UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

EXPLORING THE ACCESS/EGRESS BEHAVIOUR OF RAPID RAIL PASSENGERS THROUGH THE APPLICATION OF NESTED LOGIT MODELS

DANIEL WATTS

A project dissertation submitted in partial fulfilment of the requirements for the degree of

MASTER OF ENGINEERING (TRANSPORTATION ENGINEERING)

In the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

UNIVERSITY OF PRETORIA

November 2020

ABSTRACT

The Gautrain is a rapid rail system in Gauteng, a province of South Africa. Currently, little is known about what causes the behaviour of Gautrain passengers concerning their choice of mode for their first- and last-mile trips. This thesis is a study of the first- and last-mile mode choice behaviour of Gautrain passengers. The study had three main aims. First, it aimed to develop the most accurate and statistically significant models of both first- and last-mile behaviour. Part of this aim was to try to understand the effect of non-traders on the stated preference data. To achieve this end, models were estimated with and without non-traders in the sample. Second, the study attempted to determine if there were any differences between commuters' first- and last-mile behaviour. For this purpose, a stated preference survey was developed and put online. It was then marketed on the social media platforms of the Gautrain Management Agency, the body that manages the rail system. The results of the survey were analysed and cleaned of errors. Different discrete choice models were derived from the data to find the best fitting model structure. The last aim was to see how the developed models could be used to plan future access and egress services. Significant discrete choice models were developed on a reasonably representative sample of the Gautrain population. The best-fitting model structure was a nested one for both first- and last-mile trips. These models show that while first- and last-mile commuter behaviour was similar, there were some differences. In particular, people are much more sensitive to in-vehicle time for their last-mile journey in comparison to their first-mile journey. This exercise showed that although non-trader data reduced the effectiveness and significance of the model slightly, the data did not change the overall picture shown by the models without non-traders.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

DECLARATION

I, the undersigned, hereby declare that:

- I understand what plagiarism is and I am aware of the University's policy in this regard;
- The work contained in this thesis is my own original work;
- I did not refer to the work of current or previous students, lecture notes, handbooks or any other study material without proper referencing;
- Where other people's work has been used, this has been properly acknowledged and referenced;
- I have not allowed anyone to copy any part of my thesis;
- I have not previously in its entirety or in part submitted this thesis at any university for a degree.

DISCLAIMER

The work presented in this report is that of the student alone. Students were encouraged to take ownership of their projects and to develop and execute their experiments with limited guidance and assistance. The content of the research does not necessarily represent the views of the supervisor or any staff member of the University of Pretoria, Department of Civil Engineering. The supervisor did not read or edit the final report and is not responsible for any technical inaccuracies, statements or errors. The conclusions and recommendations given in the report are also not necessarily that of the supervisor, sponsors or companies involved in the research.

Signature of student:	
Name of student:	**Daniel Watts**
Student number:	**13161513**
Date:	**22/11/22**
Number of words in report:	**30585** words

ACKNOWLEDGEMENTS

I wish to express my appreciation to the following persons who made this thesis possible:

a) Professor Christo Venter for his guidance, thorough editing and material support through the course of this study.

b) Gary Hayes for giving such good guidance throughout the process.

c) The GMA for generously providing the funding for this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1 INTRODUCTION

1.1 BACKGROUND

Gautrain is a commuter rail system in the Gauteng area that began operating in 2010. It has 10 stations in key commercial hubs between Johannesburg and Pretoria including one at OR Tambo International Airport. The north-south rail corridor runs between Park Station in central Johannesburg and Hatfield in Pretoria. The east-west line runs between Marlboro and OR Tambo International Airport. Fig1-1 shows the map of the stations. The users gain access to the station and departure stages of their journeys by several different access modes such as walking, parking at a station, being dropped off and the Gautrain feeder/distribution bus service.

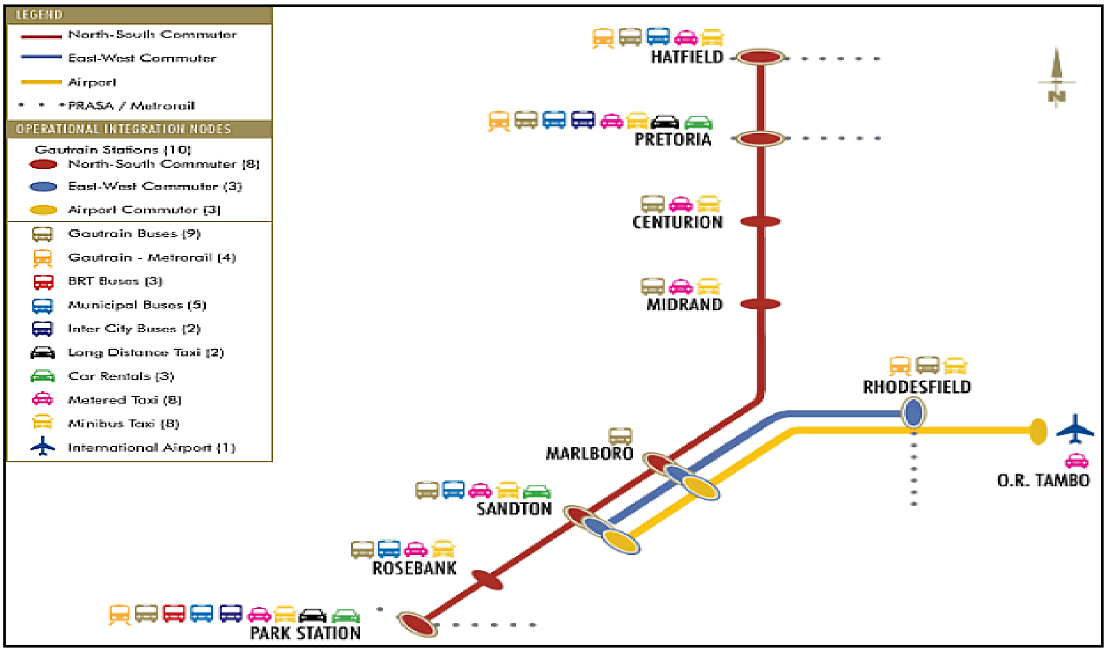

Figure 1-1: Map of the Gautrain System (GMA 2019)

The main rail system has operated efficiently for several years. Users experience an in-vehicle time saving of about 22 minutes (Gautrain Management Agency, 2017) compared to driving in peak-hour Gauteng traffic. Despite this efficiency, there are problems with the system, particularly when it comes to access and egress services to the system. For example, there is an over-demand for parking during peak periods at most Gautrain stations. Fig 1-2 shows the parking profile over the length of a day at Rhodesfield in 2014. The red line indicates the parking capacity of the station. Figure 1-3 shows the same profile at Centurion station.

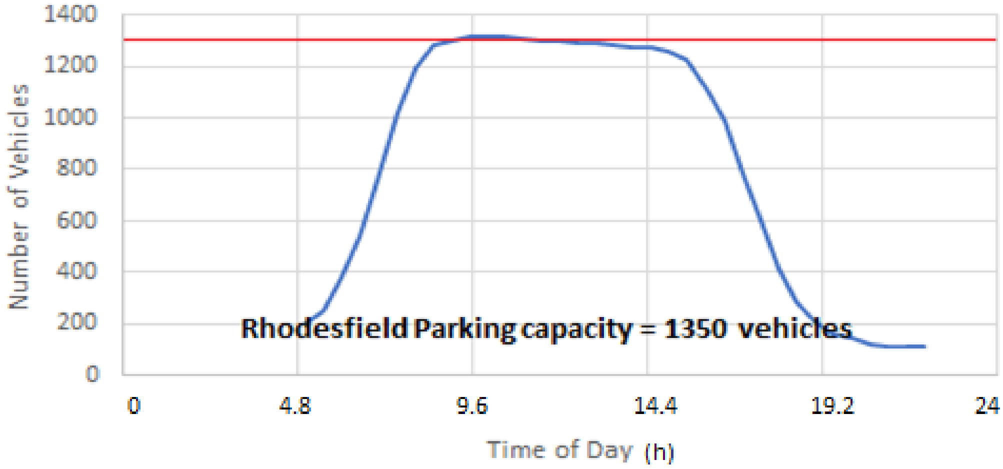

Figure 1-2: Profile of parking capacity at Rhodesfield station on a typical weekday

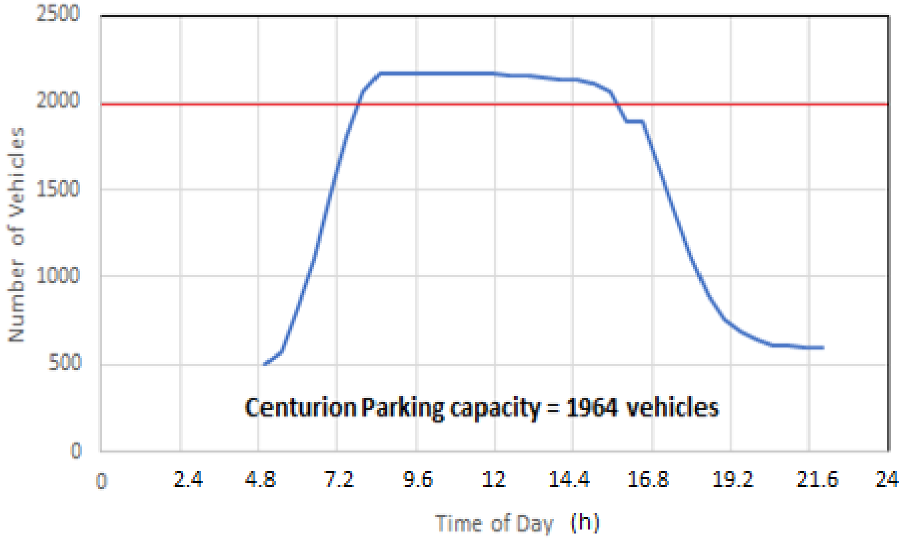

Figure 1-3: Profile of parking capacity at Centurion station on a typical weekday

Figure 1-3 shows that the Centurion station's parking is over capacity for nearly the whole day. The fact that it is over capacity shows that people are parking in spaces not meant for parking, suggesting an undersupply of parking. Figure 1-2 shows Rhodesfield is between slightly over capacity and slightly under capacity for most of the day. On the other hand, a focus group study for this research has shown that other access/egress services like the Gautrain bus are underused (see Section 3.1.1). These two factors suggest that the Gautrain service would have an interest in persuading commuters to switch from the park and ride facilities at the stations to using the Gautrain bus.

