1$ segments. Similarly, $\boldsymbol{\beta}$ is a matrix of the choice parameters of size $(r \times C)$ and $\boldsymbol{\sigma}$ is a $(C \times 1)$ sized vector of the error components. The log-likelihood function can then be formed as shown in equation 4.8. The objective is to determine the set of parameters that maximises this equation.

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} \sum_{c=1}^{C} \gamma_{nc} \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \tag{4.7}$$

$$LL(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \ln \sum_{c=1}^{C} \gamma_{nc} \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \tag{4.8}$$

4.4.2 Model estimation

Since equation 4.8 comprises the term $P_{njt|c}$, which is a unidimensional integral, we used simulation technique to estimate the log-likelihood function. Equation 4.9 shows the simulated log-likelihood function for the LCCM where R is the number of standard Halton draws used in the estimation routine (Train, 2009). Equation 4.9 can be re-arranged to get equation 4.10 which represents the simulated log-likelihood function of the LCCM. The equation can be maximised to recover the parameter estimates using the non-linear unconstrained numerical optimisation scheme proposed by Broyden-Fletcher-Goldfarb-Shanno (BFGS). The advantages of using this algorithm have been discussed earlier in section 3.5.2 of this thesis.

$$MSL(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \ln \sum_{c=1}^{C} \gamma_{nc} \left(\frac{1}{R}\right) \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \tag{4.9}$$

$$MSL(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \ln \left[\left(\frac{1}{R}\right) \sum_{c=1}^{C} \gamma_{nc} \prod_{t=1}^{T} \prod_{j=1}^{J} (P_{njt|c})^{y_{njt}} \right] \tag{4.10}$$

The estimation procedure of the LCCM was coded in Matlab which can be accessed through the weblink given in appendix B of this thesis. The simulated likelihood function given in equation 4.10 was constructed using 1,000 standard Halton draws for each of the random and error components (Train, 2009). The function was optimised using the BFGS algorithm and the standard errors of the parameters were calculated by taking the inverse of the simulated Hessian matrix.

4.4.3 Calculating the WTP estimates

Once the set of parameters ($\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}$) is estimated upon solving the likelihood function (given in equation 4.7), we evaluated the different Willingness To Pay (WTP) measures using this information. Separate sets of the WTP estimates were calculated for each of the C latent segments. The estimates were obtained using the equations 3.14 – 3.18 which have been discussed earlier in section 3.5.3 of this thesis.

4.4.4 Model results

As discussed earlier, the LCCM model involves simultaneous estimation of the class membership model and the choice model. The choice model consisted of the four attributes from the SC experiment and a normally distributed error component term capturing the correlations across the multiple choice tasks. The membership model was defined in terms of the socio-demographic characteristics of the participants. Of the different socio-economic variables that were tried in the membership model, the model with age, gender and income had a superior goodness of fit along with model parsimony. The dichotomised age and income variables signified young people (between 20-40 years) versus older (Erikson & Erikson, 1998), and low-income (below AU $25,000 p.a.) versus high-income respectively (ABS, 2015). Once the attributes for the membership model were identified, the LCCM was re-executed multiple times by incrementing the number of latent segments after each run. Table 4.12 provides a comparison between the LCCM runs with different numbers of latent segments on the basis of overall goodness of fit measures such as log-likelihood (LL), adjusted rho-squared (ρ^2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

Table 4.12: Goodness of fit measures for the different LCCM specifications

Latent segments	Estimated parameters	LL	Adj. ρ^2	AIC	BIC
2	14	-2207.74	0.192	4443.48	4532.77
3	23	-2072.81	0.242	4191.63	4338.32
4	32	-2030.98	0.257	4125.96	4330.05
5	41	-1988.19	0.273	4058.38	4319.88
6	50	-1970.72	0.279	4041.45	4360.35

The table shows that the number of estimated parameters rises sharply from 14 to 50 as the number of latent class segments increases from 2 to 6 respectively. In other words, the LCCM becomes more complex upon adding new latent classes. The final log-likelihood and the adjusted rho-squared values improve significantly between classes 2 and 3, but are relatively stable beyond 3. A monotonically decreasing trend can be seen for the AIC measure. On the other hand, the BIC measure depicts a nearly U-shaped trend, with a plateau formation between 3 and 5 latent segments. Thus, from table 4.12, we finally select the LCCM with three latent segments for analysis. The selected model is parsimonious, has the least AIC and BIC measures of fit, and a behavioural interpretation (based on the empirical analysis presented earlier in section 4.3) for each of the three latent segments.

Table 4.13: Parameter and WTP estimates from the three segment LCCM

Parameters	Class 1	Class 2	Class 3
Class Membership Model			
Constant	0.2888	0.2903	0
Females	0.5744	0.4691 **	0
Age (below 40 years)	-0.8136 ***	-0.6606 ***	0
Income (below 25K)	-0.325	-0.353	0
Choice Model			
Travel time	-0.3948 ***	-0.1913 ***	-0.0413 ***
Time spent in stop-&-go	-0.0951 ***	0.0055	-0.0488 ***
Number of stop-&-gos	-0.0374 ***	-0.0035	-0.1656 ***
Running cost	-0.7085 ***	-3.9599 ***	-0.5231 ***
Sigma (σ)	0.801 ***	0.604 ***	0.4526 ***
Representation in sample (%)	47	22	31
WTP Measures			
Running cost – Travel time (AU $/hr)	33.43	2.90	4.74
Running cost – Time in stop-&-go (AU $/hr)	41.49	2.82	10.33
Running cost – No. of stop-&-go (AU $/stop)	0.05	0.00	0.32
Travel time – No. of stop-&-go (min/stop)	0.09	0.02	4.01
Time in stop-&-go – No. of stop-&-go (min/stop)	0.08	0.02	1.84

*** significant at 99% ** significant at 90%

Table 4.13 gives the estimated parameters of the LCCM model, along with the WTP measures for each of the three latent segments. A negative coefficient on age signifies that young people are less likely to be in latent segments 1 and 2. Females are more likely to be in segment 2 than the remaining segments. A negative sign on the income coefficients signifies that the individuals in the two subgroups are less likely to be from the low-income group. However, the coefficients for income are not statistically significant at 90 percent confidence interval. One of the reasons for that could be a correlation among the categorical explanatory variables. Nonetheless, we still included the income variable in the utility equation because of two reasons: 1) a membership model primarily serves as a prediction model, thus it is acceptable to include covariates which bear correlation with one another as long as the goodness of fit is reasonably good, and 2) income is a key attribute which is generally used to stratify the population into segments. Interestingly, the model indicates segregation of older and high-income people into two different segments.

Results from the choice model show a negative and highly significant parameter values for most of the attributes across the three segments. However, the coefficient for the time spent in stop-&-go (S&G) and the number of S&Gs is statistically insignificant and close to zero for the individuals belonging to latent class 2. Moreover, this segment also has a highly negative coefficient for the running cost. A significant error component variance across all the segments confirms the presence of the unobserved correlation across the multiple choice tasks in the available SC dataset. We now discuss some interesting observations from table 4.13. The identified latent segments can be labelled as follows:

Class 1 – Travel time minimisers: These constitute 47 percent of the sample population and assign a similar weight on the disutility towards travel time and running cost. In other words, drivers in this segment have a high value of time (AU $33.43 per hour) which indicates their high productivity. They are in general less perturbed by the discomfort due to the number of S&Gs, which can be seen through a small coefficient value of -0.0374. The trade-off value observed between travel time and the number of S&Gs is also less (0.09 minutes per stop) which signifies a willingness to travel for an extra 6 seconds to reduce the number of S&Gs by one on the travelled route by the individuals belonging to this segment. This segment mainly includes individuals from the older age group who are also in the higher income bracket. These people can be considered as “time-poor” as they have additional professional and personal responsibilities upon them as well as extra disposable income. Thus, these people do not mind spending extra on the running cost in order to reduce their travel time.

Class 2 – Calm cost minimisers: Comprising 22 percent of the sample population, drivers in this segment have a very high disutility towards running cost (-3.9599) when compared to travel time (-0.1913) and an insignificant and near zero effect towards S&G traffic. In other words, drivers in this group pivot their route choice decisions mainly around cost and are indifferent towards occurrences of S&G. They are generally comfortable on departing early from home to compensate for a route with a longer travel time, but a lesser travel cost. A fairly low value of time (AU $2.90 per hour) explains their purely cost minimising behaviour. Similarly, the statistically insignificant trade-off value between travel time and the number of S&Gs (0.02 minutes per stop) suggests their calm or tolerant attitude towards the two attributes. This segment also comprises people from the old age and high-income groups. Interestingly, being a female makes one more likely to be in this segment. This finding is also backed by a study by Srinivasan (2005) which found women to spend lesser on travel than males.

Class 3 – Restless cost minimisers: At 31 percent of the sample population, drivers in this segment have a high disutility towards running cost (-0.5231) when compared to travel time (-0.0413), along with a negative and significant disutility towards S&G traffic. They are quite sensitive towards the number of S&Gs experienced, which can be seen through a high disutility coefficient of -0.1656. They generally have a less value of time (AU $4.74 per hour), but assign a very high weight on travel time to reduce an occurrence of S&G (4.01 minutes per stop). This segment mainly comprises young people who also fall in the low-income group, which explains their cost saving attitude. These people tend to get frustrated easily while experiencing alternating cycles of S&G waves. This observation is in line with the past literature which also found young drivers to exhibit higher levels of stress and frustration (Hauber, 1980; Wiesenthal et al., 2000).

4.5 Discussion

This chapter extended the previous study (the proof of concept study discussed earlier in chapter 3) by testing the research hypothesis on a much representative (at a regional level) survey sample. This study overcomes a few limitations of the previous study which are discussed as follows. Firstly, the expanded study was conducted on a sample of general car commuters residing in Sydney and its neighbouring regions, unlike the previous study which was conducted on university staff. Secondly, only the regular car commuters (who drove at least thrice a week to work) were surveyed in this study, in contrast to both regular and occasional in the previous study. Thirdly, we had better control on participant selection while collecting data in this study, which we did not witness in the last study. Thus, the data collection strategy we adopted in this study ensured a sample that: 1) better depicted the travel characteristics of car drivers in Sydney region, 2) had a better knowledge and experience of stop-&-go (S&G) traffic conditions, and 3) comprised a good mix of participant demographics in the resulting sample. Additionally, this study also explained the role of socio-demographic characteristics on the resulting route choice preferences of the individuals. The study utilised the LCCM to classify the participants among different segments (three) based on their socio-demographic information. On the contrary, these variables could not be included in the previous analysis due to the statistical challenges while including them in the RPECL model (refer to section 3.5.4 of this thesis for more information).

The quantitative analysis brought out some interesting and important findings in this study. The results show that nearly three-quarters of the sample had a disutility towards the number of S&Gs experienced. This is in accordance with the proposed research hypothesis, which was even validated in the proof of concept study. Especially, the analysis found that the *restless cost minimising* users associated a higher disutility towards the occurrence of S&G compared to the *travel time minimisers* group. Interestingly, the *calm cost minimising* group of users were found to be indifferent towards S&G conditions and mainly pivot their route choice decisions around travel cost. This find could not be identified in the previous (RPECL) model.

The findings from this study expand the existing knowledge base on understanding the intricacies associated with the route choice behaviour of car drivers. Previous studies found that drivers find it more onerous spending time in S&G traffic than in normal driving conditions (Hensher, 2001a). This study found that the number of S&Gs also contributed towards the disutility for a given route. An explanation for this observation could be that drivers need to be more focussed while undergoing alternating cycles of S&G traffic, which causes discomfort to them. This discomfort eventually becomes visible in the form of increased physiological activity and frustration level which eventually results in aggressive driving behaviour. These reasons lead to drivers considering factors other than minimising travel time while making routing decisions. Instead, a more inclusive approach would be to consider them to follow a mix of two strategies, i.e. minimising both travel time and discomfort (which can be expressed in terms of the number of S&Gs). Thus, the findings from this study would potentially be useful in modifying existing transport modelling techniques to reflect a more realistic route choice selection process of drivers.

This study provides richer and more representative information (when compared to the proof of concept study) which can be scaled up at a regional level in Sydney. The results from this study will provide meaningful inputs to planners to frame policies that ease users of the discomfort caused by S&G traffic. Two potential policies and their implications can be suggested at this point, namely the toll pricing strategies and the introduction of autonomous vehicles. We discuss these policies in chapter 6 of this thesis.

This study finds more relevance to the application in the real world as it overcomes a few restrictive assumptions that were made in its precursor study. However, the limitation with regard to the measurement bias associated with the number of S&Gs still persists in this study. In addition to that, results from the LCCM revealed a need to even include attitudinal variables in the modelling paradigm. Due to the onerous nature of S&G traffic, the factors such as the level of frustration and mental load, along with socio-demographics, often influence route choice behaviour of drivers

(Levinson et al., 2004; Malta et al., 2011). Nonetheless, our objective at the start of this study was to gradually narrow down the set of assumptions that were made in the predecessor study. Once the assumptions regarding the goodness of the sample have been accommodated, we plan to next address the remaining assumptions in the following chapter of this thesis.

CHAPTER 5

EXPERIMENT III – DRIVING SIMULATOR STUDY

The chapter aims to study the effect of stop-&-go (S&G) traffic characteristics on the levels of frustration in drivers and also to investigate the association between frustration and route choice behaviour under such conditions. Additionally, this chapter narrows down the set of assumptions which were associated with the previous experiments discussed in chapters 3 and 4 of this thesis. This chapter tests the modified research hypothesis by conducting a driving simulator experiment which was useful in minimising the measurement bias with regard to quantities like the number of stop-&-gos (S&Gs) experienced and the consequent level of frustration (a latent variable). The collected data was analysed using a Structural Equation Model (SEM), to quantify the impact of S&G traffic attributes along with individual specific characteristics on the levels of frustration experienced by drivers on a route. The obtained results would provide a better understanding about the development of driver stress in S&G traffic and whether it affects route choice.

The organisation of this chapter is as follows: Section 5.1 provides a short background on the motivation behind conducting this experiment. Section 5.2 provides a brief description of driving simulators highlighting their applications, merits and utility to this study. Section 5.3 delineates the steps that were followed during the design of the experiment. Section 5.4 presents the experimental layout that was followed during data collection. Section 5.5 discusses the participant recruitment process followed in this study. Section 5.6 presents an empirical analysis of the collected data. Section 5.7 first presents a discussion on the SEM which is then followed by its mathematical formulation, estimation and model results. Finally, a discussion of the results is taken up in section 5.8.

5.1 Motivation behind the Experiment

Results from chapters 3 and 4 indicated disutility of a route to increase as the number of stop-&-gos (S&Gs) experienced on it increased. It is the alternating nature of S&G, requiring drivers to be more focussed, that causes them to experience elevated levels of discomfort and frustration (Levinson et al., 2004). A few studies have found a positive correlation between traffic congestion and driver stress. For example, Jovanović et al. (2011) suggest traffic congestion as one of the sources of frustration. Hennessy & Wiesenthal (1999) conducted telephone interviews with the participants driving under different traffic conditions. The study found that the participants gave a higher frustration score while driving in congested traffic than normal conditions. Lazarus (1966) also found adverse (congested) driving conditions along with time pressure to contribute towards driver frustration. This induced frustration, as established in the literature, is the precursor to aggressive driving behaviour which poses a risk to safe driving (Blanchard & Blanchard, 1984; Blanchard et al., 2000; Hennessy & Wiesenthal, 1997).

The above discussion indicates that driving in congested or S&G traffic does make drivers more annoyed and frustrated. The hypothesis we tested and quantified in this experiment was: *An increase in the number of S&Gs on a route makes drivers more frustrated*. We conducted a driving simulator experiment to test this new research hypothesis. Additionally, we wanted to test whether there exists a dependency between driver frustration and their resulting route choice behaviour in the driving simulator experiment. The following section provides a discussion on driving simulators highlighting their merits and usefulness to this study.

5.2 Using Driving Simulators for Data Collection

A driving simulator is an advanced setup which presents a virtual driving world before participants in a controlled environment. The virtual scenario, which is designed by the analyst, corresponds to a traffic situation of interest which is encountered in day-to-day life. Participants are required to drive through the scenario which gives an opportunity to closely study their driving behaviour. The Stated Choice (SC) experiment approach, used in the previous chapters, is an effective tool for answering the research question, but suffers from one shortcoming. In an SC experiment, competing alternatives are defined using a limited number of attributes which are of interest to the analyst (since adding more attributes might lead to situations like attribute non-attendance and lexicographic behaviour in participants (Campbell et al., 2006)). In other words, the analyst studies the relative importance and trade-off among these key attributes *ceteris paribus*. However, the actual choice making is a much more complex process which cannot be exhaustively modelled using few key attributes. Driving simulators, on the other hand, present a wholistic scenario where a participant simultaneously interacts with the surrounding traffic, roadway infrastructure and environment. Thus, the obtained data offers more realism as it reflects close to the actual driving behaviour of the participant.

Driving simulators have been around in transportation and related fields for more than two decades. Technological advances over the years have witnessed a transition from the basic PC based driving simulator (Beede & Kass, 2006; Koutsopoulos et al., 1994) to a more sophisticated fully instrumented cabin (Haque & Washington, 2014; van Driel, et al., 2007). Driving simulators have found a wide spectrum of applications in transport research related to: traffic safety and accident (Dixit et al., 2014; Fiorentino & Parseghian, 1997; Lee et al., 2003; Lee et al., 2002), reaction time and gap acceptance (Alexander et al., 2002; Choudhary & Velaga, 2017; Farah et al., 2009; Haque & Washington, 2014; Yan et al., 2003), route choice behaviour and driver disutility (Jeihani et al., 2017; Koutsopoulos et al., 1995; Levinson et al., 2004; Tian, 2010), user acceptability towards intelligent vehicle systems (Hoedemaeker, 2000; Stanton et al., 1997; van Driel et al., 2007; Xiong & Olstam, 2015), risk attitudes of drivers (Dixit et al., 2015; 2017), and physiological and psychological aspects of driving (Abou-Zeid et al., 2011; Mehler et al., 2009; Reimer et al., 2006; Lee, 2010). These studies brought out several interesting findings, which in a few cases were quite different to the existing alternate methods. For example, Levinson et al. (2004) also conducted an online SC experiment, alongside the driving simulator experiment, to compare driver disutility towards the ramp and freeway delays. While the results from the former indicated that drivers found ramp delays as more onerous than freeway delays, the driving simulator study brought out contrasting results. The authors also compared both the methodologies and highlighted the probable reasons that could have produced strikingly different findings.

The advantages of using driving simulators are as follows. Firstly, they provide a more realistic data gathering platform when compared to SC experiments. As a result, they help in reducing the hypothetical bias (cases where the actual preferences of participants differ from what they reported in the SC experiment) and measurement errors which might get introduced in the latter technique, despite a well-designed experiment. Secondly, they provide a safe environment for data collection, unlike naturalistic (real-time) data collection methods. Although the naturalistic driving data offers the highest level of realism, it has limited availability mainly due to safety concerns. On the other hand, driving simulators collect data in a controlled environment, thus making them quite safe. Thirdly, they also facilitate the collection of various psychological and physiological factors, due to a more engaging environment, which are hard to perceive in an SC experiment. Despite its merits, the use of driving simulators is mostly limited due to the high installation and operational cost.

We decided to re-test the research hypothesis by conducting a driving simulator experiment to account for the limitations discussed at the end of chapter 4 of this thesis. The reasons behind going for a driving simulator study were: 1) to let the participants experience driving in stop-&-go (S&G) traffic, giving them a better idea about the number of S&Gs faced, and 2) to study the level of frustration which gets induced while driving in such conditions.

5.3 Design of the Driving Simulator Experiment

This section explains the methodology involved in designing the virtual scenarios for the driving simulator experiment. This section describes the design specifications of several components like the logic used to determine the attributes of interest, questionnaires, and the layout of the experiment. We start this section with a discussion of the driving simulator equipment that was used for data collection.

5.3.1 Apparatus

The study was conducted using a high fidelity driving simulator installed at the Travel Choice Simulation Laboratory (TRACSLab), UNSW Sydney. The TRACSLab facility was recently set up (in 2015) which gave us the additional boost to conduct this study. Figure 5.1 shows the driving simulator that was used in this study. The setup comprises interiors of a Holden Commodore car and includes a driver's seat, steering wheel, accelerator and brake pedals, odometer and an automatic transmission gearbox.

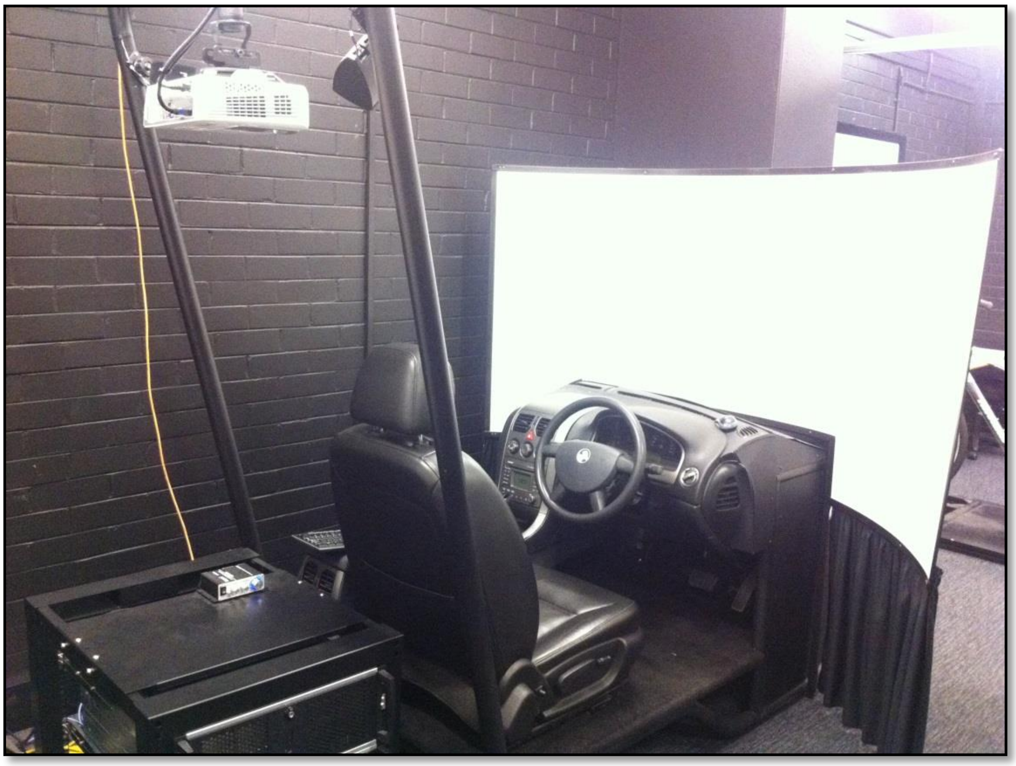

Figure 5.1: Driving simulator setup used in the experiment

The projector screen displays a 150 degree, high resolution, horizontal view of the virtual scenario to the participants. The central and right-side rear view mirrors are also displayed on the screen. The simulator exerts a resistive force while operating the steering wheel or the brake pedal. The odometer is calibrated to give a better sense of vehicle speed to the participants. The simulator also comes equipped with an audio system that recreates sounds such as vehicle noise, honking, etc. All these features collectively offer a realistic driving environment to the participants providing a platform to minutely study their behaviour. Data such as position, speed, acceleration, deceleration and braking can be collected at a very fine resolution, which was set at 10Hz (every 0.1 seconds) for this study.

5.3.2 Design of virtual scenarios

The first step involved in the design of the virtual scenarios was recreating a hypothetical city called Congestington. The built-up environment of the city consisted of houses, government offices, high rise buildings, schools, hospitals, shopping areas, etc. Figure 5.2 shows the line diagram of the road network map of the hypothetical city. The participants started the driving task on the western end of the city (Country Drive) and the objective was to reach their workplace located at the eastern end of the city (Alpha Consultancy). Two red and white chequered boxes (see figure 5.2) were placed by the roadside at the destination to demarcate the finish line. The road network consisted of two unlabelled one lane bi-directional routes, addressed as the left route and the right route henceforth, which were aesthetically similar to one another. Providing single lane roads along with no left turning lanes at the two T-intersections ensured that the subject vehicle had to stop or slow down while entering and crossing the T-intersection. Thus, it was safe to assume that the participants had to put similar amount of effort while making left or right turn and treated them equally. This assumption was empirically verified by taking feedback from the participants at the end of the experiment. None of the participants reported that they experienced trouble while making a particular turn. Providing a single lane also ensured that the participants experienced S&G traffic which would have become challenging to recreate in a multi-lane environment. The length of the left and right routes was around 4.4 and 4.0 km respectively. The posted speed limit sign boards read 50 km/h and no overtaking or lane changing was allowed on both the routes. Audio messages were played to guide the participant's vehicle (referred to as the subject vehicle henceforth) to the chequered box (destination). Alternatively, the participants could follow the vehicles in front (which were programmed to reach the destination) in case they missed the audio messages. The subject vehicle was part of a group of vehicles on each route where the leader vehicle of the platoon (a truck) was programmed to initiate S&G waves.

Different levels of S&G traffic conditions were recreated on the two routes. This was done by placing proximity sensors, denoted by shaded rectangles in figure 5.2, on both the routes. A proximity sensor corresponds to the spatial location where the leader vehicle was programmed to trigger S&G waves. A total of 3 and 5 proximity sensors were placed on the left and right routes respectively. The leader vehicle decelerated to a halt every s seconds upon entering the proximity sensor and resumed acceleration after d seconds in a stopped position.

A set of three driving scenarios were designed for the experiment where each scenario comprised the same two alternatives (routes) but with varying S&G conditions. Thus, a total of 6 driving tasks (2 routes across 3 scenarios) were developed by altering the values of s and d along with a few other parameters. The parameters were set in a way such that the left route had a lesser number of S&Gs of longer duration while the right route had a greater number of S&Gs of shorter duration. The reasons behind maintaining this design across all scenarios were: 1) given that there were 3 scenarios to be presented, the participants could only start learning about the routes from the third scenario which happened to be the last, thus causing minimal learning bias, and 2) the shuffled scenario would have required an extra effort to be put in modifying the scenarios.

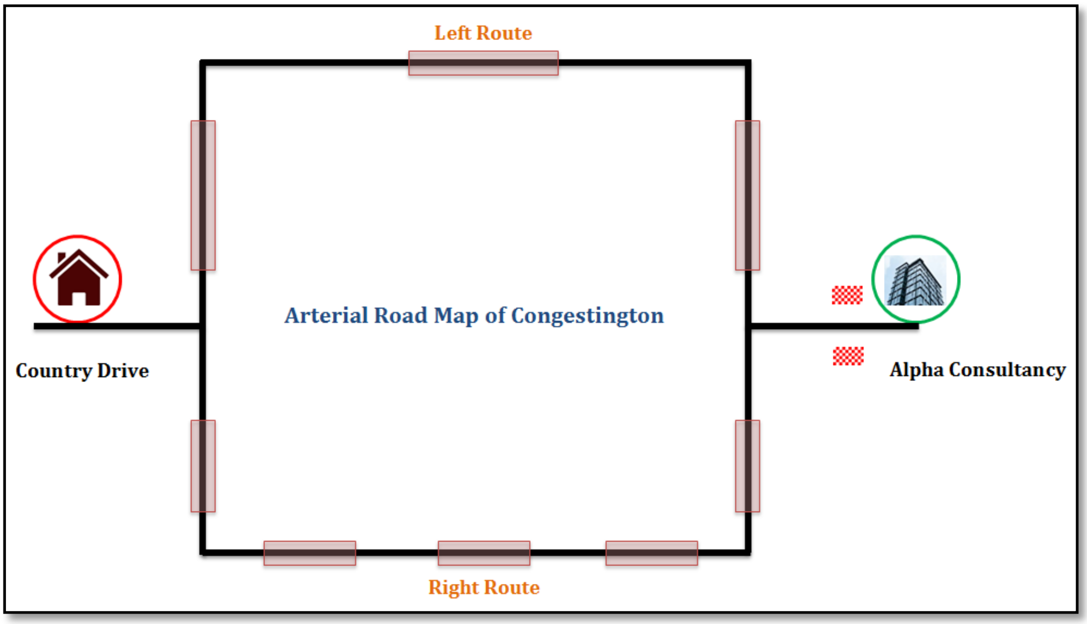

Figure 5.2: Road network map of the hypothetical city

Table 5.1 gives the parameter values which were used in the design of each driving task. The values of s and d were generally drawn from a uniform distribution to add more realism to the driving tasks. The acceleration and deceleration rates were kept constant across all the scenarios so that a variation in their intensity does not affect route choice. The parameter ambient traffic density was used to set the number of vehicles to be simulated as surrounding traffic in every driving task. While a proportion closer to 0 signifies very few vehicles, a value closer to 1 represents higher ambient traffic density. The cruising speed of the simulated vehicles forming the platoon was set at 50 km/h for all the driving tasks except for the left route in scenario 3 (42 km/h) which represented driving at a lower speed in congested traffic.

Table 5.1 also gives the resulting values of the four attributes of interest (travel time, time spent in S&G, number of S&Gs and fuel cost) for a given setting of the parameters. These values were obtained after making multiple test runs to check for the suitability of the tasks and the resulting scenarios. The exact procedure for deducing the attributes is discussed below:

i. Travel time: It was calculated as the difference between the time instants when the subject vehicle started driving and once it reached the destination (upon crossing the chequered boxes) during every task. Sensors were placed at these two locations which recorded the timestamp the moment the subject vehicle was detected. Travel time on the left route was maintained between 1.1 to 1.25 times the travel time on the right route in each of the three scenarios (refer to table 5.1).

Table 5.1: Calibration parameters used in the design of virtual scenarios

Parameter	**Scenario 1**		**Scenario 2**		**Scenario 3**	
	Left	**Right**	**Left**	**Right**	**Left**	**Right**
S&G start (s) (sec)	[13,16]	[4,4]	[8,10]	[4,5]	[20,25]	[5,6]
Stopped time (d) (sec)	[22,25]	[4,6]	[14,16]	[5,6]	[14,15]	[7,7]
Ambient traffic density	0.3	0.3	0.4	0.4	0.5	0.5
Stream speed (km/h)	50	50	50	50	42	50
Acceleration (m/s^2)	4					
Deceleration (m/s^2)	6					
Resulting values of attributes						
Travel time (mm:ss)	7:51	6:26	8:09	6:33	7:20	6:20
Time in S&G (mm:ss)	4:09	3:28	4:22	3:30	3.02	3.13
Number of S&Gs	7	22	10	19	5	14
Fuel Cost (AU $)	0.59	0.82	0.65	0.76	0.66	0.53

ii. Number of S&Gs: For this study, we adopted the definition of S&G given by Zheng et al. (2011a) which characterises S&G as a cycle involving deceleration followed by acceleration of the vehicle. Thus, we quantified one occurrence of S&G

as the deceleration of the leader vehicle from the cruising speed to zero followed by an acceleration back to its original speed. For this study, we determined the number of S&Gs initiated by the leader vehicle rather than counting the number experienced by the subject vehicle. The reason behind adopting this approach can be explained as follows: The aim of this study was to understand the participant's response to recurring cycles of S&Gs. Such conditions are mainly inflicted upon the subject vehicle due to the presence of vehicles in the front rather than the vehicle itself. Thus, we took into consideration the number of S&Gs triggered by the leader vehicle as an attribute to evaluate a route's disutility for the participant (in the subject vehicle). The total number of S&Gs experienced was accumulated for the entire route and recorded as an attribute. Furthermore, the number of S&Gs on the right route was maintained between 1.75 to 3 times (refer to table 5.1) the number experienced on the left route.

iii. Time spent in S&G: Time spent in S&G refers to the duration for which a vehicle undergoes a cycle of S&G. Timestamps were recorded each time the leader vehicle commenced deceleration towards zero speed and when it resumed its original cruising speed upon acceleration. The time difference between the two instants gave the time spent under S&G for that cycle. This quantity was accumulated over all S&Gs experienced on a route and used as an attribute for analysis. The overall time spent in S&G on the left route was maintained between 0.9 to 1.2 times (refer to table 5.1) that experienced on the right route.

iv: Fuel cost: A fuel consumption model had to be coded for this experiment since the existing driving simulator lacked an inbuilt mechanism. We adopted the fuel consumption model given by Ferreira (1982) which also took into consideration the additional fuel used up during S&G traffic. Equation 5.1 shows the mathematical expression of the model proposed by the author.

$$FC = a_1 D + a_2 T_s + a_3 S \qquad (5.1)$$

Where,

FC: Fuel consumed in litres

D: Distance travelled at cruising speed

T_s: Amount of stopped time

S: Number of stops made

a_1, a_2, a_3: Calibration parameters

Based on an empirical analysis conducted by Ferreira (1982), the values of calibration parameters were set as follows: a_1 = 0.07 l/km; a_2 = 1.2 l/hr and a_3 = 0.022 l/stop. These values were obtained for a 1500cc engine car and at a cruising speed of 48 km/h. Given the speed limit of 50 km/h used in this study, the cruising speed was expected to be around 48 km/h. Thus, we could use these calibration parameter values directly for our case. Equation 5.1 provides us with a consistent mechanism of approximating the actual fuel consumed during each driving task. Furthermore, we evaluated the fuel consumed by the leader vehicle and stored it as an attribute. It was done due to the following reasons: 1) as the subject vehicle was part of a platoon undergoing S&G cycles, the fuel consumption quantity between itself and the leader vehicle should only differ marginally, and 2) the computational ease while calculating the fuel consumed by the leader vehicle (which was programmed) in real-time (the reason for evaluating it in real-time will be discussed in section 5.4). As we discussed above, these assumptions facilitated a simplified model which was expected to consistently give us good estimates of the actual fuel consumed, in the absence of an inbuilt fuel consumption mechanism. The obtained fuel quantity was converted into cost by multiplying it by the average price of petrol in Sydney, i.e. AU $1.30. The fuel cost on the right route was higher and maintained between 1.1 to 1.4 times the cost on the left route.

The range of ratios for each of the four attributes was carefully selected after several trials and satisfied the following conditions: 1) the difference in travel time was not too small to be unnoticed by the participants, and 2) the ratio was not too large to induce non-selection of the route.

5.3.3 Design of questionnaires

Three sets of questionnaires were designed which were planned to be handed over to the participants at different times of the experiment session. This subsection elaborates on each of the questionnaires.

5.3.3.1 Self-reported frustration ratings

The participants were asked to report the levels of frustration experienced at the end of every driving task (6 in all). The questions asked were: *What was the level of frustration you experienced while driving?* and *How frustrated would you be if you have to experience such conditions during your actual commute to work?* The first question was taken from an earlier study by Lee (2010) which looked at the effect of frustration on driver's performance. The second question was asked in case the participants found the driving time on the route as too less. Thus, the second question asked them to rate their levels of frustration had they witnessed similar S&G conditions, as experienced during the task, in their actual commute. The responses (to both the questions) were recorded on a 5-point Likert scale (as used in Lee (2010)) with the anchors 1: Not at all frustrated; 2: Slightly frustrated; 3: Moderately frustrated; 4: Very frustrated and 5: Extremely frustrated. These anchors were taken from the list prepared by Brown (2010) which also provides anchors to a variety of other questions.

5.3.3.2 Route choice questions

Participants were asked to select the most preferred route at the end of the second route in every scenario. The questions asked were: *Having travelled on both routes, which route would you select for your next trip to work*, and *If you had to experience traffic conditions like the two routes for the length of your actual commute, which route would you select*. The second question was asked so as to capture their route choice if they had to undergo S&G conditions similar to what they experienced on the two routes in real-life. The options to both the questions were 1) the first route, and 2) the second route.

5.3.3.3 Socio-demographic questionnaire

A set of questions were designed to gather the socio-demographic information of the participants. The information collected from this questionnaire would later allow us to even study the role these attributes in modelling driver frustration. The questionnaire consisted of 11 questions which asked for information such as gender, age, income, occupation, driving experience, etc.

5.3.3.4 Online SC survey

This questionnaire was exactly the same as the one discussed earlier in chapter 4 of this thesis. Participants were shown a set of 10 choice tasks, where each choice task comprised the currently travelled route (status-quo) and other two hypothetical routes. The attributes (four in all) of the hypothetical routes were pivoted around the attribute values of the status-quo alternative.

5.3.3.5 Attitude towards S&G traffic

Three questions were asked to participants to understand their perception towards S&G traffic in real-life. The responses to these questions were again recorded on a 5-point Likert scale to be consistent with the other route choice questions in the experiment which also used a similar scale (eg. Question on the level of frustration experienced (Lee, 2010)). The first question asked was: *How often do you experience stop-&-go traffic during your actual commutes?* which had the options: 1: Never; 2: Rarely; 3: Sometimes; 4: Often and 5: Always. The second question was: *How frustrated do you feel when driving in such conditions?* which had the options: 1: Not at all frustrated; 2: Slightly frustrated; 3: Moderately frustrated; 4: Very frustrated and 5: Extremely frustrated. The last question was: *How likely are you to look for an alternate route with a slightly higher travel time but fewer stop-&-gos?* which had the options: 1: Extremely unlikely; 2: Unlikely; 3: Neutral; 4: Likely and 5: Extremely likely.

5.4 Experimental Procedure

The experiment comprised the introduction, driving task, online SC survey and attitudinal questionnaire sections. Figure 5.3 shows the sequence of sections presented to the participants in the entire experiment session. A warm-up task was given upfront to familiarise the participants with the mechanisms of the simulator equipment and the virtual routes played before them. The participants drove on both the routes during the warm-up task. It was then followed by the three driving scenarios. The order of the scenarios was randomised across the participants using the Latin square design technique (Williams, 1949).

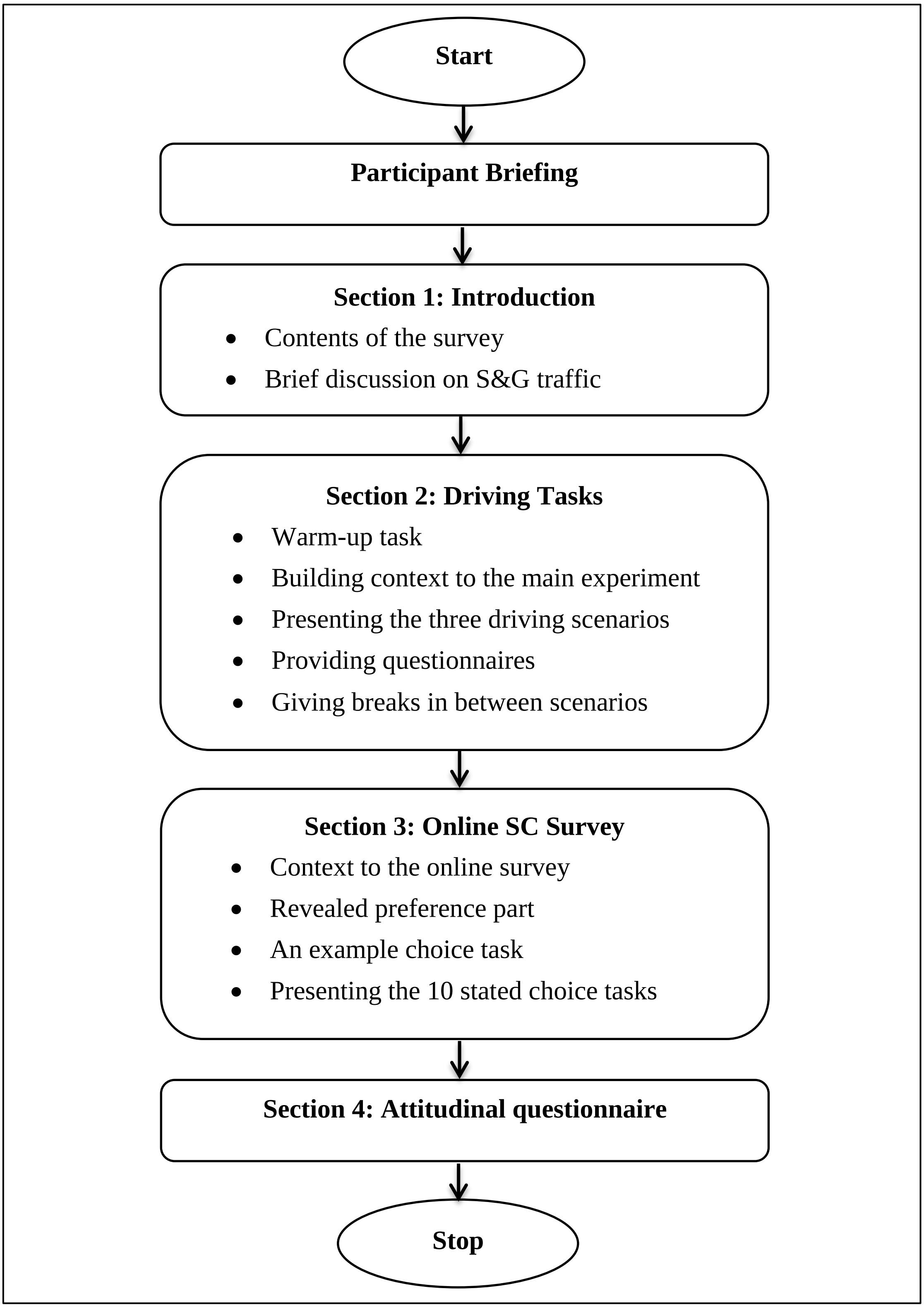

Figure 5.3: Layout of the driving simulator experiment session

The participants were informed about a set of penalties, applicable to the three scenarios, which were as follows: 1) an AU $3 deduction in case the participants violated the road rules discussed earlier in subsection 5.3.2, and 2) each task (route) could be finished within 9 minutes without breaking the road rules, inability to do so would attract a penalty at a rate of AU $1 per minute. The penalties were designed to foster real-world driving behaviour among the participants in the driving simulator experiment. Travel time and fuel cost incurred was displayed on the screen at the end of every route for their information. This was done since drivers can easily evaluate these two attributes in real-life. Information on S&G conditions was not displayed because drivers generally experience such conditions without exactly keeping a count. The participants were then asked to report the frustration ratings at the end of every route within a scenario. A 5-10 minute break was given after the warm-up task and in between the scenarios where the participants were allowed to relax and offered refreshments. A set of questionnaires regarding socio-demographics and route choice and attitude towards S&G were presented during break time and towards the end of the session respectively. A monetary reward of AU $40 was handed to participants upon completing the entire study. The expected duration of the study was around 90 minutes which included time for driving, briefing, relaxation and responding to questions.

5.5 Participant Recruitment

The invitation to the driving simulator study was mainly circulated among university staff and students. This was done by emailing the flyer of the experiment to the administrative managers of the different schools on campus and via the professional staff group email to be circulated among the members. Additionally, the invitation to the study was also posted on social media platforms, like Facebook, and community bulletin boards around campus to attract people from outside (however, despite our efforts, we could not find a good response from the people outside the campus). Individuals who drove to UNSW (Kensington campus) or work by car at least thrice a week were contacted for this study. The exclusion criteria to the study included people who: 1) did not drive to UNSW or work, 2) did not possess a driver’s license, 3) had a history of motion sickness, and 4) pregnant women. A respondent satisfying any of these criteria was not invited to the study. The reason behind using these exclusion criteria was to obtain a sample of car drivers who drove regularly to university or work and might have experienced traffic congestion (and S&G traffic) on a daily basis. The study ran for around 6 weeks in June 2017 and a total of 111 people participated in this experiment. Of these participants, 12 were not able to finish the experiment due to uneasiness or motion sickness. Approvals were again taken from the university human ethics committee, Human Research Ethics Advisory (HREA) Panel H: Science and Engineering, before commencing the study. The copy of the approval letter (HC No. 15752) is available in appendix C of this thesis.

A simulation procedure was followed to determine the minimum sample size required for the quantitative data analysis which will be discussed later in section 5.7. The procedure showed that around 70 participants would provide us with parameter estimates from the statistical analysis at a significance level of 5 percent. Thus, we collected data for 99 participants in all to capture the variance in real-world dataset missing in the simulated dataset. Results from the simulation study are provided in appendix E of this thesis.

5.6 Empirical Analysis

The effective dataset for further analysis comprised responses from 99 participants which equated to 594 rows of observations. Figure 5.4 shows the socio-demographic information of the effective dataset. The experiment saw a majority of participation from males, which comprise three-quarters of the sample. 60 percent of the participants are up to the age of 30 years which typically represents the age group of students. The remaining 40 percent of the sample is above 30 years and corresponds to the usual age group of university staff. The pie chart on the weekly income distribution shows around 50 percent of the sample to earn up to AU $800 per week (AU $41,600 per annum), which roughly equates to the weekly wages or stipend received by the students. The other income groups also have a decent share (around 10 percent representation in the sample) in the sample representing the weekly salary of staff. The employment status and the occupational distribution also shows 54 percent of the sample as students enrolled part-time or full-time and 56 percent of the sample as undergraduate, postgraduate and Ph.D. students respectively. The remaining sample comprises workers like professionals, teaching staff, administrative staff, technicians and other jobs.

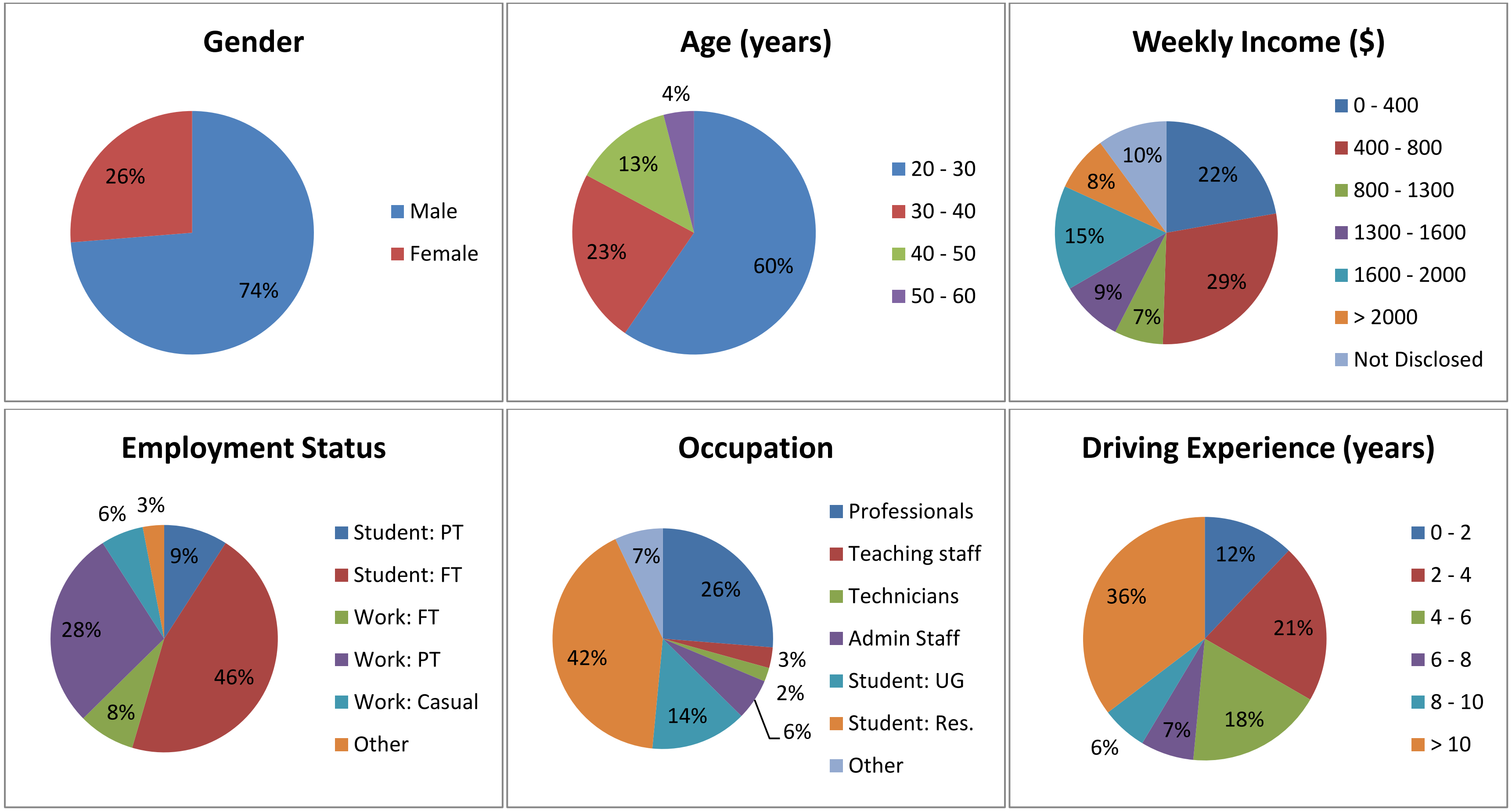

Figure 5.4: Socio-demographic information of the effective dataset

The pie chart on driving experience shows a good representation (around 10 percent representation in the sample) of different experience groups. While almost 40 percent of the sample represents experienced drivers with an experience of at least 8 years, the sample also comprises 33 percent relatively new drivers with up to 4 years of driving experience. The latter group of participants lacks driving experience as they either got their full driver's licence recently or still driving on the provisional licence. Thus, they are expected to exercise more caution when driving in general and easily get perturbed during S&G traffic conditions as it requires additional attention (Levinson et al., 2004; Lee & Winston, 2016).