A survey by the Gautrain Management Agency (GMA) in 2017 revealed some aspects of the access/egress behaviour of Gautrain users. Of 660 people surveyed 54% of people accessed the station by driving a car, and 42% of people used a lift from a colleague (pickup) as their method of egress from a station. This behaviour occurs even though 62% of respondents said that their door-to-door trip distance (that is the first-mile journey, the distance in the train, and the last-mile journey) was less than 10 km. The fact that a significant number of commuters use a private car as access mode even for short journeys contribute to putting the parking supply under strain at several stations.

Problems like these illustrate the need for a model that could assist the GMA in understanding how Gautrain users make decisions regarding access and egress to and from the Gautrain stations. Considering the GMA’s plan to expand its rail network (Business Insider, 2020), an accurate access/egress model would be particularly valuable. To maximise ridership and by extension profitability, the GMA would need to offer the most attractive overall trip package. Understanding how Gautrain users perceive the costs and benefits of different access/egress combinations would help the GMA to design its facilities for different access/egress services as attractively as possible to the average traveller.

1.2 APPLICATION OF DISCRETE CHOICE MODELS IN ACCESS/EGRESS MODELLING

Various forms of discrete choice models based on random utility maximisation have been used to estimate passenger access mode choices for a main public transport system. The multinomial logit model and improved versions like the nested and cross-nested model are the most common model types (Hensher et al., 2005).

The literature describes various studies dealing with similar problems around the world. For example, in the Netherlands, a team at the Tienberg Institute developed a model to describe a commuter’s choice of a train station and corresponding access mode (Debrezion et al., 2007). In Taiwan, a nested logit model structure was used successfully to understand how people accessed the high-speed rail network (Wen et al., 2012). Research into public transport access has also been done in Israel (Bekhor & Shiftan, 2009).

The basis of transport choice models is data from passenger surveys. It is common to present survey respondents with a hypothetical choice scenario with several choices of transport modes for a trip. The aspects of each mode that are thought to govern a person’s choice are described in detail (see section 3). Typical attributes are travel time, waiting time and bus fare. In the choice scenario, the analyst captures the choices a respondent makes between the described modes. The survey data containing all the choices made across the sample of respondents are then used to estimate the final model.

Most prior studies that have been reported feature only the access mode and main mode choices. Only one study in the literature reviewed examined the egress mode. Given the scarcity of literature on egress mode behaviour, this dissertation's study of last-mile behaviour will contribute to the literature on this topic. First-mile and last-mile behaviour on the same system was also examined, allowing a comparison of first-mile and last-mile behaviour, which has also rarely been done.

A robust station access/egress survey of the Gautrain system could be used to develop a model that estimates the attractiveness of Gautrain user access/egress combinations. Currently, no such studies seem to have been done in South Africa.

1.3 PROBLEM STATEMENT

The Gautrain rail transit system needs more information on the access/egress behaviour of its passengers. As of yet, the system does not have any effective tools that allow it to understand what drives the decisions of its users' access and egress behaviour. By extension, it is difficult to provide appropriate access/egress services. To remedy this problem, this study used the discrete choice modelling technique on data from a stated preference survey, to develop a choice model that would give a reasonable understanding of the access/egress behaviour of Gautrain users.

1.4 STUDY OBJECTIVES

The main objective of this study was to develop a suitable discrete choice model that accurately described the access/egress behaviour of Gautrain users. A secondary objective of the study was to use the model to examine predicted changes in user behaviour when the GMA adjusts variables within its control.

1.5 SCOPE OF THE STUDY

The study conducted a passenger survey among Gautrain users using an online platform that fields responses from passengers across all 10 stations. The focus group mentioned earlier, and other surveys have shown that Gautrain users are among the wealthiest members of South African society, and most frequently online. A sample size of 400 completed surveys was obtained, which is adequate for this type of study, especially when considering that each respondent produced between 8 and 12 choice responses or data points. The study used data from the survey to develop the best-fitting logit model structure. Although the study explored multinomial logit (MNL), and nested logit model forms, it did not include mixed logit models. The scenario testing exercise was limited to finding strategies for increasing feeder bus usage and the effect of parking fees on user access behaviour.

1.6 STUDY METHODOLOGY

The study applied the following methods to fulfil the objective:

- The literature on random utility theory and the different types of discrete choice model structures was reviewed. Relevant transport choice modelling literature was examined, followed by a brief overview of the literature on different types of survey data and techniques.
- Two stated preference fractional factorial experiments were designed, one for access trips to Gautrain stations and one for egress trips from Gautrain stations to passenger destinations. Sections collecting relevant respondent demographic data were added to the end of the survey. The experiments were loaded onto the Survey Monkey online platform. A small pilot sample was run to make sure it was accessible to the average Gautrain user. The GMA advertised the survey in November 2019 and a sample of about 400 people responded. After cleaning the data, the sample was modelled using N-Logit software to find the best fitting choice models.
- The estimation of various discrete model types is described in detail. The process started with a simple MNL to investigate the forms of the utility equations. Then various nested logit structures were explored to find the optimum model structure. The main criterion for determining model optimisation was the log likelihood. Secondary criteria were the Akaike information criterion, the Bayesian information criterion and the ρ^2 value (Hensher et al., 2005).
- The model explored the findings on Gautrain user behaviour using scenarios to determine how the GMA might increase feeder bus usage and reduce parking demand.

1.7 STUDY OUTLINE

The report is structured as follows.

Chapter 1 introduces the study topic, the problem statement and the relevant background information to the study.

Chapter 2 is a literature review of the topic. It covers random utility theory and different types of choice models with their relevant strengths and weaknesses. It examines the literature on choice modelling in access/egress studies and finally, it looks at research relevant to first-/last-mile travel that does not involve choice modelling.

Chapter 3 explains the design of the online survey and its choice experiment. It explains the theoretical foundation on which the stated choice experiment was designed. It describes each question that the conceptual design of the survey had to answer and how the final design was reached. The technical aspects of the survey are explained, such as the level coding for each experiment, the assumptions and calculations for each attribute’s level and the online tool used to put the survey together. It presents the rationale for hosting the survey online and discusses how the sample was attracted to take the survey. Finally, the chapter gives a description of the sample size and its demographic split, comparing it with a sample from a GMA market segmentation report two years previously as a calibration exercise.

Chapter 4 presents the results of all the models estimated on the data and discusses some of the trends that emerge in them.

Chapter 5 presents the mean elasticates of the attributes tested in Chapter 4. It also proposes a hypothetical scenario to explore different strategies that the Gautrain management might use to optimise the system's access and egress services.

Chapter 6 provides conclusions and recommendations going forward in researching access and egress behaviour in the Gautrain system

2 LITERATURE REVIEW

Chapter 2 reviews the relevant literature, starting with a discussion of random utility theory and how it develops into discrete choice models. It then discusses the different types of choice models and the history of choice modelling in transport. The review ends with a short discussion of studies that deal with walking to public transport, and the way the safety of a walking environment is dealt with in a survey.

2.1 RANDOM UTILITY THEORY

Section 2.1 reviews the underlying theory of random utility theory or choice modelling. It explains the basics of the formulation of the MNL model and the model's limitations. These limitations can be overcome by nested logit and cross-nested logit models. The theories discussed in this section primarily refer to the work of Hensher et al. (2005) and De Dios Ortuzar & Willumsen (2014). Where other references are used, they are given.

2.1.1 The foundational principles of random utility theory

Choices revolve around trade-offs; a person selects a choice by weighing up various attributes. In transport mode choice, typical trade-offs involve choosing between the time savings provided by faster modes, or the higher costs of using them. For example, a person might weigh the time of a trip by car (Johannesburg to Durban) against the cost of a trip by plane. It is helpful to understand and be able to predict mode choices in transport planning. Because all the attributes were in different units (cost, severity, distance, time), it was difficult to model them mathematically. By defining each attribute as having a certain *utility*, a solution to this problem was developed (Hensher et al., 2005).

A critical limitation of the random utility theory is that there are aspects of a choice in a survey that may not be observable to the analyst. Hence, there will always be criteria applied to choices that cannot be quantified in a survey. An element of the total utility will therefore always be random and unknown. For example, a respondent might suffer from claustrophobia and be reluctant to fly, making them less likely to choose air transport even if they valued saving time more than saving money.

In practice, people assign a particular utility to everything that goes into making a choice. For instance, spending R5 000 on a plane ticket represents x units of utility, whereas travelling 12 hours represents y units of utility. A fundamental assumption of random utility theory is that people are rational. They will choose the option that maximises their knowable utility. This assumption implies that if $x>y$ (assuming there are no unknowable random utility elements in the decision), the person will fly; conversely if $y>x$, the person will drive.

2.1.2 Formulation of the multinomial logit model

The utility of one alternative of a choice is comprised of a knowable component and an unknowable part. This concept is expressed mathematically in Equation 2-1.

$$\boldsymbol{U_i = V_i + \varepsilon_i}$$

Equation 2-1: The two components of the utility of an alternative

Equation 2-1 states that the utility ($\boldsymbol{U_i}$) of an alternative *i* in a set of *j* alternatives has a quantifiable part ($\boldsymbol{V_i}$) and an unquantifiable part ($\boldsymbol{\varepsilon_i}$) (Hensher et al., 2005).

Further, Hensher et al. show that V_i can be represented as in Equation 2-2.

$$V_i = \beta_{0i} + \beta_{1i} f(X_{1i}) + \beta_{2i} f(X_{2i})$$

Equation 2-2: Components of known utility

Where:

β_{ni}: Represents the attribute-specific coefficients estimated during a choice experiment. This constant's purpose is to convert an attribute to utility/disutility. Its units are expressed in terms of utility per unit attribute. For example, an experiment could estimate the travel time constant of a model as –0.2 utility/minute; in the application of this estimated model, everyone minute of travel time in an alternative translates into –0.2 units of utility.

β_{0i}: Represents the alternate specific constant (ASC). Its purpose is to recapture the effect of the unknowable utility in a choice model. After the attribute-specific coefficients are estimated, the ASCs are estimated to improve the rate at which the model accurately predicts the alternative chosen in the choice data. For a given choice model, the ASCs for the set of alternatives are measured relative to each other. One alternative in a choice set is given a fixed value (for example, 0). Then every other alternative's ASC is measured relative to the ASC fixed for estimation.

$f(X_{1i})$: Represents the generalised representation of the quantifiable attributes in the expression of the utility of an alternative. The attributes are represented as a function because attributes are sometimes expressed non-linearly in choice model estimation. For instance, a modeller may sometimes choose to express the travel time in log form. Hensher et al. (2005) explain the formation of an MNL as follows: one starts with the following idea (expressed mathematically) – the probability that a user will choose alternative *i* from alternatives (*i*, *j*) is equal to the probability that the utility of alternative *i* is greater than the utility of alternative *j*.

This statement is expressed mathematically in Equation 2-3.

$$\boldsymbol{P_i = P(U_i \geq U_j)}$$

Equation 2-3: MNL derivation Part 1

Substituting Equation 2-1 in Equation 2-3 yields Equation 2-4:

$$P_i = P\left[(V_i + \varepsilon_i) \geq \left(V_j + \varepsilon_j\right)\right]$$

Equation 2-4: MNL derivation Part 2

It is useful to rearrange Equation 2-4 since the uncertain part of utility is unmeasurable in the form of Equation 2-5.

$$\boldsymbol{P_i = P\left[\left(V_i - V_j\right) \geq \left(\varepsilon_j - \varepsilon_i\right)\right]}$$

Equation 2-5: MNL derivation Part 3

To put this into words, the probability that a user will choose option *i* is equal to the probability that the difference between the observed utilities of *i* and *j* is greater than or equal to the difference between the unobserved utilities of *i* and *j*.

The derivation of the standard MNL model assumes that the unknown utilities are independently and identically distributed according to the EV1 (extreme value type 1, Gumbel distribution). This assumption is known as the IID assumption. The full proof can be found in Hensher et al. (2005) and results in the expression in Equation 2-6.

$$\boldsymbol{P_i = \frac{exp^{v_i}}{\sum_{j=1}^{J} exp^{vj}}}$$

Equation 2-6: Probability of an MNL alternative

Expressed in words, Equation 2-6 states that the probability of a respondent's choosing alternative *i* is the exponent of its known utility divided by the sum of the exponents of the known utilities of all the alternatives in the choice set.

2.1.3 The limitations of the MNL model

The way the MNL is derived leaves it with a limitation, known as the independence of irrelevant alternatives (IIA), which states that the ratio of a pair of choice probabilities is independent of the presence or absence of other alternatives. This assumption has the significant implication of making all alternatives equally similar or dissimilar. Therefore, all the information in the unknown components of utility distributed across alternatives is identical in quantity and relationship. Furthermore, the unknown utilities are assumed to be independent and identically distributed according to the Gumbel Type I distribution. In specific scenarios, this assumption is fair; for example, a retired couple may only consider one or two places in which to live. However, the assumption is unrealistic in that it does not describe a choice realistically, especially in transport (Cushing & Cushing, 2007).

A well-known example for illustrating the limitation of the IIA assumption is the red bus/blue bus problem. Consider a town with two transport modes, namely a car and a red bus, with a modal split in an MNL of 0.5/0.5. If a competing bus service were introduced, which ran blue buses across the city, the model would maintain the ratio of car users to red bus users. That is, the modal split would become 0.33/0.33/0.33. However, the blue bus service would draw more users from the red bus service than from the pool of car drivers, making a mode share of 0.33/0.33/0.33 unrealistic. This example shows that the correlation between bus services is not accounted for in the MNL.

There are certain steps an analyst can take to reduce the risk of violating the IIA. They could pick attributes in such a way that they maximised the known utility components in a model. The practice of including socio-demographic attributes such as gender and income in a model is useful because these attributes can discern some unobservable elements of utility that might correlate across alternatives.

The other way of approaching this problem is to use a model type that relaxes the IIA constraints to allow some correlating components of utility in the unknown utility component of the equations. The two types of model applicable to this study are the nested logit model and the cross-nested logit model.

2.1.4 Nested logit model

The nested logit model is a way of accommodating choice scenarios that do not adhere to the IID/IIA constraints. It allows for some correlating components of utility by grouping the alternatives that correlate into nests. For example, in the problem examined in this study, people who take a feeder bus to a Gautrain station and people who walk to a Gautrain station both face a certain amount of walking through the city streets. It is a common perception that walking in city streets at certain times of the day is not safe. If respondents feared for their safety, their fear would be a disutility that appears in the equation for the bus and the equation for walking. If that fear were not captured in the survey, it would be transferred to the unknown utility component and violate the IIA/IID constraint. However, a nested logit model estimates that correlation and uses a nesting parameter for the correlating alternatives. The nesting parameter increases the magnitude of the known utility coefficients. Doing this increases the magnitude of the known utility. The increased magnitude of the known utility helps to maintain the condition specified in Equation 2-5 (Hensher & Greene, 2001)

For this reason, a nesting parameter can only be between the values of 0 and 1. At 1, the known utility components decrease and it collapses into an MNL model. Figure 2-1 shows an example of a theoretical nested logit model.

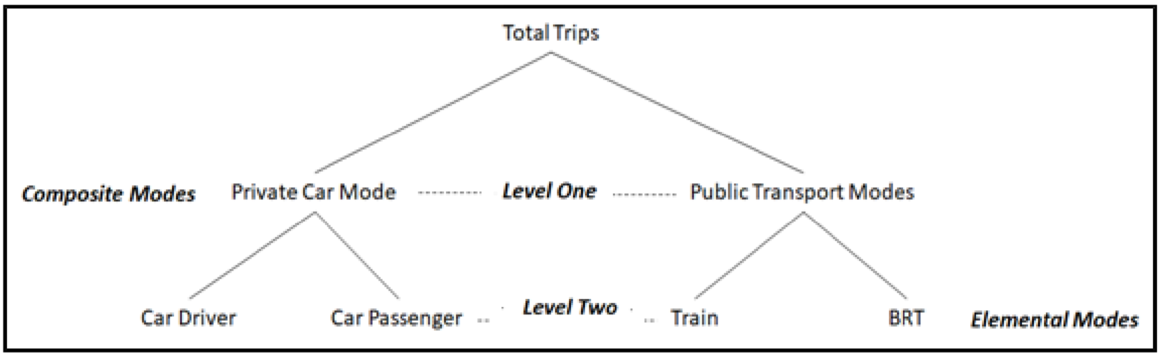

Figure 2-1: Theoretical nested logit model

One should note that at Level 2, the modes are not correlated and thus make an MNL. Further down, at Level 1, each composite alternative does have some cross-correlation with the other composite alternative. Level 1 therefore has a nesting structure. It is also essential to note that a nested logit model is not a sequential choice model, although it may seem as if a person chooses between a private car and public transport and then between the two alternatives in whichever nest they chose. However, the model has been constructed by grouping the alternatives that have a high correlation together.

The evaluation of the choice probabilities of alternatives in a nested logit model is straightforward. The choice probability is the probability of choosing a nest and the probability of choosing an alternative in the chosen nest. This process is described below. Note that the rest of the section in the equations and the text refers to Figure 2-1.

$$L_{pt} = ln \sum_{j \in Spt} e^{\frac{V_j}{\lambda pt}}$$

Equation 2-7: Log sum of a nest

Equation 2-7 describes the evaluation of the log sum of a nest. The log sum of a nest is a parameter that is used in determining the probability of a respondent choosing a particular nest. The λpt symbol represents the nesting parameter of the public transport nest (pt). V_j represents the known utility of an alternative in the public transport nest. Equation 2-8 shows how to use the log sum of each nest to work out the probability of each nest, namely the probability of choosing a nest at Level 1 of the model.

$$P_{Spt} = \frac{e^{\lambda_{pt} L pt}}{\sum e^{\lambda_n L n}}$$

Equation 2-8: Probability of choosing a nest

N represents each nest in the model. Finally, the probability of alternative j in the pt nest is evaluated in Equation 2-9.

$$P_{j|pt} = \frac{e^{\frac{V_j}{\lambda pt}}}{\sum_{j \in Spt} e^{\frac{V_j}{\lambda pt}}}$$

Equation 2-9: Probability of choosing an alternative in a given nest

These equations together are used to evaluate the nested logit model. While the nested logit model is useful, it does have some limitations. It does not allow for randomisation of parameters to account for individual heterogeneity or choice heterogeneity like the mixed logit model. The more alternatives one has, the more nesting structures one must test to find the best fitting model, and this can be very time-consuming. The nested logit model cannot handle alternatives that correlate with two uncorrelated nests. For example, if Uber were part of the model in Figure 2-1, it could fall into both nests, but the model does not accommodate that. The solution to this problem lies in the cross-nested logit model.