Table 5.2 shows the travel related information on each of the two routes across three scenarios. The statistics presented in this table were obtained after sorting the scenarios in the order we defined during the design. In other words, the scenario statistics presented in this table do not correspond to the shuffled order which was presented to the participants.

Table 5.2: Descriptive statistics of route specific information

Average attribute value	Scenario 1		Scenario 2		Scenario 3	
	Left	Right	Left	Right	Left	Right
Travel time (mm:ss)	7:24	6:07	7:43	6:25	7:26	6:13
Time in S&G (mm:ss)	3:25	3:04	3:38	3:16	2:59	2:53
Number of S&Gs	5.17	18.4	8.4	17.89	4.84	12.13
Fuel Cost (AU$)	0.54	0.74	0.61	0.73	0.53	0.61
Frustration Rating	2.23	2.62	2.32	2.64	2.14	2.45
Real-world Frustration Rating	2.25	2.70	2.44	2.71	2.13	2.46

The values for the trip related attributes provided in this table match closely to the ones reported in table 5.1 discussed earlier in subsection 5.3.2. This indicates that the parameters used for designing the scenarios were able to re-create, to a great extent, the actual driving behaviour. The table shows that the left route always has a slightly higher travel time but fewer S&Gs and vice-versa for the right route, which is consistent with the experiment design. We conducted a one-way ANOVA to check if the four route specific attributes on the left and right routes were statistically different from one another. Results from the ANOVA test showed that the mean attribute values on both the routes were statistically different from one another at 99 percent confidence interval. For example, the mean values of the number of S&Gs on the left and right routes in scenario 1 were found to be statistically different [$F(1,196)=622.905$ and p-value=0.000].

Table 5.2 also shows the self-reported frustration ratings for each of the six routes driven by the participants. It can be observed that the average frustration rating is always higher on the right route across three scenarios. Thus, the participants experienced higher levels of frustration when asked to drive on a route which had more S&Gs even though it was quicker. This observation indicates the validity of the proposed research hypothesis for this study, i.e. more S&Gs on a route makes drivers more frustrated. The observation is also in line with the previous literature which indicated the level of frustration to be higher under more intense S&G traffic (Hennessy & Wiesenthal (1999) for example).

Table 5.2 also shows the fuel cost to be higher for the right route, which has the higher frustration rating. It could be argued that fuel cost might have also affected the higher frustration rating given by the participants. However, our intuition says that drivers find S&G traffic to be more frustrating as they need to be extra focussed while driving which might not have a bearing on the cost incurred. The literature we discussed above also supports our idea since they too did not consider the effect of fuel cost on the frustration rating (Hennessy & Wiesenthal, 1999; Levinson et al., 2004). Furthermore, we conducted a Pearson correlation test between the self-reported frustration ratings (assumed as a continuous variable for simplicity) and the running cost to empirically verify the relationship. The correlation coefficient was found to be 0.29 which being low indicates weak relationship between the two quantities.

Figure 5.5 shows the route choices made by all the 99 participants across three scenarios. The figure shows that a majority of the participants preferred travelling on the left route, across three scenarios, which had a slightly higher travel time but fewer S&Gs. While the proportional split between the two routes (left: right) was 3:2 in scenarios 1 and 3, it increased to 2:1 in scenario 2. Looking at table 5.2 and figure 5.5, it can be said that a majority of the participants selected the route on which they experienced lesser levels of frustration.

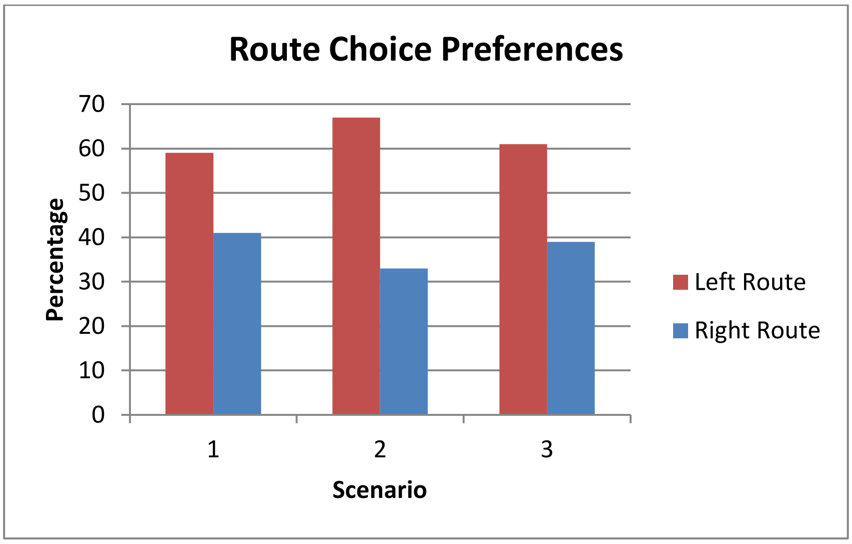

Figure 5.5: Route preferences of the participants across scenarios

Thus, the effective dataset is composed of the participants depicting a wide variety of socio-demographic characteristics (most of the segments under a socio-demographic attribute have around 10 percent representation in the sample) and driving styles which can yield significant relationships (at a statistical significance of 5 percent) and check for the validity of the proposed research hypothesis.

5.7 Structural Equation Model

Traditional discrete choice models are grounded on the principle of Random Utility Maximisation (RUM) to express the decision-making process of an individual. This decision-making process is modelled as a “black box” which takes in observed variables such as the alternate specific attributes and socio-demographic information of individuals and outputs the resulting choice (Ben-Akiva et al., 2002). In other words, this discrete choice modelling framework evaluates the direct effect of the observed attributes on the choices made. However, such models generally do not take into consideration the impact of the unobserved (latent) individual specific characteristics, such as attitudes, norms, perceptions and beliefs which also have a contribution towards their preferences. In fact, past studies have shown that these “soft” latent constructs can often outweigh the effect of the “hard” observed attributes in the actual choice making process (Anable, 2005; Bamberg & Schmidt, 2001; Gärling et al., 2003).

A Structural Equation Model (SEM) is a multivariate modelling technique used to estimate the joint association between the latent constructs (structural) and the observed variables (measurement) (MacCallum & Austin, 2000). SEM is a widely used tool for pattern identification among different variables in the domain of transportation engineering, particularly travel behaviour (Golob, 2003). Recent studies have also used this tool to measure aggressive driving behaviour (Abou-Zeid et al., 2011), develop an aggression index at a signalised intersection (Hamdar et al., 2008) and towards cyclist safety (Nair et al., 2016). The main advantage of using SEM is that it explicitly accounts for the measurement errors when jointly estimating the observed and latent variables (Schumacker & Lomax, 2012). This is particularly useful in this study as the observed frustration ratings can be subject to errors since they do not truly represent the underlying frustration propensity (a latent variable).

A typical SEM framework comprises two sub-models, namely the latent variable structural equation model and the latent variable measurement model. The structural equation model specifies a linear relationship between the latent constructs and the observed covariates. These latent constructs can be observed by the analyst in the form of indicator variables which can be measured on a continuous, nominal or an ordered scale. The measurement equation model expresses a linear relationship between the "soft" latent constructs (computed through the structural part) and the indicator variables and also allows for measurement errors in the latter in capturing the true intrinsic value of the former.

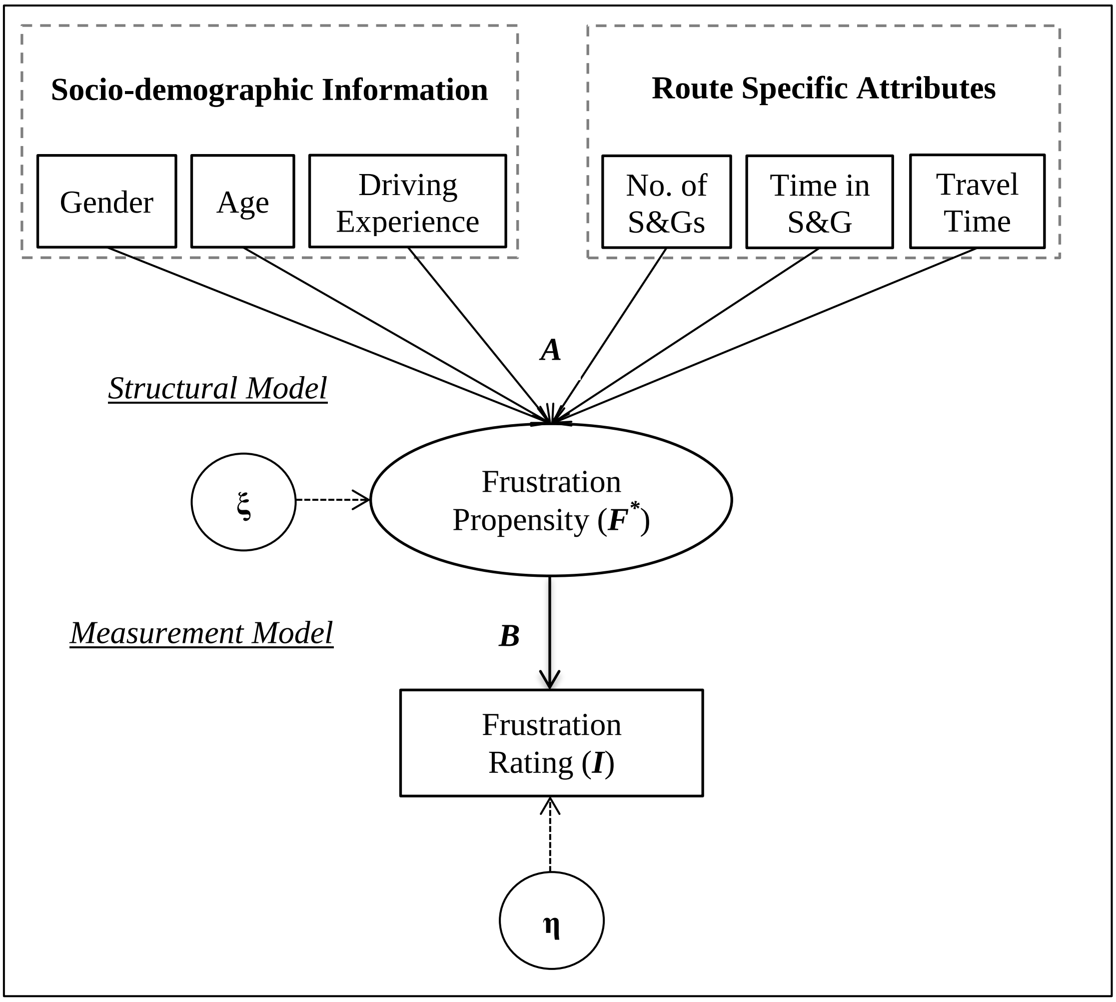

Figure 5.6: The SEM framework adopted

As discussed earlier in section 5.1, the aim of this study was to test the effect of the number of S&Gs on the levels of frustration of drivers. Thus, we decided to use the SEM to analyse the data obtained from the driving simulator experiment. Figure 5.6 shows the path diagram which presents the blueprint of the proposed research hypothesis for this study. The studies we reviewed earlier in section 5.1 showed that drivers tend to experience higher levels of frustration while driving in S&G traffic. This frustration gets developed as a result of two main latent psychological constructs: perception and attitudinal. While perceptions are usually associated with the alternate specific attributes, attitudes are related to the characteristics of an individual (Bahamonde-Birke et al., 2017). As shown in the figure, we define the latent construct, frustration propensity ($\boldsymbol{F}^*$), as the function of the route specific attributes such as the time spent in S&G and the number of S&Gs and socio-demographic information of the participant like gender, age and driving experience (this forms the structural equation model). Since this experiment comprised unlabelled alternatives, the same frustration propensity relationship can be used for both the routes. The frustration propensity is then mapped to the observed frustration ratings ($\boldsymbol{I}$) which were given by the participants in the experiment (this represents the measurement model). The term $\boldsymbol{\xi}$ represents the structural error in measuring the true frustration propensity. $\boldsymbol{\eta}$ represents the measurement error and $\boldsymbol{\varepsilon}$ signifies the idiosyncratic error term in the discrete choice model. $\boldsymbol{A}$ and $\boldsymbol{B}$ are the parameters which are estimated from the SEM. We discuss the model formulation of the SEM, followed by model identification and estimation in the following subsections. The same notations used in figure 5.6 are used to maintain uniformity. We follow the same formatting style, both in figure 5.6 and in the next subsections, while discussing matrix algebraic notations used in the model formulation, i.e. *scalar* quantities are written in italics, ***vectors*** in italics and bold face, and **matrices** in bold face.

5.7.1 Model formulation

Consider that participant $n \in N$ evaluates route $j \in J$ in choice scenario $t \in T$. The participant gives a frustration rating i_{njt} for each of the routes driven within a scenario and the route choice y_{nt} at the end of every scenario. As discussed earlier in section 5.3, the experiment comprised a total of three scenarios ($T = 3$) each involving two routes ($J = 2$). Thus, the participant provided a total of six indicator ratings (along with three route choices) in the entire experiment. We now present the mathematical formulation of each the two components of the SEM, which has been discussed in the paper by Bhat & Dubey (2014).

5.7.1.1 Latent variable structural equation model

The structural equation specifies a linear relationship between the frustration propensity f^*_{njt} (latent construct) and a set of explanatory variables. Equation 5.2 gives the expression (in scalar notations) for this linear relationship.

$$f^*_{njt} = \boldsymbol{a_1 x'_{njt}} + \boldsymbol{a_2 z'_n} + \xi_n \quad (5.2)$$

In this equation, $\boldsymbol{x_{njt}}$ is a vector of size $1 \times k$, where k is the number of route specific attributes (the time spent in S&G and the number of S&Gs, so $k = 2$ in this study as shown in figure 5.6). $\boldsymbol{z_n}$ is a $1 \times m$ vector of socio-demographic variables, where m is the number of covariates (gender, age and driving experience, so $m = 3$ in this study as shown in figure 5.6). $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ are the structural coefficients of size $1 \times k$ and $1 \times m$ respectively which are assumed to be generic parameters due to the unlabelled routes setting discussed earlier. ξ_n is the structural error term which do not have the subscripts j and t. While the structural error is the same across alternatives (j) due to the routes being unlabelled, it remains the same across scenarios (t) since it reflects the unobserved attitudes which do not change quickly for the participant over time (at least over the duration of this experiment). It is assumed that ξ_n follows a standard normal distribution since f^*_{njt} is a continuous latent variable in the range $(-\infty, \infty)$ (Bolduc et al., 2005). Equation 5.2 can be re-written using matrix notations as shown in equation 5.3.

$$\boldsymbol{F}^* = \mathbf{X}\boldsymbol{A}' + \boldsymbol{\xi} \tag{5.3}$$

$\boldsymbol{F}^* = \{f^*_{n11},\ f^*_{n21}, \ldots . f^*_{nJT}\}$ represents a $JT \times 1$ vector of frustration propensities for individual n. $\mathbf{X}$ is a $JT \times (k + m)$ matrix formed by concatenating (horizontally) k route specific attributes across JT observations and repeating m socio-demographic variables JT times for participant n. $\boldsymbol{A}$ is the vector of structural coefficients of size $1 \times (k + m)$ which is obtained upon concatenating (horizontally) $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$. $\boldsymbol{\xi}$ is formed by repeating the structural error term (ξ_n) JT times, and follows a univariate normal distribution: $\boldsymbol{\xi} \sim \mathrm{N}(0, \boldsymbol{\psi})$ where $\boldsymbol{\psi}$ is a correlation matrix which is parametrised to one for the case of a single latent construct (as in our study). Generally speaking, for the case of l latent constructs, $\boldsymbol{\xi}$ is distributed as a multivariate (l dimension) normal ($\boldsymbol{\xi} \sim \mathrm{MVN}_l\ (\boldsymbol{0}, \boldsymbol{\psi})$) where $\boldsymbol{\psi}$ is a $l \times l$ correlation matrix which captures the correlation among the different latent variables.

5.7.1.2 Latent variable measurement equation model

The frustration propensity (f^*_{njt}) thus obtained is used as an explanatory variable to model the observed frustration ratings given by the participant. Equation 5.4 gives the linear expression for the measurement model using scalar notations.

$$i^*_{njt} = \delta_j + bf^*_{njt} + \eta_{njt} \tag{5.4}$$

In this equation, i^*_{njt} is the latent variable signifying the overall frustration (a latent variable) experienced by participant n for alternative j in driving scenario t. δ_j is an alternate specific constant and b is the factor loading on the latent construct (f^*_{njt}). Like the structural coefficients, b is treated as a generic parameter due to the unlabelled routes setting followed in this experiment. η_{njt} is the measurement error term which follows a standard normal distribution. The error term generally requires normalisation for identification purpose (McKelvey & Zavoina, 1975). Since i^*_{njt} cannot be observed by the analyst, who can only observe the indicator rating (frustration rating i_{njt}) which is generally asked on an R-point Likert scale ($R = 5$ in this study). The realisation (or the probability of occurrence) of the ordered rating r can be expressed using equation 5.5.

$$\Pr(i_{njt} = r) = \Pr(\mu^j_{r-1} < i^*_{njt} < \mu^j_r) \tag{5.5}$$

In this equation, $\boldsymbol{\mu^j}$ is a $1 \times (R + 1)$ vector of threshold points within which the observed rating $r \in R$ falls. In other words, the observed rating r lies between the $(r - 1)^{\text{th}}$ and r^{th} elements of the threshold vector $\boldsymbol{\mu^j}$. The superscript j signifies a different threshold vector for every alternative (route in this study). The elements in each set are: $\boldsymbol{\mu^j} = \{-\infty, 0, e^{\alpha_{1j}}, e^{\alpha_{1j}+\alpha_{2j}}, \dots, e^{\alpha_{1j}+\alpha_{2j}\dots+\alpha_{(R-2)j}}, \infty\}$. The parameters $\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{(R-2)j}$ ($R - 2$ in number) are estimated in the model to compute $\boldsymbol{\mu^j}$. As the measurement errors are normally distributed, the expression on the right-hand side in equation 5.4 becomes a probit kernel (Train, 2009). The equations above can be re-written in matrix notation and shown in equations 5.6 and 5.7.

$$\boldsymbol{I}^* = \boldsymbol{\delta} + \boldsymbol{B}.\boldsymbol{F}^* + \boldsymbol{\eta} \tag{5.6}$$

$$\Pr(\boldsymbol{I}) = \Pr(\boldsymbol{\mu_{low}} < \boldsymbol{I}^* < \boldsymbol{\mu_{up}}) \tag{5.7}$$

In equation 5.6, $\boldsymbol{I}^* = \{i^*_{n11},\ i^*_{n21}, \dots .\, i^*_{nJT}\}$ represents a $JT \times 1$ vector of actual frustration (latent variable) experienced. $\boldsymbol{\delta}$ is a $JT \times 1$ vector made by T repetitions of δ_j, i.e. the alternate specific constant for each alternative. $\boldsymbol{B}$ is a $JT \times 1$ vector which is formed by repeating the generic factor loading (b) JT times. The dot (.) between B and $\boldsymbol{F}^*$ signifies element wise multiplication of the two quantities. $\boldsymbol{\eta} = \{\eta^*_{n11}, \eta^*_{n21}, \dots .\, \eta^*_{nJT}\}$ is a vector of measurement errors of size $JT \times 1$. $\boldsymbol{\eta}$ follows a multivariate (JT dimension) normal distribution such that: $\boldsymbol{\eta}$ ~ MVN_{JT} ($\boldsymbol{0}$,$\sum$) where $\sum$ is a correlation matrix of size $JT \times JT$. In equation 5.7, $\boldsymbol{I} = \{i_{n11},\ i_{n21}, \dots .\, i_{nJT}\}$ is a $JT \times 1$ vector of the observed ordered responses given by the participant. $\boldsymbol{\mu_{low}}$ and $\boldsymbol{\mu_{up}}$ are the lower and upper threshold vectors of size $JT \times 1$ each consisting of the lower and upper cut-off values from the set $\boldsymbol{\mu^j}$ for each element in $\boldsymbol{I}$. Given the distributional assumptions made on the measurement error, the joint probability of observing $\boldsymbol{I}$ also follows a multivariate normal distribution.

5.7.2 Model identification

Equation 5.8 can be obtained by substituting equation 5.3 in equations 5.6. Equation 5.9 gives the expression to evaluate the joint probability of observing the indicators ($\boldsymbol{I}^*$) which is a multivariate normal distribution of JT dimensions with the mean ($\ddot{\boldsymbol{\theta}}$) and covariance ($\ddot{\boldsymbol{\vartheta}}$) shown in equations 5.10 and 5.11 respectively. While $\ddot{\boldsymbol{\theta}}$ is of size $JT \times 1$, $\ddot{\boldsymbol{\vartheta}}$ is of size $JT \times JT$.

As an illustration, we show the calculation of the mean and variance for equation 5.8. The term $\boldsymbol{\delta} + \boldsymbol{B}.\boldsymbol{XA}'$ in the equation comprises all known quantities (attributes or parameter values) thus forming the mean, which is shown in equation 5.10. For evaluating the variance, we use the following properties from statistics:

- If v is a random variable and o a scalar then $Var(ov) = o^2 v$.

- Assuming O and V as vector notations, $Var(OV) = OVO'$.
- If V and N are two independent random numbers then $Var(V+N) = Var(V) + Var(N)$

The last two terms in equation 5.8, $\boldsymbol{B}.\boldsymbol{\xi} + \boldsymbol{\eta}$ are stochastic. Also, we know that the error terms $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are independent. Thus, we can calculate the variance of equation 5.8, which is shown in equation 5.11.

$$\boldsymbol{I}^* = \boldsymbol{\delta} + \boldsymbol{B}.\mathbf{X}\boldsymbol{A}' + \boldsymbol{B}.\boldsymbol{\xi} + \boldsymbol{\eta} \tag{5.8}$$

$$Pr(\boldsymbol{I}^*) = MVN_{JT}\,(\ddot{\boldsymbol{\theta}}, \ddot{\boldsymbol{\vartheta}}) \tag{5.9}$$

$$\ddot{\boldsymbol{\theta}} = [\boldsymbol{\delta} + \boldsymbol{B}.\mathbf{X}\boldsymbol{A}'] \tag{5.10}$$

$$\ddot{\boldsymbol{\vartheta}} = [\boldsymbol{B}\boldsymbol{\psi}\boldsymbol{B}' + \sum] \tag{5.11}$$

All the parameters to be estimated in $\ddot{\boldsymbol{\theta}}$ and $\ddot{\boldsymbol{\vartheta}}$ are identifiable by ensuring that the diagonal elements of $\sum$ which correspond to the ordinal variables are normalised to one. In other words, $\sum$ is normalised as an identity matrix. The value of $\boldsymbol{\psi}$ has been discussed above.

The SEM formulation presented above evaluates the joint probability of observing JT frustration ratings given by the participant in the experiment. As the participants were asked to drive on the two routes in quick succession within a scenario, we expected some dependency between the routes within and between scenarios. A Pearson Chi-Squared test between frustration ratings on left and right routes across three scenarios was conducted which also indicated a dependency between the two sets of ratings (significant at 95 percent confidence interval). Thus, we decided to jointly estimate the frustration ratings rather than treating them as independent. We now discuss the estimation procedure for the model.

5.7.3 Model estimation

The probability of jointly observing $\boldsymbol{I}$ indicator ratings responses requires the evaluation of a JT dimensional multivariate normal (MVN) cumulative distribution function (cdf). It implies that in order to solve the MVN cdf, one has to solve a JT dimensional integral. A general procedure to solve such complex functions is through the Maximum Simulated Log-likelihood (MSL) approach (Greene & Hensher, 2010; Train, 2009). However, this widely used estimation procedure poses the following limitations due to the curse of dimensionality: 1) the MSL can become infeasible as the order of integration grows bigger, 2) the numerical solution methods can become quite time-consuming and might also lead to convergence issues during estimation (Bhat et al., 2010). A recent inference approach, the Composite Marginal Log-likelihood (CML), provides a simulation free technique thus overcoming the shortcomings of the MSL. The method which is gradually becoming popular requires estimation of a pairwise likelihood function, called the marginal likelihood, which is easy to compute due to a lower dimensionality. This marginal likelihood function provides a good approximation to the full likelihood function. Furthermore, the CML estimator is consistent and asymptotically normal and is a lot quicker in computation time when compared to the MSL approach. Readers can find a good description of the CML approach and its advantages over the MSL method in the paper by Paleti & Bhat (2013).