2.1.5 Cross-nested logit model

As stated, one main limitation of the nested model is its inability to handle alternatives that belong to two different nests. This problem is expressed mathematically in the following equations.

$$\boldsymbol{Corr(\varepsilon_1, \varepsilon_2) > 0}$$

Equation 2-10: Condition 1 necessary for cross-nesting

$$\boldsymbol{Corr(\varepsilon_2, \varepsilon_3) > 0}$$

Equation 2-11: Condition 2 necessary for cross-nesting

$$\boldsymbol{Corr(\varepsilon_1, \varepsilon_3) = 0}$$

Equation 2-12: Condition 3 necessary for cross-nesting

ε_2 represents the unknown utility of an alternative in a choice set. It has some correlation with the first and third alternatives. However, the first and third alternatives do not correlate with each other. To overcome this problem, researchers developed the cross-nested logit model. Its design is similar to that of the nested logit model. However, it introduces the possibility of an alternative in both nests, described by the nest inclusion factor (α) which (using equations 2-10 to 2-12) describes how much Alternative 2 correlates with Alternative 1 and how much Alternative 3 correlates with Alternative 2. It is useful in adjusting the nesting factor when working out the probability of choosing that alternative. Figure 2-2 shows an example of a model structure that exhibits cross-nested behaviour.

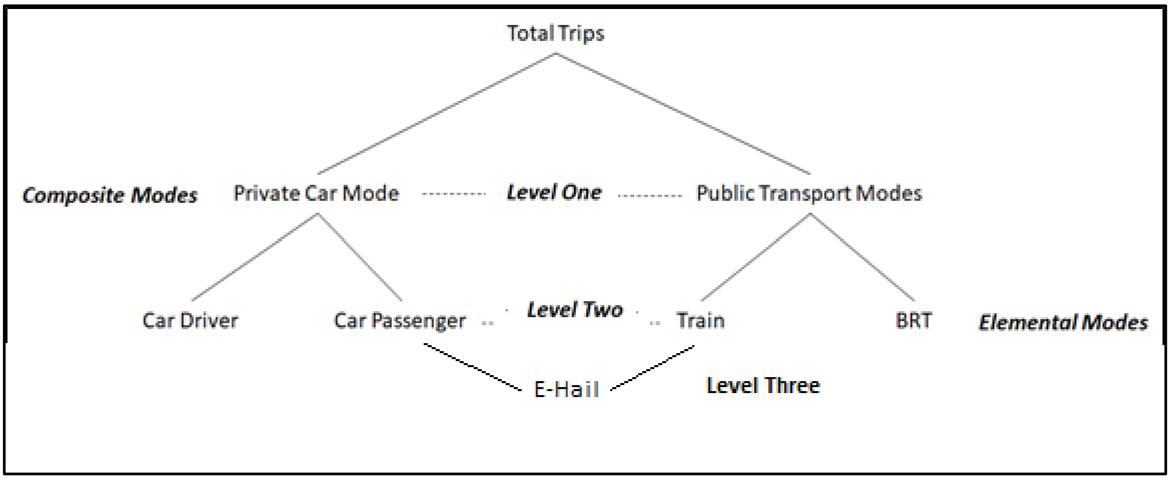

Figure 2-2: Modified model schematic to show cross-nesting behaviour

In this example, α would denote if E-hail correlated more strongly with the private car mode nest than the public transport nest or vice versa. Determining the probability of choosing an alternative in a cross-nested model is like the evaluation of the probability of an alternative in a nested logit model. However, α is included in the equation for the cross-nested alternatives only. The probability of choosing a nest that has a cross-correlated alternative is shown in Equation 2-13.

$$P_{Spt} = \frac{(\sum_{j\in Spt}(\alpha_{eh,pt} e^{V_{eh}})^{\frac{1}{\lambda pt}})^{\lambda pt}}{\sum_M(\sum_{j\in M}(\alpha_{eh,M} e^{V_j})^{\frac{1}{\lambda M}})^{\lambda M}}$$

Equation 2-13: Probability of choosing a nest with a cross-nested alternative

Where M is the total set of nests, and j represents the set of alternatives in a nest, V_{eh} represents the known utility of the E-hail alternative and $\alpha_{eh,M}$ represents the nest inclusion factor for E-hail per nest, and $\alpha_{eh,pt}$ represents the nest inclusion factor in the public transport nest. Equation 2-14 shows how to calculate the probability of choosing E-hailing.

$$P_{eh|pt} = \frac{(\alpha_{eh,pt} e^{V_{eh}})^{\frac{1}{\lambda pt}}}{\sum_{j\in pt}(\alpha_{j,pt} e^{V_j})^{\frac{1}{\lambda pt}}}$$

Equation 2-14 – Probability of choosing an alternative

One should note that other than the capability for cross-correlation, the cross-nested model has all the other limitations of the nested logit model.

2.1.6 Evaluation of discrete choice models

There are multiple measures of evaluation of discrete choice models that are used in the model. The log likelihood (LL) value describes the predictive power of the model. For instance, assume a model that describes 100 choices between option *x* and option *y*. If option *x* has a 70% probability of being chosen and option *y* has a 30% probability of being selected for choice scenario 1, and the person chooses option *y*, then the log of 0.3 is taken. This process is repeated for all the choices in a model, and the total LL is summed. The less negative the LL, the better a model is at predicting the choices of people. It is important to note that LLs can only be compared between models of the same population.

The LL parameter is useful for assessing if nested models and cross-nested models are valid for a given data set. If a nested or cross-nested structure has a more negative LL value than the LL value of an MNL model, the nesting/cross-nesting structure is not valid.

The second parameter that evaluates the strength of a discrete choice model is the rho square value (ρ^2). This value denotes how much of a model is described by a known utility. It ranges from 0 to 1; at a value of 0 the model is entirely random and dominated by unknown utilities, and at a value of 1 the model is entirely deterministic and dominated by known utilities.

The third parameter is the chi-square value. This parameter evaluates the model attributes as a group. Model estimation is an iterative process in which the starting values for the model parameters are assumed (typically they are assumed to be 0). The model estimates an LL value for when all the parameters are at 0. That value is the LL when the model of the data is random, meaning it gives an equal probability for any alternative no matter the scenario. When the model reaches convergence, it calculates a final LL value, and this is generally less negative than the starting LL value. A chi-square test is performed on the difference between the final LL model and starting LL model if the difference is significant within a 95% confidence interval.

Another parameter measured is the Akaike information criterion (AIC). This parameter compares the LL in relation to its complexity. It is defined in Equation 2-15.

$$\boldsymbol{AIC = -2LL(\theta) + 2K}$$

Equation 2-15: Akaike information criterion

In Equation 2-15, K refers to the number of estimated parameters in the model and $2LL(\theta)$ refers to the estimated LL.

Another important parameter for evaluating models is the t-test statistic. It is used to evaluate the significance of an attribute coefficient. The t-test statistic gives the probability that the β coefficient is different from a starting estimate. In practice, a 95% probability critical value is adopted, so β parameters with a ratio greater than or just under 1.96 (the value at which it is 95% certain the coefficient differs from its starting estimate) are accepted and anything less is rejected.

The t-test attribute is an important assessment tool of nested and cross-nested discrete choice model types. The t-test applied to β parameters is also applied to the λ parameter and the α parameter. If the lambda or alpha values fails the t-test ($t < 1.96$), then the nesting/cross-nesting structure applied is not viable.

The final tool in assessing the validity of a discrete choice model is a rationality test. If the β values are illogical (for example, if travel costs have a positive β value, implying that the more people value an option the more they pay) the model proposed is not useful.

2.2 CASE STUDIES OF DISCRETE CHOICE MODELLING IN TRANSPORTATION PLANNING

There is a long history of the use of choice modelling in transport engineering. This section reviews the use of choice modelling in studying access mode choice.

Wen et al. (2012) developed a latent class nested logit model to describe an access mode choice for the Taiwan High-Speed Rail network. The latent class model is a way of accounting for the segmentation of the data sample.

Segmentation is the process of splitting the data along various socio-demographic lines, occurring because different parts of a population have different sensitivities to different variables in a choice (high-income people might be more sensitive to time; low-income people might be more sensitive to cost). Developing models on a segmented data sample captures some of the individual heterogeneity across users of a system. In most modelling work, the segmentation of data is arbitrary. For example, people estimate the income range of respondents and divide the data into several segments based mainly on the modeller's judgement. This practice can often lead to inaccuracies or imprecision in the data. In contrast, the latent discrete choice model estimates the number of segments and the corresponding segment boundaries as well as the utility coefficients. In doing so, it removes some of the arbitrariness of data segmentation.

The latent choice model of Wen et al. (2012) showed a dramatic variation in attribute sensitives (where sensitivity refers to the level of change in behaviour given a level of change in a particular attribute) across segments. It had important implications for the modelling of data from a South African context. Different areas have different demographics; for example, one train station in a rail system might have students and academics in its catchment area. The catchment area of another station in the same rail system might serve mostly professional people. There are likely to be significantly different preferences between the two stations. Therefore, given a big enough data sample, it would be interesting to develop a model for each station location.

The study published by Wen et al (2012). has several interesting aspects. The authors managed to capture correlation across choice alternatives using a nested structure in conjunction with the latent style model. The most effective nesting structure they found was nesting car-based alternatives in one nest and public transport alternatives in the second nest. It suggests that this nesting structure (public transport alternatives vs private alternatives) is likely to be part of the foundational nesting structure of the models developed in the present study. Also, Wen et al. (2012) found that people were significantly more sensitive to the cost of accessing the station than the time it took to access the station. Interestingly, their best-fitting model divided the parking fee for the car trip and the petrol cost income. Lastly, the model also divided access time by distance travelled (to account for traffic) as it seemed to fit the data better.

Debrezion et al. (2007) performed a different kind of access mode choice study in the Netherlands. They built a nested model that combined the choice of departure station in conjunction with access mode. The Netherlands has a much larger public transit rail network than anything available in Africa at the moment. Stations connect to different destinations; they also have different levels of public transport supplying them, parking places and bike-locking facilities. This situation added a layer of complexity to the problem in that they could not use trip distance to the station as the sole decision criterion. Also, departure station choice affected the choice of access mode, hence the factoring of both into the study. They developed a rail service quality index to quantify the attractiveness of each station and then incorporated that into a nested logit model with access mode choice.

However, their formulation of this had some interesting design characteristics. They found bike-locking facilities to have a significant effect on people's choice of station. In the South African context, variables like bike parking may not be important. It shows how context-specific choice modelling experiments tend to be. Also, instead of using the waiting time for public transport, Debrezion et al. (2007) used frequency of service, as headway captured the same disutility (the inconvenience of the wait). Still, unlike waiting time, headway is not a random variable, so it is much easier to estimate accurately.

Unfortunately, the only service that had regular headway for consideration in this study was the Gautrain bus service. The prevalent traffic congestion detracted from the reliability of the Gautrain buses' headway timetables. The problems with the Gautrain bus service were reiterated in the focus group research carried out by Nielsen (2019).

Debrezion et al. (2007) found that in the Dutch context, car availability did not significantly influence people's choice of access mode to the station. The country is small, and if people choose to go by train, it means they generally are not interested in taking their car for a brief journey. In Gauteng, however, people make much longer total journeys than in the Netherlands. According to a survey carried out in 2017 by transport analytics firm INRIX, South Africa is ranked 8th in average peak hours spent in congestion. The Netherlands, by comparison, is not even in the top 20 countries. The INRIX survey also showed that Gauteng was among the most congested areas in South Africa. Johannesburg is ranked second in the list of hours spent in congestion, while Pretoria is ranked fourth. Also, the average travel distance to and from stations is much greater than in a small place like the Netherlands. Given the different conditions of travel and the more limited public transport services, car availability is likely to play an important role.

Debrezion et al. (2007) also made a significant assumption in developing this study, namely that the survey respondents had already decided to go by train. The authors chose to ignore the primary mode choice completely. While this is not entirely realistic, it still led to feasible results. Also, as they were only able to gain data on train users, it would be irresponsible to make a primary mode choice model with no other primary mode users. Little data is available on the primary mode choice of the segment of the Gauteng population that uses the Gautrain. The survey conducted by the author for this study only allowed for the surveying of people who already took the Gautrain. Thus, the study also ignored the primary mode choice of the respondents (whether they chose to go by private car or train). The study only considered their first- and last-mile choices, given that they had already chosen to use the Gautrain.

Interestingly, Debrizion et al. (2007) did not divide the distance by time or cost by income in the Dutch study, unlike the study by Wen et al. (2012) in Taiwan. All the utility functions were also linear; it was done to simplify the model building process.

With the gathered data, Debrizion et al. (2007) were able to graphically plot the relationship between the utility of a mode and the distance from a station (an approximated linear model) in Figure 2-3. The results were interesting. Non-motorised forms of transport started with the highest utility of all (bicycle and walking), but decreased sharply while car and public transport only suffered a gradual decrease over time. It seems quite logical, but in our context, the factoring of safety concerns would affect such relationships, causing a steeper decreasing slope on all the modes that require a walking component to the journey.

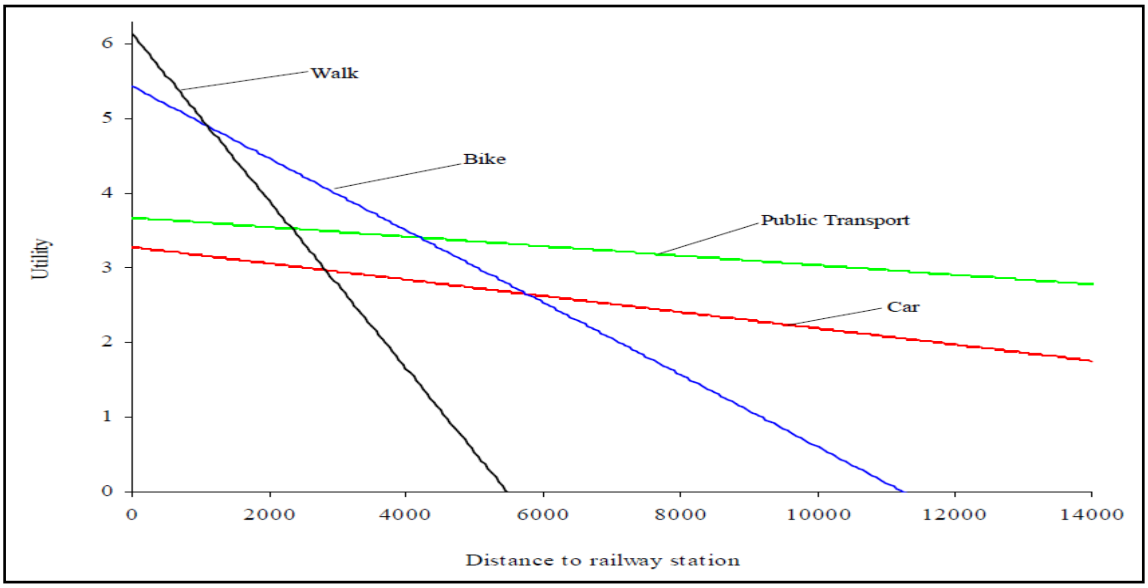

Figure 2-3: Relationship between the utility of access mode and distance to the railway station (Debrizion et al., 2007).

One thing that is strikingly apparent in the literature on access mode choice studies is the lack of research relating to egress mode choices or last-mile studies. However, one study carried out by Koh et al. (2016) in Singapore implemented a choice model that tried to describe the choice of rail commuters on the last mile of the journey.

Models sometimes have situational variables. In Singapore, bicycle travel featured strongly, so much so that it became one of the significant choice alternatives. The number of bicycles lanes surrounding the stations had a significant effect on the model. Infrastructure such as bicycle lanes and wide and well-lit sidewalks at stations were important variables for non-motorised transport modes.

A second study in the Netherlands by Giovanni & Rietveld (2007) found that car availability did not influence people's choice of access mode to the station, which is a sensible choice in the European context. The Netherlands is small, and if people choose to go by train, it means they are generally not interested in taking their car for a very short journey. In the context of the Gautrain, however, the converse applies. Parking a private car at the station is one of the most common access modes among people with cars in Gauteng.

A study in San Francisco by Cervero (2001) developed a binomial logit model on last-mile data from rail stations to test the effect of the urban environment on whether or not people walked to their destination. It found that while the distance to their destination from the station was the most significant factor in a person's decision to walk as their last-mile trip, the design of the urban environment also played a role. The number of sidewalk miles relative to street miles was the second most significant variable, followed by the median sidewalk width.

While the San Francisco study had binomial logit models for both first-mile and last-mile trips to rail transit stations, it offered a scant comparison of the two. It used variables in the access model that did not appear in the egress model.

Cervero's (2001) findings contain insights on the effect of gender. Women tended to prefer public buses over cycling or walking. Given South Africa's much higher safety concerns, the effect of gender is likely to feature strongly in Gautrain access/egress data. Also, Cervero found that as people's income went up, the chance of taking the bus became less likely. This observation might be due to the increase in car availability among higher-income demographics.

In all these studies, access and egress preferences were not compared. The lack of comparisons leaves unanswered the question of how to handle data on access and egress mode choice (of the same public transit system) concurrently. Given how the models were separate in all these studies, suggested that the best way to handle the data for the models developed in this author's study is to split it into two separate choice models. However, it is unclear if there will be a significant difference between access and egress trip preferences and sensitivities.

While this section has focused on choice modelling in first- and last-mile trips, the technique has been applied to a much broader variety of situations. The behaviour behind a person's choice of an airport has been explored in papers such as Ashford and Benchemam (1990), Yang et al. (2014), Chia-Wen et al. (2012), and Loo (2008). Discrete choice behaviour has been used to model primary mode choice for daily commutes (Cervero & Gorham, 1995; Bhat, 1997).

2.3 FACTORS THAT AFFECT PEOPLE'S WILLINGNESS TO WALK TO THEIR TRANSIT STATION

"Walkability" is the term used to describe what factors make people more likely to choose walking as their trip mode. Several studies focus on defining walkability in the context of first-/last-mile trips. Some of these studies are discussed in this section.

Agrawal, Schosslberg and Irvin (2008) outline three factors that people consider when trying to define and quantify walkability. They are density, diversity and design (colloquially referred to as the "three Ds"), although analysts tend to focus on the first two points. The literature shows that people seem to value design as the least significant factor of the three. A study by Cervero (2001) is a good example, showing that while urban design played a noticeable role, its effect on walkability was less significant than that of the distance to the station and the amount of mixed-use development around the station. Given the latter two, people might elect to walk to their access station if a shop were conveniently located on their way home from work.

Cervero (2001) analysed a survey of over 300 pedestrians in California, and the data had some interesting findings on what increased the walkability of a public transit station. Firstly, the density of development around a transit station significantly increased the walkability of a station. Increasing urban density reduces the average required walking distance to the station in the catchment area. The average distance people walked to a station was about a kilometre.