As shown in equation 5.9, the full likelihood function (for an individual) is a JT dimensional integral, which is equivalent to 6 in our experiment. According to the CML technique, this multidimensional integral can be broken down into independently observing different duplets of the observed indicator ratings. In other words, we defined the paired likelihood function of simultaneously observing two frustration ratings. This resulted in a total of $\binom{JT}{2}$ combinations to be evaluated. The resulting marginal likelihood function for individual n is thus shown in equation 5.12.

$$L_{CML}^{n}(\ddot{\boldsymbol{\theta}}, \ddot{\boldsymbol{\vartheta}}) = \prod_{i=1}^{JT-1} \prod_{j=i+1}^{JT} Pr\left(i_{ni} = r_{ni},\ i_{nj} = r_{nj}\right) \tag{5.12}$$

$$Ln\ L(\ddot{\boldsymbol{\theta}}, \ddot{\boldsymbol{\vartheta}}) = \sum_{n=1}^{N} \log L_{CML}^{n}(\ddot{\boldsymbol{\theta}}, \ddot{\boldsymbol{\vartheta}}) \tag{5.13}$$

Equation 5.12 represents the pairwise probability of observing two indicator (frustration) ratings i_{ni} and i_{nj} simultaneously. The first term, r_{ni} and r_{nj} are the ratings on a Likert scale given on i^{th} and j^{th} occasion respectively by participant n. Equation 5.13 gives the log-likelihood function by summing the logarithm of equation 5.12 across the participants. The procedure to determine the parameters $\ddot{\boldsymbol{\theta}}$ and $\ddot{\boldsymbol{\vartheta}}$ is discussed below.

To estimate equation 5.12, we need to set up the following vector. Create a selection matrix $\mathbf{Q}$ of size $2 \times JT$. Introduce the value *1* on the i^{th} and j^{th} columns of the first and second rows respectively, while the other elements are zero. Equation 5.12 represents multiplication of bivariate normal cdfs, each of which can be estimated using equation 5.14. The first two terms in the bivariate cdf on the left-hand side represent the mean value and the third term is the correlation. Equations 5.15 and 5.16 give their expressions. The threshold vector $\boldsymbol{\alpha_{low}} = \mathbf{Q}\boldsymbol{\mu_{low}}$ and $\boldsymbol{\alpha_{up}} = \mathbf{Q}\boldsymbol{\mu_{up}}$ is a column vector of size 2×1 representing the first and second rows for the ratings i and j respectively. Similarly, the mean value is selected as $\boldsymbol{\beta} = \mathbf{Q}\ddot{\boldsymbol{\theta}}$ and covariance $\boldsymbol{\omega} = \mathbf{Q}\ddot{\boldsymbol{\vartheta}}\mathbf{Q}'$ which are of sizes 2×1 and 2×2 respectively. $\boldsymbol{\omega_{11}}$ and $\boldsymbol{\omega_{22}}$ are the diagonal elements and the superscript on $\boldsymbol{\alpha_{low}}$, $\boldsymbol{\alpha_{up}}$ and $\boldsymbol{\beta}$ represent the indicator rating.

$$Pr(i_{ni} = r_{ni},\ i_{nj} = r_{nj}) = \begin{bmatrix} \Phi(\boldsymbol{\kappa_{up}^{i}}, \boldsymbol{\kappa_{up}^{j}}, \zeta_{ij}) - \Phi(\boldsymbol{\kappa_{up}^{i}}, \boldsymbol{\kappa_{low}^{j}}, \zeta_{ij}) \\ - \Phi(\boldsymbol{\kappa_{low}^{i}}, \boldsymbol{\kappa_{up}^{j}}, \zeta_{ij}) + \Phi(\boldsymbol{\kappa_{low}^{i}}, \boldsymbol{\kappa_{up}^{j}}, \zeta_{ij}) \end{bmatrix} \tag{5.14}$$

$$\begin{aligned} \boldsymbol{\kappa_{low}^{i}} &= \frac{\alpha_{low}^{i} - \beta^{i}}{sqrt(\boldsymbol{\omega_{11}})} \qquad \boldsymbol{\kappa_{low}^{j}} = \frac{\alpha_{low}^{j} - \beta^{j}}{sqrt(\boldsymbol{\omega_{22}})} \\ \boldsymbol{\kappa_{low}^{i}} &= \frac{\alpha_{low}^{i} - \beta^{i}}{sqrt(\boldsymbol{\omega_{11}})} \qquad \boldsymbol{\kappa_{low}^{j}} = \frac{\alpha_{low}^{j} - \beta^{j}}{sqrt(\boldsymbol{\omega_{22}})} \end{aligned} \tag{5.15}$$

$$\zeta_{ij} = \frac{\boldsymbol{\omega_{21}}}{sqrt(\boldsymbol{\omega_{11}})\ .\ sqrt(\boldsymbol{\omega_{22}})} \tag{5.16}$$

The model estimation routine was coded in Matlab, picking the bivariate cdf function from Professor Alen Genz's webpage (Genz, 2017). The log-likelihood function given in equation 5.13 was solved using the BFGS algorithm and the standard errors were calculated by taking the inverse of the simulated Hessian matrix upon convergence. The Matlab code is available through the weblink given in appendix B of this thesis.

5.7.4 Model results

For model estimation, the variable age was dichotomized into young (up to 40 years) and older age groups (Erikson & Erikson, 1998). Similarly, the variable driving experience was dichotomised at a threshold of 8 years since it gave a near equal representation of both the groups, i.e. less (up to 8 years) and more experienced drivers. The two dichotomised variables were found to be correlated with one another. Thus, an interaction variable between the two age and experience groups was constructed to be used as an explanatory variable in the latent variable structural equation model. Of the four resulting categories of the interaction variable, the category old (above 40 years) and less experienced (up to 8 years) did not have any observation which was intuitive. The category young and more experienced was kept as the base and the other two categories were estimated.

Table 5.3 shows the model parameters upon convergence. The parameter for the time spent in S&G and the number of S&Gs is highly significant (at 99 percent confidence interval) and bears a positive sign. In other words, the two attributes positively influence the frustration propensity of drivers. This finding indicates the validity of the hypothesis proposed in this study. The interpretation of the two parameters is as follows: While the former implies that a 1 minute increase in the time spent in S&G traffic increases the frustration propensity by 0.9153 on average, the latter increases the propensity by 0.0329, on average, with every additional repetition of S&G. This positive association is consistent with the previous literature which identified traffic congestion (eventually resulting in S&G traffic) as one of the factors instigating driver frustration (Jovanović et al., 2011; Lazarus, 1966).

Gender is found to have a significant effect on the frustration propensity, where females experience a higher frustration propensity than males. Hauber (1980) too found a similar thing, where young females reported higher stress levels than older males. We checked our dataset and found around 60 percent of the females as young (under 40 years). Thus, our finding is consistent with the previous literature.

Table 5.3: Results from the SEM estimation

Parameter	**Estimate**	
Structural Model (A)		
Time in S&G	0.9153 ***	
Number of S&Gs	0.0329 ***	
Male	-0.4442 ***	
Young & Less Exp.	0.7378 ***	
Young & More Exp.	*Base*	
Old & More Exp.	0.6729 ***	
Measurement Model (B)		
	Left Route	Right Route
Factor Loading	0.9538 ***	0.9538 ***
Constant	-2.2523 ***	-2.0149 ***
Cut-off Points		
μ_1	0.4704 ***	0.2472 ***
μ_2	0.2023 ***	0.1886 ***
μ_3	0.3829 ***	0.4756 ***
Model Fit		
LL (converged)	-3813.16	
Adjusted Rho-squared	0.1227	
Number of observations	594	

*** significant at 99%

The table also shows a positive and significant effect for the interaction variable category young and less experienced drivers (0.7378). A positive sign implies that drivers belonging to this group are expected to exhibit a higher frustration propensity when compared to the base category (young and more experienced). This can be explained as follows: As people in this group are relatively new to driving, they need to be extra careful while driving in S&G traffic, which possibly leads to a higher frustration propensity value. Similarly, the participants belonging to the older and more experienced group also show a higher frustration propensity (0.6729) than the base category. As the people belonging to this segment are usually at the peak of their career, they generally have a high value of time. Thus, they are more likely to get frustrated when driving to work in S&G traffic, which possibly explains the positive effect of this group towards the frustration propensity.

For the latent variable measurement model, the loading for frustration propensity (0.9538) is positive and highly significant. In other words, people with higher propensity value are more likely to give a higher frustration rating which is intuitive. Different constants and threshold points, which are highly significant as well, are estimated for both the routes. The threshold limits of the frustration propensity for the left route can be evaluated as $\{0, e^{0.4704}, e^{0.4704+0.2023}, e^{0.4704+0.2023+0.3829}\}$. Similarly the threshold points of the frustration propensity for the right route are $\{0, e^{0.2472}, e^{0.2472+0.1886}, e^{0.2472+0.1886+0.4756}\}$.

Several other specifications of the SEM were also tested. One of the models we tried also included travel time in the structural part. However, its parameter turned out to be insignificant (even at 80 percent confidence interval) upon model convergence. A few possible reasons behind that could be: 1) due to less variability in the attribute owing to a small sample, and 2) the travel time difference might have been perceived as too less by the participants. Thus, we dropped the travel time attribute from the model reported in table 5.3. The table also gives the overall goodness of fit of the selected SEM specification. The chosen model had a decent goodness of fit statistics and also conveyed meaningful interpretations. The results and goodness of fit measures of the alternate specifications tested are provided in appendix D (table D.5) of this thesis.

We used the parameter estimates from table 5.3 to evaluate the frustration propensity score for the available dataset. The mean propensity score was found to be 3.46 with a standard deviation of 0.59. The range of the score was between 1.17 and 5.02. We categorised the score into bins of width 1and used them to check for an association with the observed route choices across all three scenarios using the Pearson Chi-

Squared test. The result from the test was found to be statistically significant at 99 percent confidence interval which suggested a dependency between frustration propensity and the route choice of the participants.

5.8 Discussion

The study presented in this chapter was conducted with the prime objective of determining the effect of the time spent in S&G and the number of S&Gs on the psychological factors, like the levels of frustration of drivers. Previous studies in the domain of driver psychology have identified traffic congestion as one of the triggers for driver stress, which eventually gets transformed into aggression (Hennessy & Wiesenthal, 1999; Jovanović et al., 2011). As stop-&-go (S&G) traffic generally occurs in the congested regime, we expected a positive correlation between the level of frustration and the attributes such as the time spent in S&G and the number of S&Gs. Another reason behind conducting this study was to account for the measurement bias limitation which persisted since the last two studies presented in chapters 3 and 4 of this thesis.

Thus, we conducted this study using a driving simulator which gave us the opportunity to closely study the participant's driving style and route choice behaviour. The driving simulator setup not only helped in reducing the measurement bias towards the time spent and the number of S&Gs experienced, but also in observing how the level of frustration builds up while travelling through S&G traffic. The designed experiment presented a total of three scenarios to each participant, which comprised two routes within every scenario. The two routes had a similar travel distance and travel time, but different congested traffic conditions (characterised by S&G traffic). The participants were asked to drive on each route within a scenario and rate (on a 5-point Likert scale) the level of frustration experienced at the end of the drive. Similarly, at the end of every scenario, the participants were asked to select the most preferred route for travel from the two available alternatives. This way, every participant provided six frustration ratings and three route choices in all during the experiment.

The descriptive statistics of the collected data showed a good representation of staff and students, who generally depict distinct socio-demographics and driving behaviour, in the effective sample. It was observed that a majority of the participants reported a lower frustration rating for the route (left route precisely) with a slightly higher travel time but fewer S&Gs. Additionally, it was observed that the same route was preferred for travel by the participants across all three scenarios presented to them. This observation showed the level of frustration could eventually influence the route choice behaviour of drivers.

The results from the SEM showed that the attributes like the time spent in S&G and the number of S&Gs had a positive effect on the frustration propensity (a latent variable). Although the effect of the time spent in S&G (0.9153) was much higher than the effect for the number of S&Gs (0.0329), it was still useful to include this attribute to account for the case when there are more cycles of S&G within a short period of time. Thus, the proposed research hypothesis for this study was found to be valid. Additionally, the Pearson Chi-Squared test between the computed frustration propensity score (on a nominal scale) and the observed route choice indicated a dependency between the two. In other words, it can be concluded that the built up frustration also influenced the route choice behaviour of the participants. This is an interesting finding as it indicates that the number of S&Gs indirectly influences route choice (via the frustration propensity). We also conducted a simple logistic regression to check for the direct effect of S&G traffic characteristics on route choice for the current dataset. However, the parameters turned out to be highly insignificant. Thus, the current finding adds a new dimension to the earlier outcomes by even accounting for the role of psychological aspects in the route choice behaviour. The finding is more grounded in the existing literature on driver stress which also suggests that drivers become more agitated while driving in heavy congested (S&G traffic) conditions (Jovanović et al., 2011; Lazarus 1966).

The study presented in this chapter brought out some interesting findings which add to the existing knowledge base on driver stress in congested (S&G) traffic conditions. However, despite our attempts to conduct a methodologically sound experiment, this study has a limitation. We could manage to collect a small sample for analysis which mainly comprised university staff and students who do not offer much variability in the available data. We tried different channels to recruit the participants from outside the university to make the sample more homogeneous and representative of the car driving population in Sydney, but were not successful in getting many responses.

The results from this study can have few potential applications in the real world. The results from this study can be useful in managing traffic operations in a way such that S&G conditions and the resulting driver stress levels can be kept under check. We discuss the policy implications of the results in chapter 6 of this thesis.

CHAPTER 6

CONCLUSIONS, LIMITATIONS & FUTURE DIRECTIONS

Traffic congestion problem is an overgrowing concern among transport planners worldwide who are striving hard to keep this menace under check. It not only leads to an economic loss in the form of person hours wasted in congestion, but also poses a serious threat to the surrounding environment and the road user safety. Thus, in order to propose congestion mitigation measures, it is crucial to first understand the dynamics of traffic congestion and its impact on the route preferences of car drivers. This thesis looked at studying the role of stop-&-go (S&G) traffic, a phenomenon that is common in congested traffic conditions, in the intricate route choice behaviour of car drivers. Particularly, the aim of this thesis was to test the research hypothesis: *an increase in the number of S&Gs on a route increases its disutility for a driver*. Chapter 1 of this thesis presented the motivation (through an example) which laid the foundation to conceptualise the proposed research hypothesis.

Chapter 2 of this thesis presented the state-of-the-art in the three domains of knowledge where the characteristics and dynamics of S&G traffic have been extensively studied. It first reviewed the literature from traffic engineering which either deals with mathematical modelling of S&G waves (theoretical studies) or identifying their characteristics such as formation, propagation and dissipation using different techniques (empirical studies). However, while these works were able to quantify the occurrence of S&G waves, they did not study the impact of these conditions on the route choice of drivers. The chapter next presented research works on driver behaviour where the time spent in S&G was used as one of the attributes to express disutility of a route. However, as we showed through the example presented in chapter 1, it is the number of S&Gs experienced which is also expected to have an adverse effect on disutility of a route along with the time spent in such conditions. Moreover, the methodologies adopted for experiment design and data analysis in these works are also known to have a few limitations. Lastly, the chapter presented the literature on Adaptive Cruise Control (ACC) algorithms which used the measures of vehicle dynamics such as the time headway, acceleration (deceleration) and jerk for its calibration. However, these measures were generally assigned a very high value, from a safety perspective, which could not be used for quantifying the occurrences of S&G waves in a real-world scenario. Thus, this chapter was able to identify the research gap in the existing body of knowledge, i.e. determining the impact of the number of S&Gs experienced on the route choice behaviour of car drivers.

Chapter 3 presented a proof of concept study which was conducted as the first step towards testing the validity of the proposed research hypothesis. A Stated Choice (SC) experiment was conducted on a sample of respondents comprising university staff and students. The survey instrument first made the participants aware of S&G traffic which was followed by 10 SC tasks to understand their route choice behaviour. The collected dataset was analysed using an econometric approach, called the Random Parameter Error Component Logit (RPECL) model, to account for the preference heterogeneity across individuals and the serial correlation arising due to the panel nature of the dataset. Results from the analysis showed a negative sign on the coefficient for the number of S&Gs attribute which indicated an increase in disutility of a route as the attribute value increases. This finding showed that the hypothesis we set at the beginning of this study was valid. This study also found that the participants were willing to travel for 16 extra seconds to avoid one additional occurrence of S&G. It is an interesting finding which indicates that drivers are not always travel time minimisers, but rather assign some weight to discomfort (which can be explained in terms of the number of S&Gs) as well. However, this study being the first step towards thoroughly testing the research hypothesis had the following limitations: 1) a restricted sample of university staff and students which do not reflect the overall demographics and travel characteristics of the wider population, 2) the sample even included occasional drivers (drove less than thrice a week to work) who might have biased perceptions towards S&G traffic due to a less frequent exposure to such conditions, and 3) the RPECL model showed preference heterogeneity among the participants, but was not able to pinpoint the cause behind this variation.

The experiment presented in Chapter 4 of this thesis extended the proof of concept study by overcoming the limitations discussed above. It circulated another SC survey (similar to the one used in the earlier study) to the individuals residing in Sydney or its neighbouring regions who regularly drove to work. The empirical analysis of the collected dataset indicated three types of route choice behaviour upon visual inspection of the collected SC data. The Latent Class Choice Model (LCCM) used for the quantitative data analysis also confirmed the existence of three groups of people, i.e. 1) *Travel time minimisers*, 2) *Calm cost minimisers*, and 3) *Restless cost minimisers*. Nearly 78 percent of the total participants, belonging to the first two segments, were found to have a significant disutility towards the number of S&Gs attribute. Particularly, the *restless cost minimisers* group, which generally comprised young participants, had the highest disutility towards the number of S&Gs and were willing to travel for additional 4 minutes to avoid one extra occurrence of S&G. This was in line with the previous studies which also found young drivers to get easily perturbed in congested (S&G) traffic (Hauber, 1980; Wiesenthal et al., 2000). Interestingly, 22 percent of the participants, belonging to the *calm cost minimisers* group, were found to be indifferent to the number of S&Gs. The overall trade-off value between travel time and the number of S&Gs across the three segments was found to be 1.29 minutes per stop which was higher than what was found in the previous (proof of concept) study. However, this study, like the previous study, considered the number of S&Gs experienced, revealed by the participants at the start, as the true measure which might be susceptible to a measurement bias. Moreover, we also came across the literature which suggested that drivers tend to experience higher levels of frustration in congested traffic (Hennessy & Wiesenthal, 1999; Jovanović et al., 2011) which gave us the push to further extend our exploration.

Chapter 5 extended the research exploration to study the effect of the number of S&Gs and the time spent in S&G on psychological factors, like the level of frustration in drivers. The chapter also aimed to investigate the dependency between the inbuilt frustration and the resulting route choice of drivers. A driving simulator experiment was conducted which not only minimised the measurement bias associated with the attributes of S&G traffic, but also facilitated the study of driver stress under such conditions. The experiment asked the participants to drive on two different routes, within each of the three scenarios, to study their level of frustration and route choice. The descriptive statistics of the collected dataset showed that a majority of the participants preferred the route with a slightly higher travel time but fewer S&Gs. A Structural Equation Model (SEM) was used for the data analysis which showed that every minute spent in S&G and every additional occurrence of S&G on average increased the frustration propensity by 0.9153 and 0.0329 respectively. This finding not only validated the proposed hypothesis for this study, but also provided quantification between S&G traffic characteristics and the frustration propensity (a latent variable) of the participants. Furthermore, results from the Pearson Chi-Squared test, conducted on the frustration propensity scores and route choices of the participants, revealed a dependency between the two. Thus, it can be said that S&G traffic characteristics (including the number of S&Gs) have an indirect effect on route choice. This observation was in contrast to the findings from the previous two studies (presented earlier in chapters 3 and 4) which indicated a direct effect of S&G attributes on route choice. The obtained results were consistent with the previous literature on driver psychology which also hinted an indirect effect of traffic congestion on route choice (Hennessy & Wiesenthal, 1999; (Jovanović et al., 2011). This study added a new dimension in explaining the intricate route choice behaviour of drivers, i.e. the evaluating the impact of psychological constructs. However, this study has a limited scope as the collected dataset did not represent the travel and socio-demographic characteristics of the overall car driving population in Sydney.

In conclusion, the three studies presented in this thesis corroborated the validity of the proposed hypothesis which was *an increase in the number of S&Gs on a route increased its disutility for a driver.* While the findings from chapters 3 and 4 indicated a direct effect of the number of S&Gs on disutility of a route, results from chapter 5 pointed an indirect effect between the two quantities. The results from these studies, even though contrasting, still confirmed that the number of S&Gs also had an impact (association) on the route choice behaviour of drivers along with the other attributes that have been used by other researchers until now. Moreover, chapters 3 and 4 found a non-zero trade-off value between travel time and the number of S&Gs. This value shows the willingness of drivers to reduce discomfort on the travelled route and not just purely minimising travel time. Additionally, results from chapter 5 confirmed and quantified the fact that S&G traffic makes drivers more frustrated which is also the precursor to incidents like driver aggression and road rage. The quantification of these relationships can have a few policy implications which are discussed in the following section.

6.1 Policy Implications of the Findings

Each of the experiments conducted in this thesis brought out a few interesting findings. The results from these experiments can potentially inform policy decisions and measures aimed at mitigating traffic congestion and the resulting S&G traffic. Some policy implications of the results obtained from this thesis are summarised below:

6.1.1 Modifying the toll pricing strategy

Tolls are additional costs which are generally applied to curb the greedy behaviour of road users. Examples such as Braess Paradox show that the user equilibrium assignment (where users aim to minimise their travel time only) can actually put the network in a worse-off situation (Sheffi, 1984). Thus, tolls are one of the ways of shifting the network equilibrium more towards system optimum rather than user equilibrium.

The toll pricing calculation on a link is generally done considering the travel time loss caused to all other road users due to the addition of a new user. In other words, the toll price levied on the new user is equivalent to the monetary value of the delay caused to other users due to its entry. However, the price evaluation does not directly take into account the discomfort caused due to the addition of new vehicles, which often stimulates S&G traffic. The results from the experiments conducted in this thesis indicate that drivers generally consider both travel time and discomfort while selecting a route for travel. Thus, the toll calculation function should account for both delay (expressed as additional travel time) and discomfort (expressed as the number of S&Gs experienced). Equation 6.1 gives the expression for the modified toll function. In this equation, toll-cost is in dollars, TT Delay signifies the travel time delay in hours and S&G represents the average number of S&Gs experienced. The calibration parameters ω and ϑ represent the value of time (\$/h) and the WTP measure between the number of S&Gs and cost (\$/S&G). These calibration parameters can be derived from table 4.13 by taking the weighted mean (for simplicity) of the WTP measures of interest, i.e. AU \$17.80 per hour ($\omega = 0.47 \times 33.43 + 0.22 \times 2.9 + 0.31 \times 4.74$) and AU \$0.12 per stop($\vartheta = 0.47 \times 0.05 + 0.22 \times 0 + 0.31 \times 0.32$) . Alternatively, a more effective way will be to have separate toll-cost functions for each of three segments identified in the table specifically catering to the WTP of users in that segment. However, its implementation will involve more complexity and computational effort to apply class specific toll-costs to users.

$$Toll - Cost\ (\$) = \ \omega * TT\ Delay + \ \vartheta * No.of\ S\&Gs \qquad (6.1)$$

The modified toll cost function in equation 6.1 assigns some weight on discomfort which can be expressed in terms of the number of S&Gs experienced on average on a link. Implementation of this toll value would potentially lead to a smoother flow of traffic by reducing the level of congestion and discomfort caused by S&G traffic.

6.1.2 Updating existing transportation models

The transportation models in practice are a handy tool for transport planners to ascertain the response of a transportation network to any stimulus given in the form of a policy enforcement or an infrastructural change. An important component of these models is the traffic assignment module which redirects vehicles on different paths (routes) in the network. The traffic assignment procedure first determines the set of k-shortest paths between an origin destination pair and then distributes the demand (number of cars users for example) among the paths until equilibrium conditions are obtained. The shortest paths are evaluated using the generalised cost information of every path (which is made up of links (arcs)) which is expressed in terms of travel time (for the User Equilibrium (UE) case) or travel time and cost (for the System Optimal (SO) case). However, as discussed in the subsection above, the contribution of discomfort (in the form of the number of S&Gs experienced) in the generalised cost function is also noteworthy. In other words, the revised generalised link cost function as shown in equation 6.2 should be utilised as an input into the UE formulation. In this equation, μ represents the trade-off value between travel time and the number of S&Gs. Its value (weighted mean) was obtained as 1.29 minutes per stop ($\mu = 0.47 \times 0.09 + 0.22 \times 0.02 + 0.31 \times 4.01$) from table 4.13 presented in chapter 4 of this thesis.

$$Generalised\ Cost\ (mins) = TT + \mu * No.of\ S\&Gs \tag{6.2}$$

The resulting UE link flows (using the revised generalised link cost function) are not only expected to provide a better representation of the evolution and propagation of traffic congestion in a road network, but also result in smoothening of traffic flow resulting in fewer S&Gs. We also conducted an exploratory work to compare the vehicle assignment for the two cost functions, default (travel time only) and modified (travel time and the number of S&Gs), in a microscopic simulation study. Appendix F of this thesis presents the analysis from this study. The initial results looked promising where the proportion of vehicles assigned to a route at equilibrium was found to be different across the two functions. However, it is noteworthy to state that this analysis is still in its infancy and the challenges with regard to its application on a wider scale need to be addressed carefully. Some immediate caveats that we see upfront are: 1) quantifying the occurrences of S&G waves in a macroscopic or mesoscopic network loading framework which would make the problem computationally tractable for a large scale network instance (Chiu et al., 2011), and 2) accounting for equilibrium conditions with link interactions where S&G conditions on the downstream link might spill back onto the upstream link (Sheffi, 1984).

6.1.3 Segment specific schemes: Introduction of Autonomous Vehicles

An Autonomous Vehicle (AV) is a self-sufficient system that reduces the extent of human (driver) intervention required during driving operations. The society of automotive engineers (SAE) classifies vehicles into six levels based on the degree of automation they offer (SAE Mobilus, 2016). The classification ranges between a manually driven car (Level 0) to a fully automatic self-driving car (Level 5). The state-of-the-art is presently at Level 1 (Subaru and Mazda) and 2 (Mercedes Benz E-Class) and auto majors such as General Motors, Daimler and Ford plan to manufacture a fleet of Level 4 and 5 vehicles in the near future. AVs are increasingly grabbing the attention and curiosity of the road users and planning bodies across the globe as their introduction on roads is expected to invoke a major paradigm shift in the way people travel.

AVs are capable of sensing the environment and navigate without driver intervention. This facilitates significant reduction, or even elimination, of asymmetries in driving behaviour, due to the human factor, which is often the trigger for S&G waves (Laval & Leclercq, 2010b). Thus, having some proportion of AVs in the fleet of on-road vehicles can potentially lead to channelized traffic flow through a great reduction in traffic congestion and the consequent S&G traffic (Litman, 2015). One of the policy measures, as discussed earlier in chapter 4, is encouraging the use of AVs among the *travel time minimisers* user group having a high value of time. It can reap the following benefits: 1) they can still be productive while on the go, and 2) having some AVs on road will in-turn benefit other road users as well. Policy makers can potentially encourage the uptake of AVs among this user group through the following schemes: 1) Relaxing the taxes and duties involved in importing an AV into Australia, as the AV manufacturing in Australia is still growing when compared to the US, and 2) providing adept infrastructure to support AVs movement on roads (for example, V2V and V2I communication). However, it is worth mentioning at this point that although promising, the idea along with the potential schemes requires a thorough evaluation on the grounds of acceptability and safety of AVs.

The schemes discussed above can reap two potential benefits once implemented in real-life. Firstly, it would help in reducing the magnitude of traffic congestion and the resulting S&G traffic through smoother traffic conditions, which is vital for the good economic health of a city. Secondly, as shown in chapter 5 of this thesis, keeping S&G traffic under control would also encourage calm driving conditions which would considerably reduce incidents pertaining to road rage due to driver aggression.

6.2 Future Research Directions

As the saying goes that *research is a never-ending process*, we have identified a few directions to further extend the research problem, which we presented in this thesis, in future endeavours. The future works are discussed below:

Firstly, future works will aim at further enriching the data analysis presented in chapter 5 of this thesis. The current analysis quantifies the relationship between S&G traffic characteristics and the level of frustration and also shows an association between the latter and route choice of the participants. Thus, a future task will be to fit an Integrated Choice and Latent Variable (ICLV) model on the available dataset to even gauge the exact effect size of the inbuilt frustration propensity (a latent variable) on route choice. An ICLV model typically integrates the latent psychological constructs into the discrete choice modelling framework which not only facilitates a more detailed depiction of the true choice process, but also enhances the predictive power of the model (Bolduc & Alvarez-Daziano, 2010; Temme et al., 2008)

Secondly, future research will focus on the naturalistic driving data to study the relationship between the attributes of S&G traffic, such as the time spent in S&G and the number of S&Gs, and the route choice behaviour of drivers. As drivers generally consider travel time as a disutility (Cirillo & Axhausen, 2006), the naturalistic driving data will enable studying their actual behaviour when subjected to S&G traffic. We imagine that the vehicle trajectory data along with the route preferences of drivers would be available for analysis. Thus, a sub-problem of this work will be to quantify the number of S&Gs experienced from the trajectory data. As discussed in chapter 2 of this thesis, Wavelet Transformation (WT) is an effective technique which has been used by researchers to locate the initiation, propagation and dissipation of an S&G wave in space and time (Zheng et al., 2011a,b; Zheng & Washington, 2012). We conducted some initial work on this sub-problem, which is presented in appendix A of this thesis, where we used WT to find the time instant of the formation and dissipation of an S&G wave for a sample vehicle trajectory.

Thirdly, the congestion mitigation policies discussed above will be tested in a simulation environment to build a much stronger case before the planners on its merits. Several aspects of this step such as the scale of the network, amount of data required need to be accessed to test these policies.

Additionally, some other interesting explorations on the line of the research problem presented in this thesis could be:

- Studying the effect of the number of S&Gs on route choice for non-work trips
- Effect of the number of S&Gs on public transit users
- Comparing the route choice behaviour of drivers under S&G traffic across different geographies

REFERENCES

Abou-Zeid, M., Kaysi, I., Al-Naghi, H., 2011. Measuring Aggressive Driving Behavior Using a Driving Simulator: An Exploratory Study. Presented at the 3rd International Conference on Road Safety and Simulation, Purdue University, Transportation Research Board.

ABS, 2015. 6523.0 - Household Income and Wealth, Australia, 2013-14 [WWW Document]. URL http://www.abs.gov.au/ausstats/abs@.nsf/mf/6523.0 (accessed 2.9.16).

Addison, P.S., 2002. The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance. CRC Press.

Adeli, H., Ghosh-Dastidar, S., 2004. Mesoscopic-Wavelet Freeway Work Zone Flow and Congestion Feature Extraction Model. J. Transp. Eng. 130, 94–103. doi:10.1061/(ASCE)0733-947X(2004)130:1(94)

Adeli, H., Samant, A., 2000. An Adaptive Conjugate Gradient Neural Network–Wavelet Model for Traffic Incident Detection. Comput.-Aided Civ. Infrastruct. Eng. 15, 251–260. doi:10.1111/0885-9507.00189

Ahn, S., Cassidy, M.J., 2007. Freeway Traffic Oscillations and Vehicle Lane-Change Maneuvers. Presented at the Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17.

Ahn, S., Cassidy, M.J., Laval, J., 2004. Verification of a Simplified Car-Following Theory. Transp. Res. Part B Methodol. 38, 431–440. doi:10.1016/S0191-2615(03)00074-2

Alexander, J., Barham, P., Black, I., 2002. Factors Influencing the Probability of an Incident at a Junction: Results from an Interactive Driving Simulator. Accid. Anal. Prev. 34, 779–792.

Anable, J., 2005. "complacent Car Addicts" or "aspiring Environmentalists"? Identifying Travel Behaviour Segments Using Attitude Theory. Transp. Policy 12, 65–78. doi:10.1016/j.tranpol.2004.11.004

Antoine, A.L.C., Molenaar, K.R., 2016. Latent Class Analysis for Highway Design and Construction Project Categorization. Procedia Eng., ICSDEC 2016 – Integrating Data Science, Construction and Sustainability 145, 1314–1321. doi:10.1016/j.proeng.2016.04.169

Apparies, R.J., Riniolo, T.C., Porges, S.W., 1998. A psychophysiological investigation of the effects of driving longer-combination vehicles. Ergonomics 41, 581–592. doi:10.1080/001401398186766

Infrastructure Australia, 2015. Australian Infrastructure Audit: Our Infrastructure Challenges (No. Volume 1). Infrastructure Australia, Sydney, Australia.

Bahamonde-Birke, F.J., Kunert, U., Link, H., Ortúzar, J. de D., 2017. About Attitudes and Perceptions: Finding the Proper Way to Consider Latent Variables in Discrete Choice Models. Transportation 44, 475–493. doi:10.1007/s11116-015-9663-5

Bamberg, S., Schmidt, P., 2001. Theory-Driven Subgroup-Specific Evaluation of an Intervention to Reduce Private Car Use. J. Appl. Soc. Psychol. 31, 1300–1329. doi:10.1111/j.1559-1816.2001.tb02675.x

Bando, M., Hasebe, K., Nakanishi, K., Nakayama, A., 1998. Analysis of Optimal Velocity Model with Explicit Delay. Phys. Rev. E 58, 5429–5435. doi:10.1103/PhysRevE.58.5429

Bando, M., Hasebe, K., Nakayama, A., Shibata, A., Sugiyama, Y., 1995. Dynamical Model of Traffic Congestion and Numerical Simulation. Phys. Rev. E 51, 1035–1042. doi:10.1103/PhysRevE.51.1035

Beede, K.E., Kass, S.J., 2006. Engrossed in Conversation: The Impact of Cell Phones on Simulated Driving Performance. Accid. Anal. Prev. 38, 415–421. doi:10.1016/j.aap.2005.10.015

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of Choice and Latent Variable Models, in: Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges. pp. 431–470.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press.

Benz, T., Christen, F., Lerner, G., Schulze, M., Vollmer, D., 2003. Traffic Effects of Driver Assistance Systems – The Approach within INVENT.

Bertini, R.L., Leal, M.T., 2005. Empirical Study of Traffic Features at a Freeway Lane Drop. J. Transp. Eng. 131, 397–407. doi:10.1061/(ASCE)0733-947X(2005)131:6(397)

Bhat, C.R., 2001. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. Transp. Res. Part B Methodol. 35, 677–693. doi:10.1016/S0191-2615(00)00014-X

Bhat, C.R., 1995. A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. Transp. Res. Part B Methodol. 29, 471–483. doi:10.1016/0191-2615(95)00015-6

Bhat, C.R., Castelar, S., 2002. A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area. Transp. Res. Part B Methodol. 36, 593–616. doi:10.1016/S0191-2615(01)00020-0

Bhat, C.R., Dubey, S.K., 2014. A New Estimation Approach to Integrate Latent Psychological Constructs in Choice Modeling. Transp. Res. Part B Methodol. 67, 68–85. doi:10.1016/j.trb.2014.04.011

Bhat, C.R., Ferdous, N., Varin, C., 2010. A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered-Response Model, in: Maximum Simulated Likelihood Methods and Applications, Advances in Econometrics. Emerald Group Publishing Limited, pp. 65–106.

Bhat, C.R., Gossen, R., 2004. A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice. Transp. Res. Part B Methodol. 38, 767–787. doi:10.1016/j.trb.2003.10.003

BITRE, 2015. Traffic and Congestion Cost Trends for Australian Capital Cities. Department of Infrastructure and Regional Development, Canberra, Australia.

BITRE, 2014. Trends: Infrastructure and Transport to 2030. Department of Infrastructure and Regional Development, Canberra, Australia.

BITRE, 2011. Public Road-Related Expenditure and Revenue in Australia. Department of Infrastructure and Transport, Canberra, Australia.

Blanchard, D.C., Blanchard, R.J., 1984. Affect and Aggression: An Animal Model Applied to Human Behavior, in: Advances in the Study of Aggression. Elsevier, pp. 1–62.

Blanchard, E.B., Barton, K.A., Malta, L., 2000. Psychometric Properties of a Measure of Aggressive Driving: The Larson Driver’s Stress Profile. Psychol. Rep. 87, 881–892. doi:10.2466/pr0.2000.87.3.881

Bliemer, M.C., Rose, J.M., 2005. Efficiency and Sample Size Requirements for Stated Choice Studies. Inst. Transp. Stud. Logist. Work. Pap.

Bliemer, M.C.J., 2015a. How Many Choice Tasks to Keep in Pivot Design. Choice-Metricscom Forum Users Ngene.

Bliemer, M.C.J., 2015b. Getting 0% Ood Optimality for My Design Script. Choice-Metricscom Forum Users Ngene.

Bliemer, M.C.J., 2014. Efficent Rppanel or Bayesian Rp Model? Choice-Metricscom Forum Users Ngene.

Bliemer, M.C.J., Rose, J.M., 2011. Experimental Design Influences on Stated Choice Outputs: An Empirical Study in Air Travel Choice. Transp. Res. Part Policy Pract. 45, 63–79. doi:10.1016/j.tra.2010.09.003

Bolduc, D., Alvarez-Daziano, R., 2010. On Estimation of Hybrid Choice Models, in: Choice Modelling: The State-of-the-Art and The State-of-Practice. Emerald Group Publishing Limited, pp. 259–287.

Bolduc, D., Ben-Akiva, M., Walker, J., Michaud, A., 2005. Hybrid Choice Models with Logit Kernel: Applicability to Large Scale Models1, in: Integrated Land-Use and Transportation Models. Emerald Group Publishing Limited, pp. 275–302.

Boto-Giralda, D., Díaz-Pernas, F.J., González-Ortega, D., Díez-Higuera, J.F., Antón-Rodríguez, M., Martínez-Zarzuela, M., Torre-Díez, I., 2010. Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks. Comput.-Aided Civ. Infrastruct. Eng. 25, 530–545. doi:10.1111/j.1467-8667.2010.00668.x

Brown, S., 2010. Likert Scale Examples for Surveys. ANR Program Evaluation. Iowa State University Extension.

BTS, 2015. Bureau of Transport Statistics [WWW Document]. Househ. Travel Surv. URL http://www.bts.nsw.gov.au/Statistics/Household-Travel-Survey/default.aspx#top (accessed 2.5.16).

BTS, 2014. Household Travel Survey Report: Sydney 2012/13. Transport for NSW.

Burgess, L., Street, D.J., 2003. Optimal Designs for 2 k Choice Experiments. Commun. Stat. - Theory Methods 32, 2185–2206. doi:10.1081/STA-120024475

Campbell, D., Hutchinson, W.G., Scarpa, R., 2006. Lexicographic Preferences in Discrete Choice Experiments: Consequences on Individual-Specific Willingness to Pay Estimates (No. 2006.128), Working Papers. Fondazione Eni Enrico Mattei.

Cassidy, M.J., Bertini, R.L., 1999a. Some Traffic Features at Freeway Bottlenecks. Transp. Res. Part B Methodol. 33, 25–42. doi:10.1016/S0191-2615(98)00023-X

Cassidy, M.J., Bertini, R.L., 1999b. Observations at a Freeway Bottleneck. Presented at the 14th International Symposium on Transportation and Traffic Theory.

Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the Capacity of an Isolated Merge by Metering Its on-Ramp. Transp. Res. Part B Methodol. 39, 896–913. doi:10.1016/j.trb.2004.12.001

Chandler, R.E., Herman, R., Montroll, E.W., 1958. Traffic Dynamics: Studies in Car Following. Oper. Res. 6, 165–184.

Chen, D., Laval, J., Zheng, Z., Ahn, S., 2012. A Behavioral Car-Following Model That Captures Traffic Oscillations. Transp. Res. Part B Methodol. 46, 744–761. doi:10.1016/j.trb.2012.01.009

Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., Hicks, J., 2011. Dynamic Traffic Assignment: A Primer. Transp. Res. E-Circ.

ChoiceMetrics, 2012. Ngene 1.1.1 User Manual & Reference Guide, Australia.

Choudhary, P., Velaga, N.R., 2017. Modelling Driver Distraction Effects Due to Mobile Phone Use on Reaction Time. Transp. Res. Part C Emerg. Technol. 77, 351–365. doi:10.1016/j.trc.2017.02.007

Cirillo, C., Axhausen, K.W., 2006. Evidence on the Distribution of Values of Travel Time Savings from a Six-Week Diary. Transp. Res. Part Policy Pract. 40, 444–457. doi:10.1016/j.tra.2005.06.007

Collins, A., Rose, J., Hess, S., 2007. Stated Preference Survey Design in Air Travel Choice Behaviour Modelling. Proc. Eur. Transp. Conf. 2007 HELD 17-19 Oct. 2007 LEIDEN Neth.

Cook, R.D., Nachtsheim, C.J., 1980. A Comparison of Algorithms for Constructing Exact D-Optimal Designs. Technometrics 22, 315–324. doi:10.2307/1268315

Daubechies, I., 1992. Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics.

Davis, L.C., 2003. Modifications of the Optimal Velocity Traffic Model to Include Delay Due to Driver Reaction Time. Phys. Stat. Mech. Its Appl. 319, 557–567. doi:10.1016/S0378-4371(02)01457-7

Del Castillo, J.M., 2001. Propagation of Perturbations in Dense Traffic Flow: A Model and Its Implications. Transp. Res. Part B Methodol. 35, 367–389. doi:10.1016/S0191-2615(99)00055-7

Dixit, V., Harrison, G.W., Rutström, E.E., 2014. Estimating the Subjective Risks of Driving Simulator Accidents. Accid. Anal. Prev. 62, 63–78. doi:10.1016/j.aap.2013.08.023

Dixit, V.V., Harb, R.C., Martínez-Correa, J., Rutström, E.E., 2015. Measuring risk aversion to guide transportation policy: Contexts, incentives, and respondents. Transp. Res. Part Policy Pract. 80, 15–34. doi:10.1016/j.tra.2015.07.002

Dixit, V.V., Ortmann, A., Rutström, E.E., Ukkusuri, S.V., 2017. Experimental Economics and Choice in Transportation: Incentives and Context. Transp. Res. Part C Emerg. Technol. 77, 161–184. doi:10.1016/j.trc.2017.01.011

Edie, L.C., 1965. Discussion of Traffic Upstream Measurements and Definitions, in: The Second International Symposium on The Theory of Road Traffic Flow. London, pp. 139–154.

Edie, L.C., 1961. Car-Following and Steady-State Theory for Noncongested Traffic. Oper. Res. 9, 66–76.

Edie, L.C., Baverez, E., 1967. Generation and Propagation of Stop-Start Traffic Waves. in Vehicular Traffic Science. Presented at the Proceedings of the Third International Symposium on the Theory of Traffic Flow.

Erikson, E.H., Erikson, J.M., 1998. The Life Cycle Completed (Extended Version). W. W. Norton.

Farah, H., Bekhor, S., Polus, A., 2009. Risk Evaluation by Modeling of Passing Behavior on Two-Lane Rural Highways. Accid. Anal. Prev. 41, 887–894. doi:10.1016/j.aap.2009.05.006

Ferreira, L.J.A., 1982. Car Fuel Consumption in Urban Traffic. The Results of a Survey in Leeds using Instrumented Vehicles. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Fiorentino, D.D., Parseghian, Z., 1997. Time-To-Collision: A Sensitive Measure of Driver Interaction with Traffic in a Simulated Driving Task. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 41, 1028–1031. doi:10.1177/107118139704100267

Foote, R.S., 1965. Single Lane Traffic Flow Control, in: The Second International Symposium on The Theory of Road Traffic Flow. London, pp. 84–103.

Fosgerau, M., 2006. Investigating the Distribution of the Value of Travel Time Savings. Transp. Res. Part B Methodol. 40, 688–707. doi:10.1016/j.trb.2005.09.007

Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A Link Based Network Route Choice Model with Unrestricted Choice Set. Transp. Res. Part B Methodol. 56, 70–80. doi:10.1016/j.trb.2013.07.012

Frejinger, E., Bierlaire, M., 2007. Capturing Correlation with Subnetworks in Route Choice Models. Transp. Res. Part B Methodol. 41, 363–378. doi:10.1016/j.trb.2006.06.003

Gärling, T., Fujii, S., Gärling, A., Jakobsson, C., 2003. Moderating Effects of Social Value Orientation on Determinants of Proenvironmental Behavior Intention. J. Environ. Psychol. 23, 1–9. doi:10.1016/S0272-4944(02)00081-6

Gazis, D.C., Herman, R., Potts, R.B., 1959. Car-Following Theory of Steady-State Traffic Flow. Oper. Res. 7, 499–505.

Gazis, D.C., Herman, R., Rothery, R.W., 1961. Nonlinear Follow-The-Leader Models of Traffic Flow. Oper. Res. 9, 545–567.

Genz, 2017. Alan Genz Website [WWW Document]. URL http://www.math.wsu.edu/faculty/genz/homepage (accessed 8.5.17).

Ghosh-Dastidar, S., Adeli, H., 2006. Neural Network-Wavelet Microsimulation Model for Delay and Queue Length Estimation at Freeway Work Zones. J. Transp. Eng. 132, 331–341. doi:10.1061/(ASCE)0733-947X(2006)132:4(331)

Ghosh-Dastidar, S., Adeli, H., 2003. Wavelet-Clustering-Neural Network Model for Freeway Incident Detection. Comput.-Aided Civ. Infrastruct. Eng. 18, 325–338. doi:10.1111/1467-8667.t01-1-00311

Gipps, P.G., 1981. A Behavioural Car-Following Model for Computer Simulation. Transp. Res. Part B Methodol. 15, 105–111. doi:10.1016/0191-2615(81)90037-0

Golob, T.F., 2003. Structural Equation Modeling for Travel Behavior Research. Transp. Res. Part B Methodol. 37, 1–25. doi:10.1016/S0191-2615(01)00046-7

Greene, W.H., Hensher, D.A., 2010. Modeling Ordered Choices: A Primer, 1 edition. ed. Cambridge University Press, Cambridge England ; New York.

Gritzinger, B., 2015. Autoweb's Guide to… Adaptive Cruise Control [WWW Document]. Autoweb. URL http://editorial.autoweb.com/autowebs-guide-to-adaptive-cruise-control/ (accessed 4.9.17).

Hagenaars, J.A., McCutcheon, A.L. (Eds.), 2009. Applied Latent Class Analysis, 1 edition. ed. Cambridge University Press, Cambridge; New York.

Hamdar, S.H., Mahmassani, H.S., Chen, R.B., 2008. Aggressiveness Propensity Index for Driving Behavior at Signalized Intersections. Accid. Anal. Prev. 40, 315–326. doi:10.1016/j.aap.2007.06.013

Han, D., Yi, K., 2006. A Driver-Adaptive Range Policy for Adaptive Cruise Control. Proc. Inst. Mech. Eng. Part J. Automob. Eng. 220, 321–334. doi:10.1243/09544070JAUTO75

Haque, M.M., Washington, S., 2014. A Parametric Duration Model of the Reaction Times of Drivers Distracted by Mobile Phone Conversations. Accid. Anal. Prev. 62, 42–53. doi:10.1016/j.aap.2013.09.010

Hauber, A.R., 1980. The Social Psychology of Driving Behaviour and the Traffic Environment: Research on Aggressive Behaviour in Traffic. Appl. Psychol. 29, 461–474. doi:10.1111/j.1464-0597.1980.tb01106.x

Helbing, D., 1997. Fundamentals of Traffic Flow. Phys. Rev. E 55, 3735–3738. doi:10.1103/PhysRevE.55.3735

Hennessy, D.A., Wiesenthal, D.L., 1999. Traffic Congestion, Driver Stress, and Driver Aggression. Aggress. Behav. 25, 409–423. doi:10.1002/(SICI)1098-2337(1999)25:6<409::AID-AB2>3.0.CO;2-0

Hennessy, D.A., Wiesenthal, D.L., 1997. The Relationship Between Traffic Congestion, Driver Stress and Direct Versus Indirect Coping Behaviours. Ergonomics 40, 348–361. doi:10.1080/001401397188198

Hensher, D.A., 2008. Joint Estimation of Process and Outcome in Choice Experiments and Implications for Willingness to Pay. J. Transp. Econ. Policy 42, 297–322.

Hensher, D.A., 2004. Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. J. Transp. Econ. Policy 38, 425–446.

Hensher, D.A., 2001a. The Valuation of Commuter Travel Time Savings for Car Drivers: Evaluating Alternative Model Specifications. Transportation 28, 101–118. doi:10.1023/A:1010302117979

Hensher, D.A., 2001b. Measurement of the Valuation of Travel Time Savings. J. Transp. Econ. Policy 35, 71–98.