One of the questions posed to respondents was "What are the main reasons why you chose your route today?". The responses raised an interesting point. First, safety was the second most important priority in route choice after distance to the station. Security is even more critical in the South African context. According to Numbeo's (2019) global crime level index, South Africa is 3rd in the world, while America only ranks 50th in the world of crime levels. Given the local crime levels concerns for safety were factored into the design of the current study's survey.

The results of this question of route choice also seem to support the findings in Cervero's study, and environmental attractiveness was the fourth most significant variable in determining why respondents took their route that morning. However, each person might define environmental attractiveness differently, therefore making the finding a little more tenuous.

Lastly, the survey respondents were asked, "Below is a list of factors that other researchers have found to influence the routes people walk along. For each one, please mark how important it is to you." Again, the distance was the most important factor (82% of respondents listed it as very important). About half of the respondents considered traffic conditions as the second most important factor, mentioning how responsibly people drove, and traffic devices (like signalised intersections). Environmental factors were considered especially important by about 30% of respondents.

Ryan and Frank (2009) conducted a study of the effect of the surrounding environment on the tendency of people to walk to the bus station. The authors constructed a walkability index based on environmental data gathered from the San Francisco area. The authors collected data on parameters such as residential density, vacant lot concentration, intersection density, level of lighting, as well as the level of pedestrian activity. Using these parameters, they developed a metric for how "walkable" an environment was. The authors found that there was a statistically significant slight positive correlation between the walkability of an environment and public transport ridership. Given that there were gaps in the data available on public record in San Francisco, the correlation might be more positive. The finding is important as it shows factors like sidewalk condition, and level of visibility will have an impact on ridership of access and egress modes that have a walking component.

Ryan and Frank's (2009) paper shows that access/egress behaviour will change based on walking conditions throughout the day. The correlation indicates that safety is an important variable in the San Francisco context. Given how much deeper the fear of crime runs in South Africa, it makes it doubly important to include safety in the choice modelling work. Also, the parameters used to describe walkability give some idea of how to describe safety to survey respondents.

Chia and Lee (2015) investigated the effect of mode captivity on acceptable walking distance for transport users in Australia. The need for this investigation came about because previous studies were often overestimating ridership. An old rule of thumb was that a bus station's catchment area was circular with a radius of 400 m (Kittleson & Associates Inc et al., 2003). This rule has come under criticism in recent years as multiple studies have shown quite significant variations from this rule in their findings. Furthermore, there is evidence to show that walking behaviour is complex and based on several variables (El-Geneidy et al., 2014). These findings further affirm that logit modelling is a significant improvement on earlier attempts at modelling mode choice due in part to its ability to handle complex multi-variable scenarios easily.

Chia and Lee (2015) refined their captivity definition to true captive and non-true captive bus riders. True captives were bus riders without a driver's licence or a private vehicle in their home. Non-true captives were people with private vehicles at home and a driver's licence, who always took the bus. Chia and Lee (2015) then tried to fit distance decay functions to the two segments.

The authors segmented the study further by splitting non-true captives into clusters based on age, income and employment status. Employment status correlated with the ability of a person to afford the expense of a private vehicle, age correlated with a person's sensitivity to walking distance, and income correlated with cost sensitivity (all different levels of non-true captivity). The authors found that overall, as a person's captivity level decreased, their sensitivity to walking time increased. That pattern suggested that in this project, there would be a high walking time sensitivity as Gautrain users had many access/egress modes available and were one of the higher-earning income groups in South Africa.

Cervero (2001) defined walkability using factors such as the ratio of sidewalk miles to street miles, the amount of mixed-use development around the transit station and the median street width of the area surrounding the rail station. As mentioned, this significantly affects a person's decision on whether to walk to the station.

Chakour and Elru (2014) measured walkability as the distance to the station and the density of the urban development surrounding the rail station. This measure also had a significant effect on whether or not a traveller walked to the rail transit station.

Van Soest, Tight and Rogers (2019) published a study that aggregated a significant portion of the literature on the distance that people were willing to walk to the transit station and the factors that influenced the distance. One of the main takeaways from their study was just how context-specific the willingness to walk to a transit station was. The authors cited studies in Australia and North America that indicated a positive relationship between the distance people were willing to walk and household income, while Asian and European studies showed the opposite. In North America, white people were ready to walk further than other ethnicities, but Europe showed the opposite trend. However, some factors always showed the same relationship. For example, the frequency of trains always had a positive relationship with the distance people were willing to walk to the station.

2.4 SURVEY DESIGN THEORY

The discussion deals with the theoretical background of the designed experiment. It is mainly based on the work of Hensher et al. (2005). It also refers to work done by Foddy (1994), Frazer and Lawley (2000), and Rose and Bliemer (2004). The following terms that are common in the field of choice experiment design are defined.

- Alternative – One possible choice in a set of finite choices.
- Attribute – A property that describes part of an alternative (for example parking cost in park and ride).
- Level – The possible values an attribute can take on in the different choice experiments. For example, if the parking fee takes on a value of R18 and R24 in a group of choices, then

it has two levels. More levels mean that the relationship between data points can be estimated better.

- Choice scenario – a hypothetical scenario where a respondent chooses between alternatives. Many choice scenarios make up a single-choice experiment.

2.4.1 Revealed preference/stated preference data

Two types of data are used in choice modelling, namely revealed preference (RP) choice data and stated preference (SP) choice data. RP data are information collected on real choices people make in everyday life. SP choice data is based on hypothetical choice scenarios presented to people. The characteristics of each alternative in a choice scenario are presented to the respondent, and the respondent's decision is recorded. Both types of data have several advantages and disadvantages.

RP data's first advantage is that it is a smaller-scale sample of a real-life market split. This trait is advantageous because the sample results could be scaled up to represent the balance in the total market. RP data are imbued with the constraints of the real-world situation. For example, taking a real-world sample of alcohol purchases at a store will result in a data sample that takes the constraints of a person's income into account. People with a "beer income" would therefore not buy champagne regularly. However, if people with a lower income were asked to make hypothetical choices, they might choose champagne more regularly than in real life, leading to a less accurate model describing alcohol choices. RP data have inherent reliability, which implies that repeated samples of the same population will result in very similar choice splits. This trait means much smaller samples can be taken in comparison to SP data.

RP does have some significant disadvantages. Firstly, it has a low potential for attribute variation. For example, if people only ever pay R20 for a particular brand of beer, the model estimated from that data will not give a reliable estimate of the market share of that beer if its price becomes R25 or R15, which makes an RP model weaker as a forecasting and predictive tool. One of the challenges of RP data is obtaining information on what alternatives were rejected by a respondent for a given choice. If a store stocks 10 different beers and the shopper only buys one type, the modeller cannot say if the shopper had been considering all the beers as options. They may have excluded six beers because their alcohol percentage was too low, or the price was too high, but there is no way to know that.

The advantages and disadvantages of SP data are almost the inverse of the advantages and disadvantages of RP data. As RP data consist of choices made in hypothetical scenarios, they may not represent market equilibrium. This problem can occur because people often behave somewhat differently in hypothetical choice scenarios when compared to their real-life behaviour. This problem means that SP data are often less reliable than RP data if the scenario design of the survey is not considered carefully. For instance, in transport, one of the main problems with SP experiments is survey fatigue. People are often asked to take such surveys at public transport stations. If respondents are extremely rushed, they will tend to fly through their choices towards the end of the survey. Their survey fatigue causes them to make irrational decisions and leads to distorted data. Another behavioural problem in collecting SP data is the mind's tendency to exaggerate its behaviour. If given a hypothetically superior transport option, the respondent could think, "I would switch to that any day", but because they are comfortable in their routine when the option does become available, they do not choose the superior option.

However, SP data can capture a much broader spectrum of behaviour due to being able to vary attributes over a broader range. A survey could therefore ask someone if they would still buy the R20 beer at R25, or how much more beer they would buy at R15. The approach allows a modeller to test the market share at every price value in that range. However, the hypothetical values still have to be within a reasonable range. For instance, if the survey presented a scenario that gave the beer a hypothetical value of R5, the modeller would get a massive market share, but it is quite unlikely that a company would charge that price as it would be selling at a loss. The other advantage of SP data sets is that it gives the analyst complete information on alternatives not chosen by the respondent. Having a complete set of alternatives and all the necessary attribute information of each alternative makes estimating a choice model much more straightforward.

2.4.2 Attribute correlation (statistical and perceptive)

Regardless of the data type, an analyst must be careful of attribute correlation when designing a choice experiment. There are two kinds of correlation that can distort the quality of choice data, namely statistical correlation and perceptive correlation.

Statistical correlation is a direct result of experimental design. If the choices in the experiment are set up randomly or without careful thought, the attributes of alternatives can be correlated over the experiment, resulting in statistically insignificant data. If in one choice scenario, for example, two attributes (A and B) for Alternative C are at their lower-level value, and Option C is not chosen, and in the next scenario, A and B are increased to their higher-value level and Option C is chosen, it becomes unclear in the data whether it was A or B which caused the change in the chosen alternative. Orthogonal experiment design is required to avoid statistical correlation (explained in 3.1.3). Orthogonal experiment design is a design where the attributes of each choice are arranged in such a way that there is no inter-attribute correlation.

The next kind of correlation that must be considered is the perceptual correlation between attributes. Even if the experiment has been designed orthogonally, care must be taken that attributes are not used that the user would perceive to be correlated even though they are presented as statistically independent. For example, in the design of a transport mode choice survey, an analyst might initially include both a categorical traffic level variable and trip time as part of vehicle-based alternatives. Every respondent will naturally perceive a high traffic level to be associated with a high trip time. If these attributes are made statistically independent in the experiment design, a user will find a situation where traffic level is low, but trip time is high. As these circumstances are not believable, they would probably make an irrational decision resulting in distorted data. Thus, in such cases, a design must only include one of the two perceptually correlated attributes.

2.4.3 Full factorial design and orthogonality

The first possible design for a choice experiment to consider is the full factorial choice set design. The full factorial design considers every possible unique combination of attributes and alternatives, given the levels as defined by an analyst. To work out the total number of choices to present in a full factorial design, Equation 2-16 is used.

$$\boldsymbol{Ch} = \boldsymbol{L}^{\boldsymbol{MA}} (\boldsymbol{Eq}\ \mathbf{3.1})$$

Equation 2-16: Number of choices required for full factorial design

$\boldsymbol{Ch}$ is the number of choices required, and $\boldsymbol{L}$ is the maximum number of levels of the attributes in the design. $\boldsymbol{M}$ refers to the number of alternatives available to the respondent. $\boldsymbol{A}$ represents the number of attributes in the design. Please note the rest of the tables in Chapter 2 are from an exercise performed by the author to demonstrate this technique. Consider a choice experiment between two E-hailing apps, Uber and Bolt. Also, assume the only attribute that differentiates the alternatives is the fee for the ride. Consider a trip where Uber charges R60 or R40 depending on the time of year/day. Bolt, on the other hand, charges R50 or R30 depending on the time of year/day. A full factorial design choice experiment would present the respondent with four choices (using Equation 2-16). The choice scenarios of the experiment are outlined in Table 2-1.

Table 2-1: Full factorial design of a choice experiment

Choice no.	Uber fare	Bolt fare
1	R60	R50
2	R40	R50
3	R60	R30
4	R40	R30

The main advantage of the full factorial design is that it is statistically orthogonal. If the experiment were not designed as a full factorial choice set, there would be a significant statistical correlation. Consider the experiment above, if Uber and Bolt fares went up and down together concurrently. If the respondent kept choosing alternate apps, it would be difficult to determine whether it was the change in the Uber fare or the change in Bolt fare that drove the change in decision-making.

Orthogonal level coding is a method of checking if an experiment has no attribute correlation. Orthogonal coding is a type of symmetrical coding that checks the correlation between attributes. If there are two levels of an attribute, the high and low levels are represented by 1 and –1 respectively. When there are three levels, the high, mid and low levels are represented by 1, 0 and –1 respectively. The orthogonal code of the experiment in Table 2-1 is shown in Table 2-2. The +1 corresponds to the higher fare, and the –1 corresponds to the lower fare.

Table 2-2: Orthogonal coding

Choice no.	Uber fare	Bolt fare
1	1	1
2	–1	1
3	1	–1
4	–1	–1
Corr	**0**	**0**

When coded orthogonally, if the sum of each column gives 0, then the experiment will not have an inter-attribute correlation. While orthogonality is valuable, the problem with the full factorial design is that the more complicated an experiment becomes, the number of choices increases exponentially. For instance, if the experiment above had one more alternative to choose from and the addition of a waiting time attribute, the experiment would require 36 choices to be made.

2.4.4 Strategies for dealing with large choice sets

Methods have been developed to deal with the excessively long choice sets produced in full factorial designs. Two are relevant to this project, namely fractional factorial design and blocking. The first strategy to deal with this problem is blocking. Blocking is dividing the survey up into smaller orthogonal sections (“blocks”). Blocking is done by adding an orthogonal blocking column to the experiment. The orthogonal blocking variable is used to split the survey up into smaller blocks. Each respondent then responds to only one block of choice sets. If enough people respond to each block, the analysis produces an orthogonal data set of the larger whole of a survey.

Consider, for example, a choice experiment that has people choosing between driving a car to Durban or taking a plane. Each mode of travel is defined in terms of travel time and cost. Both travel time and cost have two levels assigned to them. A situation like this results in a full factorial design of 16 possible choice sets. Table 2-3 shows the orthogonal level coding for this experiment as well as the addition of a blocking variable and its orthogonal coding.

The number of levels in the blocking variable determines the number of blocks, as seen above. Each respondent is assigned only to Block 1 or Block 2. All the blocking variable values of 1 are assigned to one block, all the values of –1 to the other block. Now each respondent only has to respond to 8 choices instead of 16 choices. However, two respondents are now needed to get one complete survey response.

This technique can be scaled up to more extensive experiments to allow for more complicated designs while reducing the chance for survey fatigue to occur (see Section 2.4.1). Tables 2-4 and 2-5 show Block 1 and Block 2 of Table 2-3, and these smaller blocks are also orthogonal on their own.

Table 2-3: Orthogonal coding of a 16-choice set experiment with a blocking variable with its orthogonal coding

	Car		Plane		Block	
Choice	TT	Cost	TT	Cost	Block number	Orthogonal code
1	–1	–1	–1	–1	1	1
2	–1	1	1	1	2	–1
3	1	–1	1	1	1	1
4	1	1	–1	1	2	–1
5	1	1	1	–1	2	–1
6	–1	–1	1	1	1	1
7	–1	1	–1	1	2	–1
8	–1	1	1	–1	2	–1
9	1	–1	–1	1	1	1
10	1	–1	1	–1	1	1
11	1	1	–1	–1	2	–1
12	1	–1	–1	–1	2	–1
13	–1	1	–1	–1	1	1
14	–1	–1	1	–1	1	1
15	–1	–1	–1	1	2	–1
16	1	1	1	1	1	1
sum	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>		<u>0</u>

Table 2-4: Block one of the SP experiment shown in Table 2-3

	Car		Plane		Block	
Choice	TT	Cost	TT	Cost	Num	Orthogonal code
1	–1	–1	–1	–1	1	1
3	1	–1	1	1	1	1
6	–1	–1	1	1	1	1
9	1	–1	–1	1	1	1
10	1	–1	1	–1	1	1
13	–1	1	–1	–1	1	1
14	–1	–1	1	–1	1	1
16	1	1	1	1	1	1
Sum	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>		

Table 2-5: Block 2 of the SP experiment shown in Table 2-3

	Car		Plane		Block	
Choice	TT	Cost	TT	Cost	num	orthogonal code
2	–1	1	1	1	2	–1
4	1	1	–1	1	2	–1
5	1	1	1	–1	2	–1
7	–1	1	–1	1	2	–1
8	–1	1	1	–1	2	–1
11	1	1	–1	–1	2	–1
12	1	–1	–1	–1	2	–1
15	–1	–1	–1	1	2	–1
Sum	0	0	0	0		

The next strategy to discuss is the fractional factorial design of choice experiments. The fractional factorial design revolves around reducing the size of choice sets by removing choice sets in such a way that the orthogonality of the design is maintained.

The grounds for the removal of choice sets come from some assumptions made about the interaction effects of explanatory variables.

Interaction effects are the effects that two or more explanatory variables have on a response variable when they work together. Consider a response variable, *V*, that has two explanatory variables,X_1 and X_2. A model that estimates interaction effects is depicted in Equation 2-17.

$$\boldsymbol{V = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2}$$

Equation 2-17: Known utility equation with interaction effects

$\boldsymbol{\beta_n}$ represents the regression coefficient of the response variable, $\boldsymbol{\beta_0}$ represents the regression constant of the response variable and $\boldsymbol{X_n}$ represents the explanatory variables. Consider a situation where $\boldsymbol{V}$ represents the volatility of a chemical solution. $\boldsymbol{X_1}$ refers to the amount of nitric acid in solution. $\boldsymbol{X_2}$ refers to the amount of glycerine in solution. Individually, both compounds are inert, but they explode when combined. In this model, then, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ on their own would not contribute any volatility at all. However, the interaction between $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ would be hugely significant. Interaction effects do not have anything to do with correlation. Correlation means that variables increase and decrease concurrently. If nitric acid and glycerine were correlated, then when nitric acid in the solution increased or decreased, the amount of glycerine would increase or decrease in sync with the acid. Therefore, correlation does not show the effect that nitric acid and glycerine have on volatility when they working together. The interaction effect does.

A full factorial design of a choice experiment allows for accurate estimation of all interaction effects in a set of explanatory variables. One of the strategies employed to reduce choice sets to a manageable size is to assume that these interaction effects are not statistically significant. Choice scenarios are removed selectively from the experiment in such a way that they do not compromise the orthogonality of the experiment. For instance, a fractional factorial choice set (with its orthogonal coding) from the example of Table 2-6 is shown below. However, the reduction of choice sets means that interaction effects cannot be estimated.

A fractional factorial design, like blocking, reduces the set of choices one respondent has to make. Fractional factorial design's main advantage compared to blocking is that fewer respondents are needed. If there is a three-block survey, three people are needed for one complete survey response. Conversely, if the fractional factorial design is employed, one person is needed for one complete survey. In that case, however, such a fractional factorial design makes it impossible to estimate the effect of most interaction effects in the model. The missing choice sets lose the information that allows for the estimation of interaction effects.

Table 2-6: A fractional factorial design of the experiment shown in Table 2.4.3

	Car		Plane	
Choice	**TT**	**Cost**	**TT**	**Cost**
1	–1	1	1	1
2	1	–1	1	1
3	1	1	–1	1
4	1	1	1	–1
5	1	–1	–1	–1
6	–1	1	–1	–1
7	–1	–1	1	–1
8	–1	–1	–1	1
sum	**0**	**0**	**0**	**0**

3 SURVEY DESIGN

Chapter 3 outlines the design of the survey based on the theory in Section 2.4. It begins by discussing the results of a focus group and the reward scheme decided upon for the survey. Then the technical design of the survey is explained, describing how attributes and levels were chosen. In conclusion, the chapter provides a breakdown of the demographic results of the survey and compares the trader sample to the GMA segmentation report. The full questionnaire is provided in Appendix A.