Hensher, D.A., 2001c. The Sensitivity of the Valuation of Travel Time Savings to the Specification of Unobserved Effects. Transp. Res. Part E Logist. Transp. Rev., Advances in the Valuation of Travel Time Savings 37, 129–142. doi:10.1016/S1366-5545(00)00015-6

Hensher, D.A., Greene, W.H., 2003. The Mixed Logit Model: The State of Practice. Transportation 30, 133–176. doi:10.1023/A:1022558715350

Hensher, D.A., Rose, J.M., Ortúzar, J. de D., Rizzi, L.I., 2009. Estimating the Willingness to Pay and Value of Risk Reduction for Car Occupants in the Road Environment. Transp. Res. Part Policy Pract. 43, 692–707. doi:10.1016/j.tra.2009.06.001

Herman, R., Montroll, E.W., Potts, R.B., Rothery, R.W., 1959. Traffic Dynamics: Analysis of Stability in Car Following. Oper. Res. 7, 86–106.

Hess, S., Ben-Akiva, M.E., Gopinath, D., Walker, J.L., 2009. Taste Heterogeneity, Correlation, and Elasticities in Latent Class Choice Models. Presented at the Transportation Research Board 88th Annual Meeting.

Hess, S., Bierlaire, M., Polak, J.W., 2005. Estimation of Value of Travel-Time Savings Using Mixed Logit Models. Transp. Res. Part Policy Pract., Positive Utility of Travel Positive Utility of Travel 39, 221–236. doi:10.1016/j.tra.2004.09.007

Hess, S., Rose, J.M., 2008. Should Reference Alternatives in Pivot Design SC Surveys be Treated Differently? Environ. Resour. Econ. 42, 297–317. doi:10.1007/s10640-008-9244-6

Hess, S., Rose, J.M., 2007. Intra-Respondent Taste Heterogeneity in Instantaneous Panel Surveys. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.

Hoedemaeker, M., 2000. Driving Behaviour with Acc and the Acceptance by Individual Drivers, in: ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493). Presented at the ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493), pp. 506–509. doi:10.1109/ITSC.2000.881121

Hohwü, L., Lyshol, H., Gissler, M., Jonsson, S.H., Petzold, M., Obel, C., 2013. Web-Based Versus Traditional Paper Questionnaires: A Mixed-Mode Survey With a Nordic Perspective. J. Med. Internet Res. 15. doi:10.2196/jmir.2595

Hole, A.R., 2007. Fitting Mixed Logit Models by Using Maximum Simulated Likelihood. Stata J. 7, 388–401.

Jeihani, M., NarooieNezhad, S., Bakhsh Kelarestaghi, K., 2017. Integration of a Driving Simulator and a Traffic Simulator Case Study: Exploring Drivers' Behavior in Response to Variable Message Signs. IATSS Res. doi:10.1016/j.iatssr.2017.03.001

Jiang, R., Wu, Q., Zhu, Z., 2001. Full Velocity Difference Model for a Car-Following Theory. Phys. Rev. E 64, 17101. doi:10.1103/PhysRevE.64.017101

Jiang, X., Adeli, H., 2005. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. J. Transp. Eng. 131, 771–779. doi:10.1061/(ASCE)0733-947X(2005)131:10(771)

Jiang, X., Adeli, H., 2004. Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis. Comput.-Aided Civ. Infrastruct. Eng. 19, 324–337. doi:10.1111/j.1467-8667.2004.00360.x

Jin, W.-L., Zhang, Y., 2005. Paramics Simulation of Periodic Oscillations Caused by Network Geometry. Transp. Res. Rec. J. Transp. Res. Board 1934, 188–196. doi:10.3141/1934-20

Jovanović, D., Lipovac, K., Stanojević, P., Stanojević, D., 2011. The Effects of Personality Traits on Driving-Related Anger and Aggressive Behaviour in Traffic Among Serbian Drivers. Transp. Res. Part F Traffic Psychol. Behav. 14, 43–53. doi:10.1016/j.trf.2010.09.005

Kamakura, W.A., Russell, G.J., 1989. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. J. Mark. Res. 26, 379–390. doi:10.2307/3172759

Kerner, B.S., 2004. Three-Phase Traffic Theory and Highway Capacity. Phys. Stat. Mech. Its Appl. 333, 379–440. doi:10.1016/j.physa.2003.10.017

Kim, T., Zhang, H.M., 2008. A Stochastic Wave Propagation Model. Transp. Res. Part B Methodol. 42, 619–634. doi:10.1016/j.trb.2007.12.002

Koshi, M., Masao, K., Hirokazu, A., 1992. Capacity of Sags and Tunnels in Japanese Motorways. Ite J. 62, 17–22.

Koutsopoulos, H.N., Lotan, T., Yang, Q., 1994. A Driving Simulator and Its Application for Modeling Route Choice in the Presence of Information. Transp. Res. Part C Emerg. Technol. 2, 91–107. doi:10.1016/0968-090X(94)90002-7

Koutsopoulos, H.N., Polydoropoulou, A., Ben-Akiva, M., 1995. Travel Simulators for Data Collection on Driver Behavior in the Presence of Information. Transp. Res. Part C Emerg. Technol. 3, 143–159. doi:10.1016/0968-090X(95)00003-2

Krinsky, I., Robb, A., 1990. On Approximating the Statistical Properties of Elasticities: A Correction. Rev. Econ. Stat. 72, 189–90.

Krueger, R., Vij, A., Rashidi, T.H., 2016. Normative Beliefs and Modality Styles: A Latent Class and Latent Variable Model of Travel Behaviour (SSRN Scholarly Paper No. ID 2790047). Social Science Research Network, Rochester, NY.

Kuhne, R.D., 1987. Freeway Speed Distribution and Acceleration Noise: Calculations from a Stochastic Continuum Theory and Comparison with Measurements. Transp. Traffic Theory.

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet Analysis for Geophysical Applications. Rev. Geophys. 35, 385–412.

Laval, J.A., 2011. Hysteresis in Traffic Flow Revisited: An Improved Measurement Method. Transp. Res. Part B Methodol. 45, 385–391. doi:10.1016/j.trb.2010.07.006

Laval, J.A., 2007. Linking Synchronized Flow and Kinematic Waves, in: Schadschneider, A., Pöschel, T., Kühne, R., Schreckenberg, M., Wolf, D.E. (Eds.), Traffic and Granular Flow’05. Springer Berlin Heidelberg, pp. 521–526.

Laval, J.A., Daganzo, C.F., 2006. Lane-Changing in Traffic Streams. Transp. Res. Part B Methodol. 40, 251–264. doi:10.1016/j.trb.2005.04.003

Laval, J.A., Leclercq, L., 2010a. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci. 368, 4519–4541. doi:10.1098/rsta.2010.0138

Laval, J.A., Leclercq, L., 2010b. A Mechanism to Describe the Formation and Propagation of Stop-and-Go Waves in Congested Freeway Traffic. Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci. 368, 4519–4541. doi:10.1098/rsta.2010.0138

Laval, J., Cassidy, M., Daganzo, C., 2007. Impacts of Lane Changes at Merge Bottlenecks: A Theory and Strategies to Maximize Capacity, in: Traffic and Granular Flow’05. Springer, Berlin, Heidelberg, pp. 577–586.

Laval, J., Chen, D., Amer, K., Guin, A., Ahn, S., 2009. Evolution of Oscillations in Congested Traffic. Transp. Res. Rec. J. Transp. Res. Board 2124, 194–202. doi:10.3141/2124-19

Lazarus, R.S., 1966. Psychological Stress and the Coping Process. McGraw-Hill.

Lee, H.C., Lee, A.H., Cameron, D., Li-Tsang, C., 2003. Using a Driving Simulator to Identify Older Drivers at Inflated Risk of Motor Vehicle Crashes. J. Safety Res. 34, 453–459.

Lee, J.D., McGehee, D.V., Brown, T.L., Reyes, M.L., 2002. Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator. Hum. Factors 44, 314–334. doi:10.1518/0018720024497844

Lee, Y.-C., 2010. Measuring Drivers’ Frustration in a Driving Simulator. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 54, 1531–1535. doi:10.1177/154193121005401937

Lee, Y.-C., Winston, F.K., 2016. Stress induction techniques in a driving simulator and reactions from newly licensed drivers. Transportation Research Part F: Traffic Psychology and Behaviour 42, Part 1, 44–55. https://doi.org/10.1016/j.trf.2016.06.019

Levinson, D., Harder, K., Bloomfield, J., Winiarczyk, K., 2004. Weighting Waiting: Evaluating Perception of In-Vehicle Travel Time Under Moving and Stopped Conditions. Transp. Res. Rec. J. Transp. Res. Board 1898, 61–68. doi:10.3141/1898-08

Li, X., Ouyang, Y., 2011. Characterization of Traffic Oscillation Propagation Under Nonlinear Car-Following Laws. Transp. Res. Part B Methodol., Select Papers from the 19th ISTTT 45, 1346–1361. doi:10.1016/j.trb.2011.05.010

Li, X., Peng, F., Ouyang, Y., 2010. Measurement and Estimation of Traffic Oscillation Properties. Transp. Res. Part B Methodol. 44, 1–14. doi:10.1016/j.trb.2009.05.003

Lighthill, M.J., Whitham, G.B., 1955. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. Proc. R. Soc. Lond. Math. Phys. Eng. Sci. 229, 317–345. doi:10.1098/rspa.1955.0089

Litman, T., 2015. Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. Presented at the Transportation Research Board 94th Annual Meeting.

Louviere, J.J., Hensher, D.A., 1983. Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. J. Consum. Res. 10, 348–361.

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press.

Louviere, J.J., Woodworth, G., 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. J. Mark. Res. 20, 350–367. doi:10.2307/3151440

MacCallum, R.C., Austin, J.T., 2000. Applications of Structural Equation Modeling in Psychological Research. Annu. Rev. Psychol. 51, 201–226. doi:10.1146/annurev.psych.51.1.201

Malta, L., Miyajima, C., Kitaoka, N., Takeda, K., 2011. Analysis of Real-World Driver's Frustration. IEEE Trans. Intell. Transp. Syst. 12, 109–118. doi:10.1109/TITS.2010.2070839

Marsden, G., Brackstone, M., McDonald, M., 2001. Assessment of the Stop and Go Function Using Real Driving Behaviour, in: Advanced Driver Assistance Systems, 2001. ADAS. International Conference on (IEE Conf. Publ. No. 483). Presented at the Advanced Driver Assistance Systems, 2001. ADAS. International Conference on (IEE Conf. Publ. No. 483), pp. 76–80. doi:10.1049/cp:20010503

Martinez, J.J., Canudas-de-Wit, C., 2007. A Safe Longitudinal Control for Adaptive Cruise Control and Stop-and-Go Scenarios. IEEE Trans. Control Syst. Technol. 15, 246–258. doi:10.1109/TCST.2006.886432

Mauch, M., Cassidy, M.J., 2004. Freeway Traffic Oscillations: Observations and Predictions. Univ. Calif. Transp. Cent.

McFadden, D.L., 1973. Conditional Logit Analysis of Qualitative Choice Behavior, in: Frontiers in Econometrics. Wiley, New York.

McKelvey, R.D., Zavoina, W., 1975. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. J. Math. Sociol. 4, 103–120. doi:10.1080/0022250X.1975.9989847

Mehler, B., Reimer, B., Coughlin, J., Dusek, J., 2009. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. Transp. Res. Rec. J. Transp. Res. Board 2138, 6–12. doi:10.3141/2138-02

Menendez, M., 2006. An Analysis of Hov Lanes: Their Impact on Traffic. University of California, Berkeley, Berkeley, US.

Muñoz, J.C., Daganzo, C.F., 2003. Structure of the Transition Zone Behind Freeway Queues. Transp. Sci. 37, 312–329. doi:10.1287/trsc.37.3.312.16043

Nagel, K., Nelson, P., 2005. A Critical Comparison of the Kinematic-Wave Model with Observational Data. Presented at the Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic TheoryUniversity of Maryland, College Park.

Nair, D.J., Saxena, N., Dixit, V.V., 2016. Evaluating the Perception Index for Cyclists' Safety Using Structural Equation Modelling. Presented at the 2016 International Conference on Transportation and Development, American Society of Civil Engineers, Houston, Texas. doi:http://dx.doi.org/10.1061/9780784479926.098

Naus, G., Bleek, R. van den, Ploeg, J., Scheepers, B., Molengraft, R. van de, Steinbuch, M., 2008. Explicit Mpc Design and Performance Evaluation of an Acc Stop- #x00026;-Go, in: 2008 American Control Conference. Presented at the 2008 American Control Conference, pp. 224–229. doi:10.1109/ACC.2008.4586495

Neubert, L., Santen, L., Schadschneider, A., Schreckenberg, M., 1999. Single-vehicle Data of Highway Traffic: A Statistical Analysis. Phys. Rev. E 60, 6480–6490. doi:10.1103/PhysRevE.60.6480

Newell, G.F., 2002. A Simplified Car-Following Theory: A Lower Order Model. Transp. Res. Part B Methodol. 36, 195–205. doi:10.1016/S0191-2615(00)00044-8

Newell, G.F., 1965. Instability in Dense Highway Traffic, a Review, in: The Second International Symposium on The Theory of Road Traffic Flow. London, pp. 73–83.

NGSIM, 2010. Next Generation Simulation (No. FHWA-HOP-13-015).

Norman, P., 1999. Putting Iterative Proportional Fitting on the Researcher's Desk [WWW Document]. URL http://www.geog.leeds.ac.uk/wpapers/ (accessed 3.7.17).

Nulty, D.D., 2008. The Adequacy of Response Rates to Online and Paper Surveys: What Can Be Done? Assess. Eval. High. Educ. 33, 301–314. doi:10.1080/02602930701293231

Orosz, G., Wilson, R.E., Szalai, R., Stépán, G., 2009. Exciting Traffic Jams: Nonlinear Phenomena Behind Traffic Jam Formation on Highways. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 80, 46205. doi:10.1103/PhysRevE.80.046205

Oshyani, M.F., Sundberg, M., Karlström, A., 2012. Estimating Flexible Route Choice Models Using Sparse Data, in: 2012 15th International IEEE Conference on Intelligent Transportation Systems. Presented at the 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 1215–1220. doi:10.1109/ITSC.2012.6338676

Paleti, R., Bhat, C.R., 2013. The Composite Marginal Likelihood (cml) Estimation of Panel Ordered-Response Models. J. Choice Model. 7, 24–43. doi:10.1016/j.jocm.2013.04.006

Phanikumar, C.V., Maitra, B., 2006. Modeling Willingness-to-Pay Values for Rural Bus Attributes Under Different Trip Purposes. J. Transp. Res. Forum 45, 31–44.

Pipes, L.A., 1950. A Proposed Dynamic Analogy of Traffic. Institute of Transportation and Traffic Engineering, University of California.

Qualtrics, 2016. Qualtrics [WWW Document]. Qualtrics. URL https://www.qualtrics.com/ (accessed 2.20.16).

RACQ, 2014. RACQ – Vehicle Running Cost 2014 [WWW Document]. URL http://www.racq.com.au/cars-and-driving/cars/owning-and-maintaining-a-car/car-running-costs (accessed 10.26.14).

Ramos, G., Frejinger, E., Daamen, W., Hoogendoorn, S., 2012. Route Choice Model Estimation in a Dynamic Network Based on Gps Data. Presented at the 1st European Symposium on Quantitive Methods in Transportation Systems, Lausanne (Switzerland).

Reimer, B., Mehler, B.L., Pohlmeyer, A.E., Coughlin, J.F., Dusek, J.A., 2006. The Use of Heart Rate in a Driving Simulator as an Indicator of Age-Related Differences in Driver Workload. Adv. Transp. Stud.

Research Lifeline, 2012. Four Survey Methodologies: A Comparison of Pros & Cons. Polaris Marketing Research, Inc., Atlanta.

Revelt, D., Train, K., 2000. Customer-Specific Taste Parameters and Mixed Logit: Households’ Choice of Electricity Supplier. Dep. Econ. UCB.

Richards, P.I., 1956. Shock Waves on the Highway. Oper. Res. 4, 42–51. doi:10.1287/opre.4.1.42

Rose, J.M., Bliemer, M.C.J., Hensher, D.A., Collins, A.T., 2008. Designing Efficient Stated Choice Experiments in the Presence of Reference Alternatives. Transp. Res. Part B Methodol. 42, 395–406. doi:10.1016/j.trb.2007.09.002

Rose, J.M., Hensher, D.A., Caussade, S., Ortúzar, J. de D., Jou, R.-C., 2009. Identifying Differences in Willingness to Pay Due to Dimensionality in Stated Choice Experiments: A Cross Country Analysis. J. Transp. Geogr. 17, 21–29. doi:10.1016/j.jtrangeo.2008.05.001

Rouphail, N., Tarko, A., Li, J., 2005. Traffic Flow at Signalized Intersections, in: In Revised Monograph on Traffic Flow Theory, Chapter 9.

SAE Mobilus, 2016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016 Ground Vehicle Standard) [WWW Document]. URL https://saemobilus.sae.org/content/j3016_201609 (accessed 5.3.17).

Sándor, Z., Wedel, M., 2002. Profile Construction in Experimental Choice Designs for Mixed Logit Models. Mark. Sci. 21, 455–475.

Sawada, S., 2002. Generalized Optimal Velocity Model for Traffic Flow. Int. J. Mod. Phys. C 13, 1–12. doi:10.1142/S0129183102002894

Scarpa, R., Ferrini, S., Willis, K., 2005. Performance of Error Component Models for Status-Quo Effects in Choice Experiments, in: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics, The Economics of Non-Market Goods and Resources. Springer Netherlands, pp. 247–273.

Schlereth, C., Eckert, C., Skiera, B., 2012. Using Discrete Choice Experiments to Estimate Willingness-to-Pay Intervals. Mark. Lett. 23, 761–776. doi:10.1007/s11002-012-9177-2

Schumacker, R.E., Lomax, R.G., 2012. A Beginner's Guide to Structural Equation Modeling: Third Edition. Routledge.

Shakeel, K., Rashidi, T.H., Waller, T.S., 2016. Choice Set Formation Behavior: Joint Mode and Route Choice Selection Model. Transp. Res. Rec. J. Transp. Res. Board 2563, 96–104. doi:10.3141/2563-14

Sheffi, Y., 1984. Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods. Prentice-Hall.

Shott, M.J., 2011. Traffic Oscillations Due To Topology and Route Choice in Elemental Freeway Networks. University of California, Davis.

Sieler, P., Song, B., Hedrick, J.K., 1998. Development of a Collision Avoidance System. Development 4, 17–22.

Small, K.A., 1999. Valuation of Travel-time Savings and Predictability in Congested Conditions for Highway User-cost Estimation. Transportation Research Board.

Srinivasan, S., 2005. Research on Women's Issues in Transportation, Report of a Conference: Technical papers. Transportation Research Board.

Stanton, N.A., Dunoyer, A., Leatherland, A., 2011. Detection of New In-path Targets by Drivers using Stop & Go Adaptive Cruise Control. Appl. Ergon., Applied Ergonomics and Transportation Safety 42, 592–601. doi:10.1016/j.apergo.2010.08.016

Stanton, N.A., Young, M., McCaulder, B., 1997. Drive-by-Wire: The Case of Driver Workload and Reclaiming Control with Adaptive Cruise Control. Saf. Sci. 27, 149–159. doi:10.1016/S0925-7535(97)00054-4

StataCorp., 2013, 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.

Street, D.J., Bunch, D.S., Moore, B.J., 2001. Optimal Designs for 2k Paired Comparison Experiments. Commun. Stat. - Theory Methods 30, 2149–2171. doi:10.1081/STA-100106068

Street, D.J., Burgess, L., 2004. Optimal and Near-Optimal Pairs for the Estimation of Effects in 2-Level Choice Experiments. J. Stat. Plan. Inference 118, 185–199. doi:10.1016/S0378-3758(02)00399-3

Street, D.J., Burgess, L., Louviere, J.J., 2005. Quick and Easy Choice Sets: Constructing Optimal and Nearly Optimal Stated Choice Experiments. Int. J. Res. Mark. 22, 459–470. doi:10.1016/j.ijresmar.2005.09.003

Suh, J., Yeo, H., Skabardonis, A., 2012. A Study on the Wave Development and Evolution Characteristics of Stop-and-Go Traffic. Presented at the Transportation Research Board 91st Annual MeetingTransportation Research Board.

Tanaka, M., Nakatsuji, T., 2011. Shock Waves and Speed Peak Propagations in Stop and Go Car-Following Conditions within Short Distances. Presented at the Transportation Research Board 90th Annual Meeting.

Temme, D., Paulssen, M., Dannewald, T., 2008. Incorporating Latent Variables into Discrete Choice Models — A Simultaneous Estimation Approach Using SEM Software. Bus. Res. 1, 220–237. doi:10.1007/BF03343535

The White House, 2014. An Economic Analysis of Transportation Infrastructure Investment. The White House, Washington DC, US.

Tian, H., 2010. Route Choice Behavior in a Driving Simulator with Real-Time Information (MS Thesis). University of Massachusetts, Amherst.

Train, K., Wilson, W.W., 2008. Estimation on Stated-preference Experiments Constructed from Revealed-preference Choices. Transp. Res. Part B Methodol., A Tribute to the Career of Frank Koppelman 42, 191–203. doi:10.1016/j.trb.2007.04.012

Train, K.E., 2009. Discrete Choice Methods with Simulation, 2 edition. ed. Cambridge University Press, Cambridge ; New York.

TRB, 2000. Highway Capacity Manual (HCM). Transportation Research Board, National Research Council, Washington DC, USA.

UNSW, 2014. Travel Survey and Campus Counts Analysis summary [WWW Document]. URL http://www.facilities.unsw.edu.au/sites/all/files/page_file_attachment/20141017%20Travel%20Survey%20-%20Analysis%20Summary.pdf (accessed 1.24.15).

van Driel, C.J.G., Hoedemaeker, M., van Arem, B., 2007. Impacts of a Congestion Assistant on driving behaviour and acceptance using a driving simulator. Transp. Res. Part F Traffic Psychol. Behav. 10, 139–152. doi:10.1016/j.trf.2006.08.003

Venhovens, P., Naab, K., Adiprastito, B., 2000. Stop and go cruise control. Int J Automot. Technol 1, 61–69.

Vermunt, J.K., Magidson, J., 2002. Latent Class Cluster Analysis, in: Applied Latent Class Analysis. Cambridge University Press, Cambridge, pp. 89–106.

Vij, A., Carrel, A., Walker, J.L., 2013. Incorporating the Influence of Latent Modal Preferences on Travel Mode Choice Behavior. Transp. Res. Part Policy Pract. 54, 164–178. doi:10.1016/j.tra.2013.07.008

Vlahogianni, E.I., Geroliminis, N., Skabardonis, A., 2008. Empirical and Analytical Investigation of Traffic Flow Regimes and Transitions in Signalized Arterials. J. Transp. Eng. 134, 512–522. doi:10.1061/(ASCE)0733-947X(2008)134:12(512)

Vogel, K., 2003. A Comparison of Headway and Time to Collision as Safety Indicators. Accid. Anal. Prev. 35, 427–433. doi:10.1016/S0001-4575(02)00022-2

Walker, J.L., Ben-Akiva, M., 2011. Advances in Discrete Choice: Mixture Models (Chapters). Edward Elgar Publishing.

WampServer, 2017. WampServer [WWW Document]. SourceForge. URL https://sourceforge.net/projects/wampserver/ (accessed 2.9.17).

Wardrop, J.G., 1952. Some Theoretical Aspects of Road Traffic Research. Presented at the Inst Civil Engineers Proc London /UK/.

Wells-Parker, E., Ceminsky, J., Hallberg, V., Snow, R.W., Dunaway, G., Guiling, S., Williams, M., Anderson, B., 2002. An Exploratory Study of the Relationship Between Road Rage and Crash Experience in a Representative Sample of Us Drivers. Accid. Anal. Prev. 34, 271–278. doi:10.1016/S0001-4575(01)00021-5

Wiesenthal, D.L., Hennessy, D., Gibson, P.M., 2000. The Driving Vengeance Questionnaire (DVQ): the development of a scale to measure deviant drivers' attitudes. Violence Vict. 15, 115–136.

Wijayaratna, S., Wijayaratna, K.P., 2016. Quantifying the Impact of On-Street Parking on Road Capacity: A Case Study of Sydney Arterial Roads. Presented at the Transportation Research Board 95th Annual MeetingTransportation Research Board.

Williams, E.J., 1949. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. Aust. J. Chem. 2, 149–168. doi:10.1071/ch9490149

Xie, Y., Zhang, Y., Ye, Z., 2007. Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. Comput.-Aided Civ. Infrastruct. Eng. 22, 326–334. doi:10.1111/j.1467-8667.2007.00489.x

Xiong, Z., Olstam, J., 2015. Orchestration of Driving Simulator Scenarios Based on Dynamic Actor Preparation and Automated Action Planning. Transp. Res. Part C Emerg. Technol. 56, 120–131. doi:10.1016/j.trc.2015.02.008

Yamamura, Y., Tabe, M., Kanehira, M., Murakami, T., 2001. Development of an Adaptive Cruise Control System with Stop-and-Go Capability. Presented at the SAE Paper No. 2001-01-0798.

Yan, X., Radwan, E., Klee, H., 2003. Driving Simulators Application to Left Turn Gap Acceptance, in: Summer Computer Simulation Conference. Presented at the Society for Computer Simulation International; 1998, pp. 590–595.

Yang, Q., Koutsopoulos, H.N., 1996. A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management Systems. Transp. Res. Part C Emerg. Technol. 4, 113–129. doi:10.1016/S0968-090X(96)00006-X

Yang, Y., Reimer, B., Mehler, B., Dobres, J., 2013. A Field Study Assessing Driving Performance, Visual Attention, Heart Rate and Subjective Ratings in Response to Two Types of Cognitive Workload. Presented at the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, New York.

Yeo, H., 2008. Asymmetric Microscopic Driving Behavior Theory. University of California, Berkeley, Berkeley, US.

Yeo, H., Skabardonis, A., 2009. Understanding Stop-and-go Traffic in View of Asymmetric Traffic Theory, in: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee. Springer US, pp. 99–115.

Yi, K., Moon, I.-K., 2004. A Driver-adaptive Stop-and-Go Cruise Control Strategy, in: 2004 IEEE International Conference on Networking, Sensing and Control. Presented at the 2004 IEEE International Conference on Networking, Sensing and Control, p. 601–606 Vol.1. doi:10.1109/ICNSC.2004.1297507

Zhao, X., Gao, Z., 2005. A New Car-following Model: Full Velocity and Acceleration Difference Model. Eur. Phys. J. B 47, 145–150. doi:10.1140/epjb/e2005-00304-3

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011a. Applications of Wavelet Transform for Analysis of Freeway Traffic: Bottlenecks, Transient Traffic, and Traffic Oscillations. Transp. Res. Part B Methodol. 45, 372–384. doi:10.1016/j.trb.2010.08.002

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011b. Freeway Traffic Oscillations: Microscopic Analysis of Formations and Propagations Using Wavelet Transform. Transp. Res. Part B Methodol., Select Papers from the 19th ISTTT 45, 1378–1388. doi:10.1016/j.trb.2011.05.012

Zheng, Z., Ahn, S., Monsere, C.M., 2010. Impact of Traffic Oscillations on Freeway Crash Occurrences. Accid. Anal. Prev. 42, 626–636. doi:10.1016/j.aap.2009.10.009

Zheng, Z., Washington, S., 2012. On Selecting an Optimal Wavelet for Detecting Singularities in Traffic and Vehicular Data. Transp. Res. Part C Emerg. Technol. 25, 18–33. doi:10.1016/j.trc.2012.03.006

Zielke, B., Bertini, R., Treiber, M., 2008. Empirical Measurement of Freeway Oscillation Characteristics: An International Comparison. Transp. Res. Rec. J. Transp. Res. Board 2088, 57–67. doi:10.3141/2088-07

APPENDIX A

WAVELET TRANSFORMATION

For identifying transient locations in a non-stationary signal, Wavelet Transform (WT) has evolved as a widely used technique over time. WT is useful in discerning both the time and frequency components of a signal fluctuation, which is more detailed than its stationary signal processing counterparts. Experiments suggest the superiority of WT over popular techniques with regards to accuracy, robustness and consistency (Addison, 2002; Zheng & Washington, 2012). Figure A.1 explains the key differences in the degree of resolution among the Fourier, Short Time Fourier Transform (STFT), and WT techniques, based on Heisenberg's uncertainty principle. Heisenberg's uncertainty principle states that the area of each rectangular box, as shown in figure A.1, must be equal (which is $f \times t$ in this case). The thin, long rectangles in figure A.1(a) signify that Fourier transform gives a high frequency resolution, but compromises on the temporal detail. An STFT in figure A.1(b) improves the former technique by capturing some temporal detail through a smaller time window at an expense of losing some frequency resolution. However, the window length is decided subjectively by an analyst, and is fixed for the entire length of a signal. Wavelet transform (WT) in figure A.1(c) outperforms the other two techniques by using time and frequency windows of variable lengths to analyse transient points in a non-stationary signal. As a general rule, WT provides good frequency and temporal resolution by using long and short time windows for low and high frequency signals respectively. As WT technique requires no subjective judgement in selecting the size and shape of a time window, the results can easily be replicated across analysts.

A.1 Properties of a Wavelet

Wavelet transformation identifies the location and frequency of a signal fluctuation by scaling (dilation and compression) and translating a suitable wavelet function over the time domain of a signal. A wavelet is represented by a complex mathematical function that should satisfy the conditions stated in equations A.1 and A.2

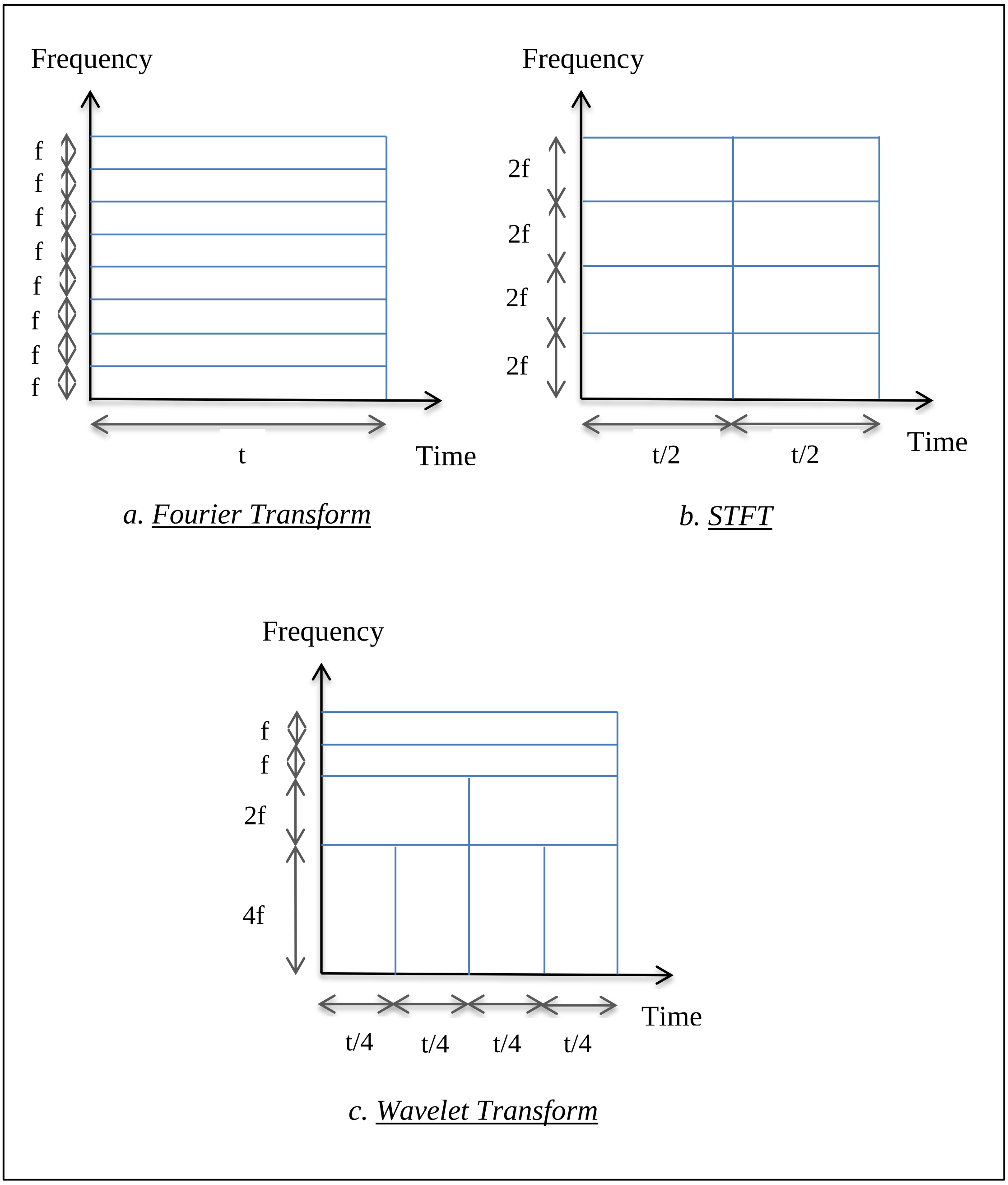

Figure A.1: Frequency-time plots in accordance with Heisenberg's uncertainty principle for (a) Fourier, (b) STFT and (c) Wavelet transformations

$$E = \int_{-\infty}^{\infty} |\varphi(t)|^2 \, dt < \infty \qquad \text{(A.1)}$$

$$\int_{-\infty}^{\infty} \varphi(t)\, dt = 0 \qquad \text{(A.2)}$$

Equation A.1, which is also called the admissibility condition, states that a wavelet must have a finite energy (E) at each point within a signal domain (Daubechies, 1992). This energy value is represented as brightness in the time-frequency plot shown in figure A.1(c). The level of brightness increases as one nears a localised fluctuation. Equation A.2 suggests that a wavelet should have a zero mean value, which implies that the total area under a given wavelet should be zero. This property also aids in recognising localised fluctuations.

Wavelet analysis can be broadly classified into continuous and discrete wavelet transformation. While Discrete Wavelet Transform (DWT) provides an accurate location of fluctuation in space and time at lesser computational cost, Continuous Wavelet Transform (CWT) is considered ideal for detecting sharp changes in a signal (Kumar & Foufoula-Georgiou, 1997). A wide range of wavelets can be used to identify the traffic oscillation information from the given data. Zheng & Washington (2012) provide exhaustive guidelines for selecting an appropriate wavelet from a candidate set. According to the guidelines, it is vital to know the mathematical properties of a given wavelet, which further paves the way for its application in a given context. The two important properties of a wavelet are: 1) vanishing moments and 2) compact support. A wavelet function $\varphi(t)$ is said to possess n vanishing moments if equation A.3 is satisfied.

$$\int_{-\infty}^{\infty} x^k\, \varphi(t)\, dt = 0 \qquad \forall\, k\ \in [1, (n-1)] \qquad \text{(A.3)}$$

A wavelet with more vanishing moments is capable of analysing a more complex signal. For example, a wavelet with 2 vanishing moments can identify only a linear discontinuity (degree 1). Similarly, a wavelet with 3 vanishing moments can identify up to a quadratic fluctuation (degree 2). Compact support is defined as an interval within which a wavelet is defined (or is non-zero). The wavelet function is evaluated at each point in the support domain for CWT case and at discrete points in case of DWT. Some interesting points about vanishing moments and compact support are as follows:

- High frequency wavelets (at a smaller scale) have a compact support
- More vanishing moments denotes a complex wavelet, which ensures an accurate representation of an input signal. However, it leads to a sharp increase in the computation time
- Wavelets with more vanishing moments are high on regularity (more smooth wavelet function), but require a wider support domain
- Wavelets with a smaller support size are more efficient in detecting transient locations in a signal

A good wavelet should preferably have fewer vanishing moments that are adequate enough for analysing the signal fluctuation of interest.

A.2 Application in Identifying Stop-&-go Waves

A Mexican hat wavelet is found to perform reasonably well in identifying stop-&-go waves from the vehicle trajectory data (Zheng et al., 2011a,b; Zheng & Washington, 2012). A Mexican hat wavelet represents the second derivative of the standard Gaussian function and its shape resembles a traffic oscillation pattern. Equation A.4 gives a general equation for this wavelet where μ and σ represent the translation and scale parameter respectively. The value $\frac{2}{\pi^{1/4}\sqrt{3\sigma}}$ ensures that the wavelet function at different scales has the same energy. Properties of other wavelets like Haar, Gauss, Daubechies, Meyer, Morlet, etc. are thoroughly reviewed in Zheng & Washington (2012).

$$\varphi(\mu,\sigma,t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}\left(\left(\frac{t-\mu}{\sigma}\right)^2 - 1\right)\exp\left(-\left(\frac{t-\mu}{\sigma}\right)^2\right) \quad \text{(A.4)}$$

$$T(\mu, \sigma) = \int_{-\infty}^{\infty} \varphi(t).\, v(t)\, dt \tag{A.5}$$

$$E_b = \frac{1}{\max(\sigma)} \int_{0}^{\infty} \big(T(\mu, \sigma)\big)^2\, d\sigma \tag{A.6}$$

A wavelet is translated over the time domain of an input signal $v(t)$ and the correlation coefficient $T(\mu, \sigma)$ is determined at a given scale (refer to equation A.5). The wavelet is then dilated and is made to run over the entire signal again. Thus, we get a plot between the scale and translation parameter, where $T(\mu, \sigma)$ is represented by the level of brightness at that scale and translation. This plot is known as a scalogram. A brighter area on the scalogram signifies the spatio-temporal location of the point of singularity in an input signal. A scalogram plot requires a visual inspection to identify the transient points, which becomes a cumbersome task while analysing bigger datasets. Zheng et al. (2011a) proposed an automation procedure, by calculating average wavelet based energy (E_b) from the scalogram plot. Equation A.6 gives an expression for evaluating this metric. An average wavelet based energy (E_b) at a given translation (μ) is defined as the average of squared correlation coefficients ($T(\mu, \sigma)$) across all scales (σ). A peak in the energy profile represents an approximate location of a transient point in data.

Zheng et al. (2011a) used the speed-time plot of individual vehicles (from NGSIM dataset) to locate traffic oscillations. The authors defined traffic oscillation as a cyclic pattern characterised by: 1) arrival of deceleration wave, 2) arrival of acceleration wave, and 3) arrival of another deceleration wave. A deceleration wave is identified by a sudden change in the speed of a vehicle which causes a sharp spike in the average wavelet based energy. Hence, one can precisely determine the occurrence of traffic oscillations experienced by individual vehicles using the wavelet transform.

A.3 Analysis using the Vehicle Trajectory Data: An Example

We applied WT on the vehicle trajectory data to identify locations of traffic oscillation, followed by evaluating the surrogate measures, like the time headway, peak acceleration (deceleration) and jerk values, around these spots. We select the time headway over the Time To Collision (TTC), as it accounts for potential hazards, which is unlike TTC that is primarily used for evaluating safety (Vogel, 2003). We present the analysis for a single vehicle undergoing S&G for illustration.

NGSIM data on the interstate US-101 is used for analysis in this study. The NGSIM dataset provides a rich information on the microscopic features of a vehicle like speed, location, acceleration, etc., collected at a fine temporal resolution of 10 Hertz (NGSIM, 2010). The data was collected during the morning peak traffic between 07:50am to 08:35am. The study site on US-101 is a 2100 feet long section in the southbound direction, Los Angeles, California, US. The section has 5 traffic lanes and also includes an on and off ramp. Lane-4, being in the vicinity of ramps, is selected because of a higher likelihood of witnessing S&G waves. Figure A.2(a) shows the speed profile of the vehicle id 540 travelling on lane-4. The figure shows a major speed fluctuation between 810 and 845 seconds from the start of data collection.

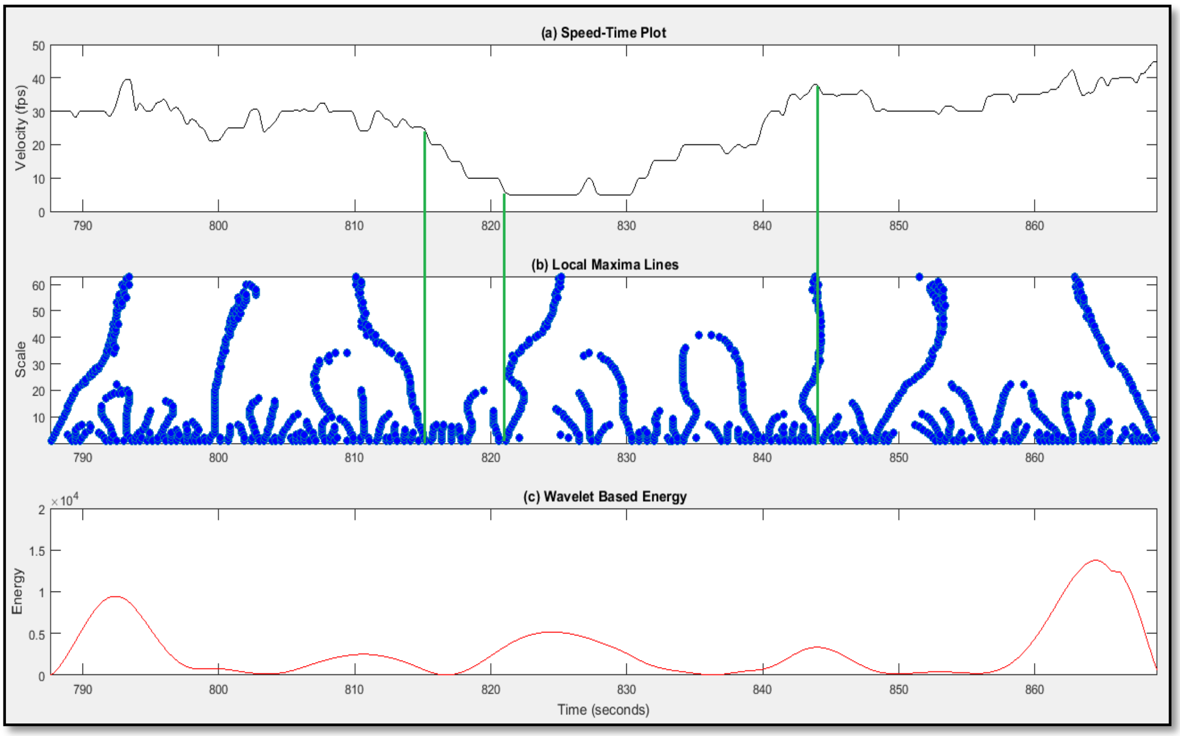

Figure A.2: Plots generated for vehicle id. 540 (a) speed-time, (b) Local maxima lines and (c) Average wavelet based energy

A wavelet transformation is conducted by selecting the Mexican hat mother wavelet with a scale range between 1 and 64 (Zheng & Washington, 2012). Figure A.2(b) shows the local maxima lines obtained upon analysing the speed trajectory signal given in figure A.2(a). Like the scalogram, local maxima lines are another way of representing singularities in the input signal. A local maxima line is a locus of points across scales where the correlation coefficient is a local maximum. Only the lines that are formed over the entire scale spectrum are considered. Partial lines and scattered points are ignored as these are generated due to noise in the signal. The oscillation location can be accurately determined by tracing the location of a maxima line at the finest scale, which is 1 in this figure (Zheng & Washington, 2012). For example, the green lines show a mapping of oscillation points on the speed time plot. Figure A.2(c) shows the plot of average wavelet based energy that was calculated using equation A.6. The figure shows the formation of three energy peaks, which jointly define a traffic oscillation (Zheng et al., 2011a). Thus, it can be inferred that vehicle 540 experienced one cycle of stop-&-go (S&G). The unexpected peaks at the start and end of figures A.2(b) and A.2(c) constitute the boundary effect (Addison, 2002), which are generally ignored. The analysis is done using the wavelet toolbox in Matlab.

Once the traffic oscillation location in time is determined from the plots, we then evaluate the surrogate measures around this location from the given data for vehicle id 540. The time headway, acceleration (deceleration) and experienced jerk value at an instant when the S&G wave is initiated are found as 1.90 seconds, -4.06 m/s^2 and -2.35 m/s^3 respectively. The negative value of jerk signifies that the driver hit hard on the brakes causing a sudden reduction in its speed. Not surprisingly, the observed deceleration value is much lower than what is used in the design of ACC systems. Thus, finding out these surrogate measures from the real-world data can provide us with alternate metrics of quantifying S&G waves.

APPENDIX B

WEB URL TO THE RESOURCES

I. Ngene Scripts for the Design of the SC Experiments

https://github.com/neerajsaxena040885/Ngene-Scripts.git

II. Online Survey Pages

1. URL to the Proof of concept study discussed in Chapter 3 of this thesis: http://rcitiunsw.epizy.com/sngotest/index.php

2. URL to the Extended study discussed in Chapter 4 of this thesis: http://rcitiunsw.epizy.com/sngosydney/index.php

3. URL to the Driving Simulator study discussed in Chapter 5 of this thesis: http://rcitiunsw.epizy.com/drivsim/index.php

III. Matlab Code of the Statistical Models used

1. RPECL model discussed in Chapter 3 of this thesis: https://github.com/neerajsaxena040885/RPECL.git

2. LCCM model discussed in Chapter 4 of this thesis: https://github.com/neerajsaxena040885/LCCM.git

3. SEM model discussed in Chapter 5 of this thesis: https://github.com/neerajsaxena040885/SEM.git

4. Data Generation Process using the SEM model https://github.com/neerajsaxena040885/SEM-DGP-100.git

APPENDIX C

ETHICS APPROVAL LETTERS

C.1 Online Stated Choice Study

UNSW AUSTRALIA

Human Research Ethics Advisory (HREA) Panel H: Science and Engineering
The University of New South Wales
UNSW Sydney, NSW, Australia, 2052
E: HREA8@unsw.edu.au

02-Apr-2015

Dear Dr Vinayak Dixit,

Project Title	Route choice of car drivers based on stop & go conditions on urban road networks
HC No	HC15007
Re	Notification of Ethics Approval
Approval Period	02-Apr-2015 - 01-Apr-2020

Thank you for submitting the above research project to the **HREAP H: Science/Engineering** for ethical review. This project was considered by the **HREAP H: Science/Engineering** at its meeting on 02-Apr-2015.

I am pleased to advise you that the **HREAP H: Science/Engineering** has granted ethical approval of this research project, subject to the following conditions being met:

Conditions of Approval Specific to Project:
1. Application form, section 2.1: "interview" should not be checked, "self-report questionnaire" should be checked. Please make this revision.

2. Initial feedback, point 6: The HREAP advises that the response provided has not adequately addressed the question that was asked. Given that the target population for the study will be UNSW employees, please justify the reasons for asking questions regarding employment status and occupation.

3. In addition to the above, the HREAP asks that the questionnaire be revised to ensure that it is relevant to the targeted study population.

Conditions of Approval – All Projects:

- The Chief Investigator will immediately report anything that might warrant review of ethical approval of the project.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of any event that requires a modification to the protocol or other project documents and submit any required amendments in accordance with the instructions provided by the **HREAP H: Science/Engineering**. These instructions can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will submit any necessary reports related to the safety of research participants in accordance with **HREAP H: Science/Engineering** policy and procedures. These instructions can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will report to the **HREAP H: Science/Engineering** annually in the specified format and notify the HREC when the project is completed at all sites.

- The Chief Investigator will notify the **HREAP H: Science/Engineering** if the project is discontinued at a participating site before the expected completion date, with reasons provided.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of any plan to extend the duration of the project past the approval period listed above and will submit any associated required documentation. Iinstructions for obtaining an extension of approval can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of his or her inability to continue as Coordinating Chief Investigator including the name of and contact information for a replacement.

A copy of this ethical approval letter must be submitted to all Investigators and sites prior to commencing the project.

The **HREAP H: Science/Engineering** Terms of Reference, Standard Operating Procedures, membership and standard forms are available from https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.

Should you require any further information, please contact the Ethics Administrator at:

E: HREA8@unsw.edu.au
W:https://research.unsw.edu.au/human-research-ethics-home

The HREAP H: Science/Engineering wishes you every continued success in your research.

Kind Regards

Dr Bruno Gaeta
Convenor HREA Panel H: Science and Engineering

C.2 Driving Simulator Study

Human Research Ethics Advisory (HREA) Panel H: Science and Engineering
The University of New South Wales
UNSW Sydney, NSW, Australia, 2052
E: HREA8@unsw.edu.au

18-Nov-2015

Dear Dr Vinayak Dixit,

Project Title	Route choice of car drivers based on stop-&-go conditions on urban road networks: A Driving Simulator Study
HC No	HC15752
Re	Notification of Ethics Approval
Approval Period	18-Nov-2015 - 17-Nov-2020

Thank you for submitting the above research project to the **HREAP H: Science/Engineering** for ethical review. This project was considered by the **HREAP H: Science/Engineering** at its meeting on 10-Nov-2015.

I am pleased to advise you that the **HREAP H: Science/Engineering** has granted ethical approval of this research project, subject to the following conditions being met:

Conditions of Approval Specific to Project:
N/A

Conditions of Approval – All Projects:

- The Chief Investigator will immediately report anything that might warrant review of ethical approval of the project.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of any event that requires a modification to the protocol or other project documents and submit any required amendments in accordance with the instructions provided by the **HREAP H: Science/Engineering**. These instructions can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will submit any necessary reports related to the safety of research participants in accordance with **HREAP H: Science/Engineering** policy and procedures. These instructions can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will report to the **HREAP H: Science/Engineering** annually in the specified format and notify the HREC when the project is completed at all sites.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** if the project is discontinued at a participating site before the expected completion date, with reasons provided.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of any plan to extend the duration of the project past the approval period listed above and will submit any associated required documentation. Iinstructions for obtaining an extension of approval can be found at https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.
- The Chief Investigator will notify the **HREAP H: Science/Engineering** of his or her inability to continue

as Coordinating Chief Investigator including the name of and contact information for a replacement.

A copy of this ethical approval letter must be submitted to all Investigators and sites prior to commencing the project.

The **HREAP H: Science/Engineering** Terms of Reference, Standard Operating Procedures, membership and standard forms are available from https://research.unsw.edu.au/research-ethics-and-compliance-support-recs.

Should you require any further information, please contact the Ethics Administrator at:

E: HREA8@unsw.edu.au
W:https://research.unsw.edu.au/human-research-ethics-home

The HREAP H: Science/Engineering wishes you every continued success in your research.

Kind Regards

Dr Bruno Gaeta
Convenor HREA Panel H: Science and Engineering

APPENDIX D

RESULTS FROM THE OTHER MODEL SPECIFICATIONS

D.1 Chapter 3: Proof of Concept Study

Different model specifications were tried on the dataset discussed in chapter 3. The first alternate model specification tried was the Random Parameter Error Component Logit (RPECL) model with a triangular distribution. The parameter estimates obtained from this model are shown in table D.1.

Table D.1: Results from the RPECL model with a triangular distribution

Attribute	Estimated parameters
Mean of random parameters	
Travel time	-0.3048 ***
Time spent in stop-&-go	-0.1291 ***
Number of stop-&-go	-0.0852 ***
Standard deviation of random parameters	
Travel time	0.4369 ***
Time spent in stop-&-go	0.2691 ***
Number of stop-&-go	0.1553 ***
Non-random parameters	
Running cost	-1.1965 ***
Sigma (σ)	-0.2604
Log-likelihood at convergence	-1115.1818
Adjusted Rho-squared	0.1835

*** significant at 99%

We then estimated the Latent Class Choice Model (LCCM) specification as discussed in chapter 4 of this thesis. In this model, the class membership model comprised gender, age and a constant term as the covariates. We also used income as the covariate but it turned out insignificant in all the LCCMs tested by us. Thus, we dropped the variable from the model. The choice model in this specification was an error component logit with the same number of attributes (four) as used in the RPECL model. We tried the LCCM with a different number of segments. Tables D.2 and D.3 show the model estimates with two and three latent segments.

Table D.2: Results from the two segment LCCM

Parameters	**Class 1**	**Class 2**
Class Membership Model		
Constant	-0.4515 *	0
Females	0.8921 ***	0
Age (below 40 years)	-0.9103 ***	0
Choice Model		
Travel time	-0.4118 ***	-0.0928 ***
Time spent in stop-&-go	-0.2177 ***	0.0221 ***
Number of stop-&-gos	-0.1013 ***	-0.0177 ***
Running cost	-0.9954 ***	-1.0181 ***
Sigma (σ)	0.8143 ***	0.7209 ***
Log-likelihood at convergence	-1087.7194	
Adjusted Rho-squared	0.2104	

*** significant at 99%   * significant at 90%

Table D.3: Results from the three segment LCCM

Parameters	**Class 1**	**Class 2**	**Class 3**
Class Membership Model			
Constant	-1.0652 *	-0.9528	0
Females	0.8457	0.6510	0
Age (below 40 years)	-0.8197	1.0653	0
Choice Model			
Travel time	-0.3611 ***	-0.6262 ***	-0.0904 ***
Time spent in stop-&-go	-0.3720 ***	-0.2058 ***	-0.0182 ***
Number of stop-&-gos	-0.2029 ***	-0.0724 ***	-0.0228 ***
Running cost	-0.9490 ***	-1.6764 ***	-1.0216 ***
Sigma (σ)	0.6475 ***	1.1471 **	0.4526 ***
Log-likelihood at convergence	-1059.6004		
Adjusted Rho-squared	0.2304		

*** significant at 99% ** significant at 95% * significant at 90%

Comparison of the Models

Table D.4 shows a comparison of the different model specifications based on the number of estimated parameters, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). All the tested models show a significant disutility towards the four attributes of interest, particularly the number of S&Gs which validates the proposed hypothesis.

Table D.4: Comparison of the models tested

Model specification	**No. of parameters**	**AIC**	**BIC**
RPL	7	2254.74	2299.38
RPECL (normal distribution)	8	2241.38	2292.40
RPECL (triangular distribution)	8	2146.36	2297.38
LCCM (2 segments)	13	2201.44	2284.35
LCCM (3 segments)	21	2161.20	2295.14

The table shows that the 2 segment LCCM has a slightly lower BIC value (2284.35) when compared to the RPECL (normal distribution) model (2292.40) reported in chapter 3 of this thesis. However, the 2 segment LCCM estimates a higher number of parameters (13) when compared to the RPECL model (8). Thus, the RPECL model with the normal distribution is a reasonable option as it is not only parsimonious, but also has a decent BIC value. The other models tested have a higher BIC value and are thus discarded for further consideration.

Looking at table D.2 again, one of the segments (class 1) in the 2 segment LCCM shows a statistically significant positive parameter value (0.8921) for the covariate female. This implies that being a female makes a person more likely to belong to this segment. The value of time for this segment is quite high, when compared to the second segment, and is evaluated as AU$ 24.82 $(-0.4118/-0.9954 \times 60)$. This observation is not consistent with the previous literature which found females to have a lower value of time (Srinivasan, 2005). Thus, we finally selected the RPECL model with the normal distribution, which is reported in chapter 3, as a tool to prove the validity of the proposed research hypothesis.

D.2 Chapter 5: Driving Simulator Study

Different model specifications of the Structural Equation Model (SEM) were tried to find the model which satisfied the following criteria: 1) model parsimony, 2) the best goodness of fit statistics, and 3) conveyed meaningful interpretations. Table D.5 shows the specifications tried by us. The colour coding scheme used in this table is presented at the bottom of the table.

Table D.5: Alternate specifications of the SEM tested

Parameters	Model 1	Model 2	Model 3	Model 4
Structural part				
Travel Time	+	+	+	
Time in S&G	+	+	+	+
Number of S&Gs	+	+	+	+
Fuel Cost		-		
Male	-	-	-	-
Age <= 40 Exp <= 8	+	+	+	+
Age > 40 Exp > 8	+	+	+	+
Measurement part				
Travel Time				+
Fuel Cost			-	
Factor Loading	+	+	+	+
Left Route				
Constant	-	+	+	-
μ_1	+	+	+	+
μ_2	+	+	+	+
μ_3	+	+	+	+
Right Route				
Constant	-	+	+	-
μ_1	+	+	+	+
μ_2	+	+	+	+
μ_3	+	+	+	+
Goodness of Fit				
Log-likelihood	-3813.16	-3808.42	-3808.42	-3813.16
Adj. Rho-squared	0.1180	0.1146	0.1300	0.1498
AIC	7656.32	7648.84	7648.84	7656.32
BIC	7722.12	7719.03	7719.03	7722.12

Legend:

	Significant at 90%		Insignificant at 90%		Not considered

APPENDIX E

THE SEM MODEL: SIMULATION STUDY

We generated simulated datasets of varying sizes (50, 70, 100 and 500 participants) using the Structural Equation Model (SEM) specification to determine the minimum sample size required for the driving simulator experiment. The first step was to generate the matrix of route specific and socio-demographic attributes. Each of the route specific attributes was drawn from a uniform distribution with the range as follows: travel time {min: 1.0; max: 3.75}, time spent in S&G {min: 1.0; max: 2.5}, number of S&Gs {min: 1.0; max: 3.0}, and fuel cost {min: 1.5; max: 3.0}. Binary socio-demographic variables like age less than 40 years and driving experience less than 8 years were generated using the following procedure: 1) set the cumulative density as 0.40 and 0.48 respectively for the two variables, 2) generate uniformly distributed random numbers between 0 and 1, and 3) if the random number is less than the cumulative density value then assign 0 else 1. True values for the structural and measurement parameters were assumed to generate the resulting frustration ratings. Additionally, the unobserved error terms were drawn from a standard normal distribution as follows: 1) the structural unobserved error remains the same for an individual across scenarios, and 2) the measurement errors are different for every individual, route and scenario.

We executed the SEM on a set of 50 different simulated datasets for a given sample size. The resulting parameter values from each run were averaged to find the mean parameter value for that sample size, as shown in Table E.1. The performance of a sample size was determined using the Absolute Percentage Bias (APB) statistic (for every parameter) proposed by Paleti & Bhat (2013) which is shown in equation E.1.

$$APB\ (\%) = \left| \frac{Mean\ Estimate - True\ Value}{True\ Value} \times 100 \right| \qquad \text{(E.1)}$$

Table E.1: Model performance by sample size

Parameter	True value	Sample size = 50		Sample size = 70		Sample size = 100		Sample size = 500	
		Mean	**APB**	**Mean**	**APB**	**Mean**	**APB**	**Mean**	**APB**
a_{TTS}	0.55	0.5217	5.145	0.5575	1.364	0.6004	9.164	0.563	2.433
a_{SnGo}	0.2	0.1896	5.2	0.1934	3.3	0.1749	12.55	0.198	0.891
$a_{AgeLT40}$	0.5	0.4323	13.54	0.4305	13.9	0.4627	7.46	0.493	1.485
a_{ExpLT8}	0.65	0.6441	0.908	0.6646	2.246	0.6971	7.246	0.662	1.914
b_{L_Cons}	-0.95	-1.0216	7.537	-1.0332	8.758	-1.0321	8.642	-0.973	2.377
b_{R_Cons}	-1.10	-1.2083	9.845	-1.1675	6.136	-1.2115	10.136	-1.117	1.557
d	1.05	0.8323	20.733	0.9801	6.657	0.9715	7.476	1.048	0.224
μ_{L_1}	-0.5	-0.5235	4.7	-0.4639	7.22	-0.4793	4.14	-0.516	3.131
μ_{L_2}	-0.6	-0.6105	1.75	-0.6019	0.317	-0.6467	7.783	-0.583	2.897
μ_{L_3}	-0.65	-0.6453	0.723	-0.7198	10.738	-0.6536	0.554	-0.65	0.045
μ_{R_1}	-0.4	-0.4393	9.825	-0.3968	0.8	-0.4192	4.8	-0.406	1.536
μ_{R_2}	-0.55	-0.5168	6.036	-0.5468	0.582	-0.5937	7.945	-0.527	4.213
μ_{R_3}	-0.7	-0.7901	12.871	-0.6807	2.757	-0.7323	4.614	-0.709	1.218
Mean APB			**7.805**		**4.982**		**7.116**		**1.840**

Table E.1 shows the mean APB for the sample size of 70 participants as 4.982 which is quite low. Given the time and budget constraints, we decided to collect data for at least 70 participants in our driving simulator study. We finally conducted the study on 99 participants to account for any variability in the responses which was probably not captured in the simulated dataset.

APPENDIX F

ANALYSING THE EFFECT OF STOP-&-GO CONDITIONS IN TRAFFIC ASSIGNMENT MODELS

F.1 Hypothetical Example

In order to test the effect of stop-&-go (S&G) waves on vehicle assignment, a hypothetical network was built in Aimsun. The network comprised 4 centroids (namely A, B, C and D), 4 nodes (node ids 362, 377, 355 and 370) and 16 links (all links are bidirectional). A dynamic scenario was built using the Stochastic Route Choice (SRC) model for traffic assignment and microsimulation model for network loading. The simulation period was set for 1 hour during which a demand of 1800 cars and 450 trucks was loaded from A to B. Shortest paths were computed every 10 minutes during the simulation period. Figure F.1 shows the snapshot of the hypothetical network where the length of sections is in metres.

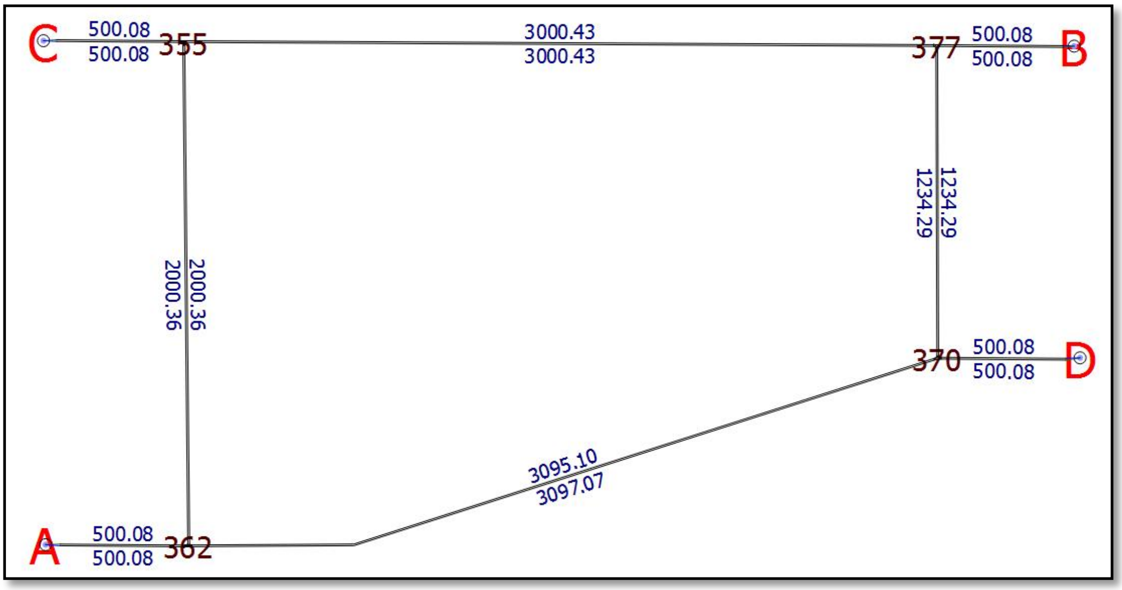

Figure F.1: The hypothetical network considered

There exist two possible routes in the given network, the upper route (A-362-355-377-B) and the lower route (A-362-370-377-B). Table F.1 shows the other route specific details.

Table F.1: Route specific details of the hypothetical network

Route Name	Node sequence	Length (m)	Signals on nodes	Total green time for movement (sec)
Upper	A-362-355-377-B	6000.95	355, 377	40+25=65
Lower	A-362-370-377-B	5331.52	370, 377	25+25=50

The table shows that the distance from A to B is greater while travelling through the upper route. In order to create S&G waves on the sections, traffic signals were placed at the nodes 355, 370 and 377. The green time of the signals on both the routes was adjusted so as to have a lesser green time on the lower route. This was done to increase the number of occurrences of S&Gs on the lower route, which was comparatively shorter in trip length.

F.2 Dynamic Scenarios Created

We created 2 dynamic scenarios for traffic assignment in this study. The two scenarios are discussed next.

1. Default case:

A dynamic scenario was created using the default link cost function. The default link cost function comprises link travel time (TT) only, i.e. $Z = f(TT)$. Similarly, default link costs were set for the initial, dynamic and K-initial shortest path functions as well.

2. Modified case:

A python script was written so as to define a new link cost function $Z = TT + \mu * no. of\ S\&Gs$. This script was linked to the dynamic scenario through the dynamic and K-initial shortest path cost functions. The parameter μ which represents the willingness to shift to another route was assumed to be 60 seconds per S&G for the study.

F.3 Quantification of S&G Waves

In Aimsun, the analyst can define two threshold speed limits, namely the queue entry speed and the queue exit speed. A queue entry speed is the lower threshold below which a vehicle is considered as stopped. The vehicle then remains in this stop condition until its speed goes beyond the upper threshold which is the queue exit speed. Figure F.2 shows the diagrammatic representation of the speed profile of a vehicle. The queue entry and exit speeds were set as 1 m/s and 4m/s respectively in this study. The duration of the stop is the difference between the queue exit and entry times. Aimsun considers this duration as one stop. For example, for the given speed profile, the vehicle underwent 2 stops, with the second stop still ongoing.

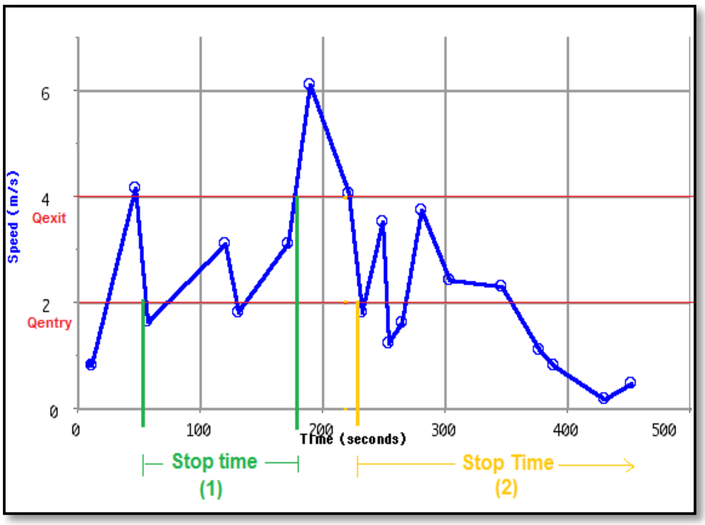

Figure F.2: S&G identification from the speed profile of a vehicle

F.4 Results and Discussion

Both the default and modified cases were executed and the following results were obtained. Figures F.3 and F.4 represent the percentage vehicles assigned during the 10 minute interval on the upper and lower routes respectively. Figures F.5 and F.6 depict the average number of S&Gs experienced on the upper and lower routes respectively.

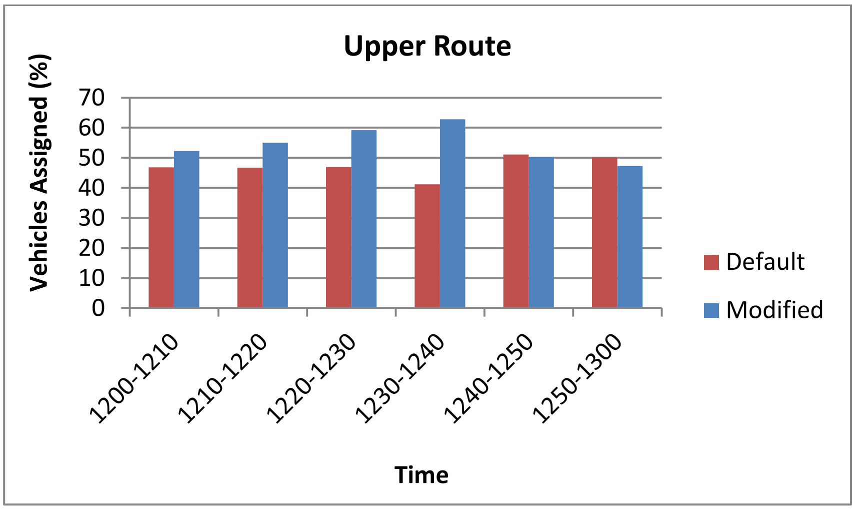

Figure F.3: Vehicles assigned on the upper route during 10 minute intervals

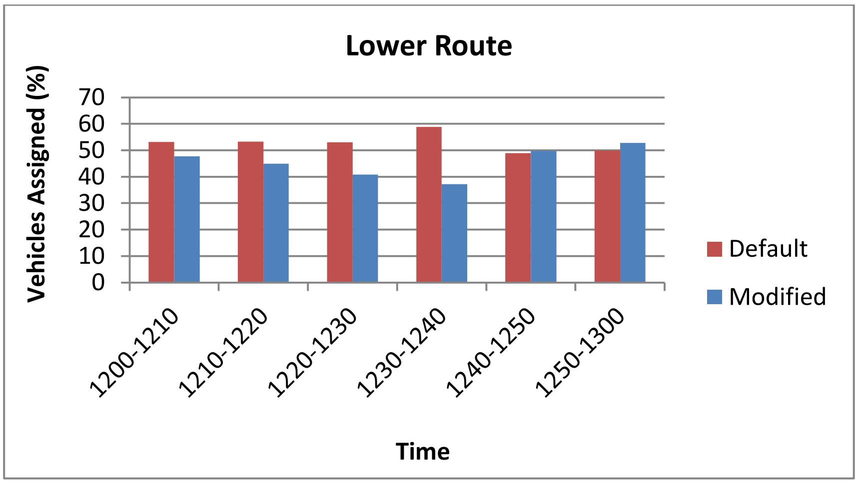

Figure F.4: Vehicles assigned on the lower route during 10 minute intervals

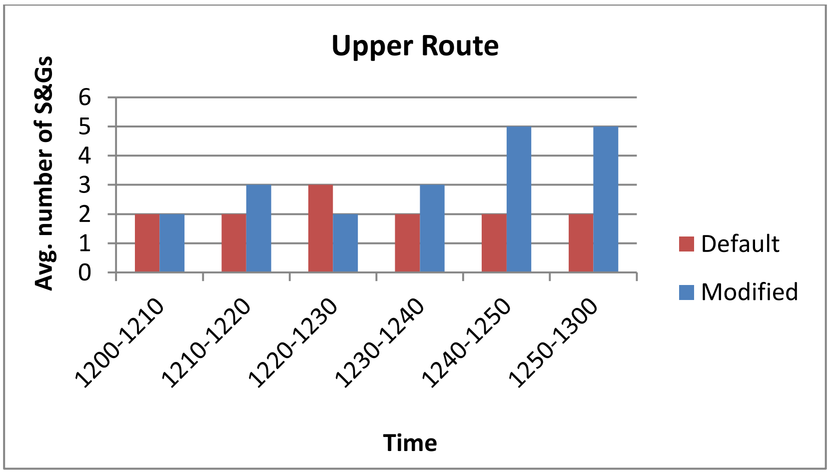

Figure F.5: Average number of S&Gs on the upper route in 10 minute intervals

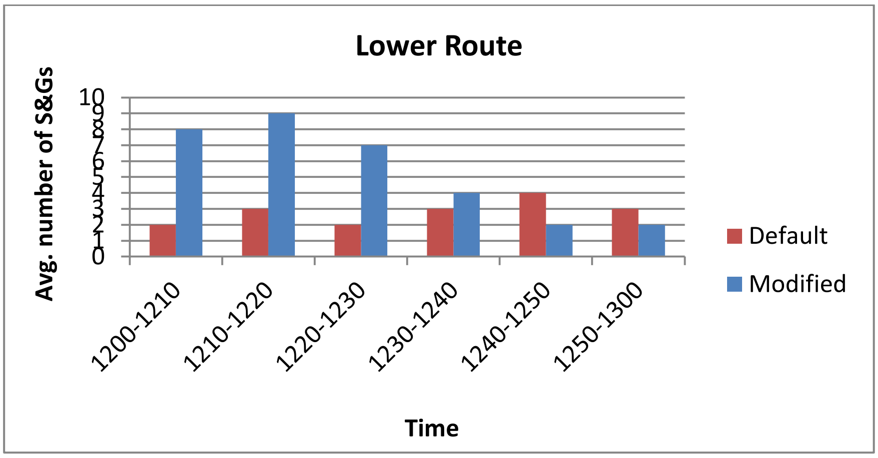

Figure F.6: Average number of S&Gs on the lower route in 10 minute intervals

From figures F.3 and F.4 it can be observed that path assignment was higher on the upper path during the first 40 minutes of the simulation. This is because there was a higher number of S&Gs experienced over the lower path, which can be seen in figure F.6. However, the percentage path assignment dropped a bit in the last 20 minutes of the simulation as the number of S&Gs increased on the upper path resulting in more vehicles assigned to the lower path. On the other hand, the default case had roughly similar percentage of vehicles assigned as those were determined purely on the basis of link travel time. Thus, it was observed from the results that a greater proportion of vehicles were assigned on the upper path during the first 40 minutes, even though it was longer but had fewer S&Gs existing over it. This indicates the validity of the hypothesis we were testing in this study. However, it was also observed that travel time was also increasing with an increase in the number of S&Gs. Future works will try to address this aspect of the model.