3.1 FOCUS GROUP AND REWARD SCHEME

Section 3.1 outlines the results of the Gautrain focus group carried out by Nielsen (2019) marketing research and the incentive offered to attract the survey sample.

3.1.1 Nielsen focus group sample and results

At the request of the GMA, part of this study involved engaging a focus group of 10 Gautrain users from Pretoria. The objective of the focus group was to obtain an anecdotal sense of Gautrain users' access and egress perceptions of the system to see if the survey design was in line with perceptions on the ground.

Nielsen, a globally respected market research firm, was asked to recruit a 50/50 gender split of Gautrain users. They had to be between 25 and 45 years old. Also, they had to travel on the Gautrain frequently, at least twice a week between 06:00 and 09:00.

The access mode behaviour of the recruited group was somewhat anomalous as none of the recruits used their private car to reach the Gautrain. Their perceptions of the Gautrain were mixed. While they found the train system fast and efficient, they found the system as a whole pricey and stressful, given the inefficiency of the access/egress modes.

The responses to questions about the park and ride facilities were likewise mixed. On the one hand, some respondents thought the car parks looked like a safe option, but others thought they did not look secure. One respondent thought that car parking cost R100 per day for Gautrain commuters (it costs R20 for commuters; non-commuters pay R100). This finding raised the issue of false perceptions of the system that might hamper the accuracy of the results. If a survey participant is faced with a park and ride choice that has a parking fare of R30, they would probably choose it, whereas in real life they might not pick the park and ride because of their misconception.

The perceptions of the drop-off/pickup mode were mostly positive. It was perceived to be faster than the Gautrain bus and very cost-effective. The main concern was that people could not always secure a lift.

There respondents' perceptions of the Gautrain bus service were somewhat negative. While it was seen as convenient when the service ran smoothly, it was perceived that it rarely ran smoothly. Focus group members said they often could not find space on the bus, the bus stops were in unsafe areas, bus drivers did not keep to the schedule, were often rude, and the buses often got stuck in traffic. These perceptions are difficult to quantify in an SP survey; even so, they will still influence the unknown utility component of the Gautrain bus component.

When the survey respondents were pressed on what they did when their primary mode was unavailable, they mostly said that e-hailing was a backup. They preferred e-hailing to a normal taxi because of the perceived convenience of e-hailing, and they would only consider using the Gautrain bus system. No other public transport system seemed good enough for them.

3.1.2 Reward scheme

To get a big enough sample size for the main survey, an incentive was needed to generate interest in the survey. The simplest solution to the problem was to give respondents vouchers on Takealot (South Africa's equivalent to Amazon). People were offered a R50 coupon on their next purchase at Takealot. They had the option to provide an email address for delivery of the coupon code. Given the budget of the survey, 400 vouchers were available. To generate interest in the survey quickly, it was decided that the first 300 respondents would get a voucher, then the next 120 vouchers would be given to every fifth person that completed the survey. To avoid people repeating the survey for extra cash, they were asked to provide a valid email address and Gautrain Card number. These were checked for duplicates.

There was a risk that the survey reward would introduce bias, especially with regards to the value of the voucher. The lower the value of the voucher, the less likely it was that the survey would draw high-income respondents. Conversely, however, increasing the value of the voucher would reduce the risk of bias toward lower-income users. However, if fewer vouchers were offered it would reduce the overall sample size.

3.2 TECHNICAL DESIGN

Section 3.2 describes the technical design of the survey, outlining the design problems and how they were solved. The choice of instrument is also described along with the orthogonal coding of all the experiments.

3.2.1 Design problems and assumptions

The main goal of the survey was to estimate a model or models that would describe how people chose to access the Gautrain system and then egress from it to their destination.

The first question to address was whether people's choice of access and egress mode should be regarded as only one decision (choosing their modes together after choosing Gautrain) or whether they chose each mode independently. This led to a number of other questions. What modes should be considered in the experiment? Also, would changes in access and egress mode circumstances force people off the Gautrain? If so, how could it be dealt with realistically? Including the main mode choice in the questionnaire was plausible but getting real answers would be difficult. The survey was due to be marketed on GMA’s social media platforms. It would most likely only attract people who already used the Gautrain as a primary mode. Even if wider marketing of the survey were possible, there would still be challenges. Only a certain portion of the driving population of Gauteng could use the Gautrain. If the survey were marketed to a broader audience, it would probably include people for whom using the Gautrain was not a feasible option, resulting in a high percentage of non-trader respondents. It would therefore be unclear if the non-traders in question simply could not use the Gautrain or did not want to.

For the design of the questionnaire, it was assumed to treat the access and egress trips as separate decisions. There are three reasons for this assumption. The first is that from an intuitive point of view, it makes sense that these choices are independent. Consider a passenger who drives his car to the Gautrain station as his access mode of choice. That choice of access mode has no bearing on whether he will walk to his destination or catch a Gautrain bus. The dominant factor there is probably the distance to his destination. The access mode of the car therefore has little to no bearing on his decision of egress mode. There is a possibility that car drivers might have increased safety concerns (as they tend to be middle to upper class) and are less likely to walk to their destination. However, this type of effect is assumed not to be significant enough to consider the choice of access and egress mode as one combined choice.

Secondly, most discrete choice modelling exercises look at either first-mile or last-mile mode choices in isolation. The two trips are not considered a combined decision. This tendency is in every study referenced in Section 2.2, with one exception. Cervero (2001) developed binomial logit models on whether a person chose to walk for their first-mile trip to a station or their last-mile trip to their destination. The first-/last-mile choices were considered separately. It is also worth noting that the primary mode was not considered in that study either.

The last reason for this assumption was a practical one. If access and egress mode choice is one decision, then the choice experiments needed for that become massive. Consider a user who chooses between private car/e-hailing on the access side and walking/bus on the egress side. Also, assume all modes have two descriptive attributes each, and that they each have two descriptive levels. When considered as separate choices, the access and egress experiments will be 16 choices for each experiment, with a total of 32 responses required per person. In an experiment where they are considered as one choice, instead of two alternatives for each experiment, there are now four alternatives for one experiment. This one experiment now has 256 choices required per respondent. The size of the survey is too large to be practical even with fractional factorial design and blocking.

Then the survey needed a list of modes to consider. The modes were obtained from a 2017 Gautrain market segmentation report commissioned by the GMA and conducted by Plus 94 research. In their sample, 95% of the trips to and from the station were gained by considering E-hailing, private car, drop-off/pickup, taxi, walking and the Gautrain bus service. The last 5% consisted of people using other public transport like bus rapid transit (BRT) systems. These were considered a preliminary set of modes. The data from the report is shown in Figure 3-1.

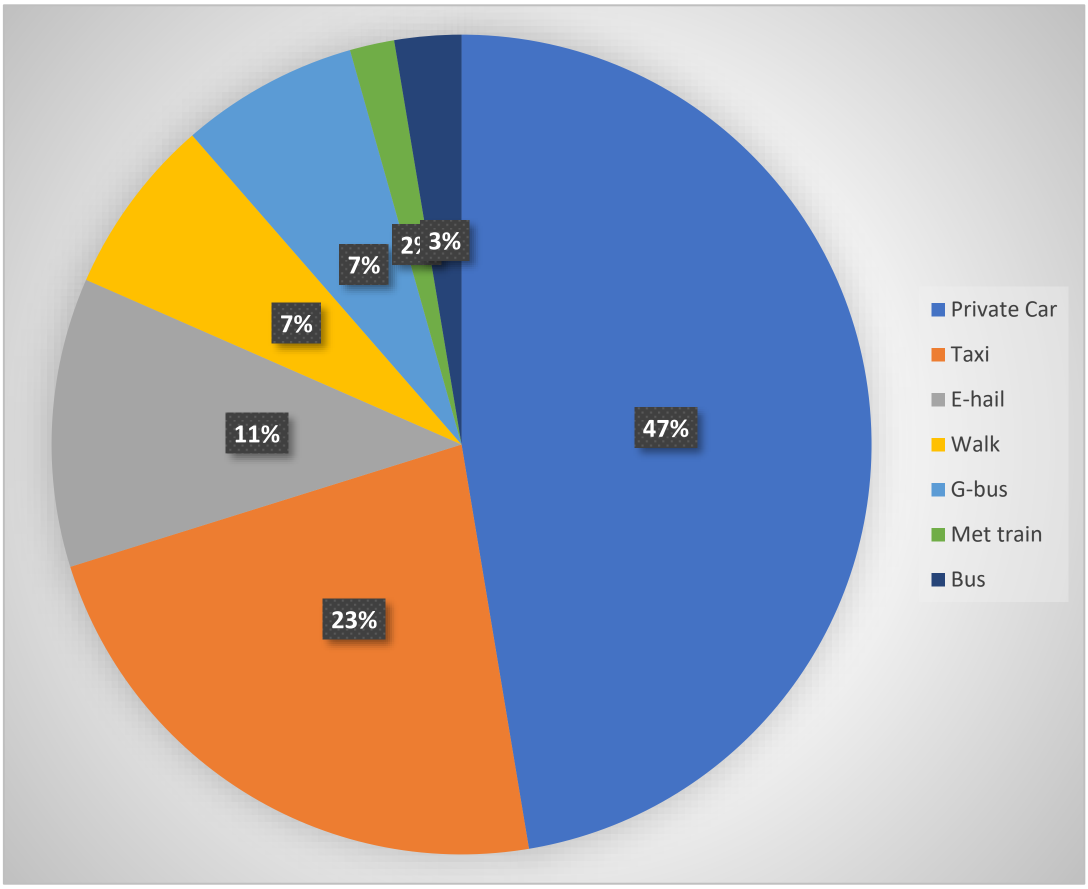

Figure 3-1: Current access modes from Gautrain market segmentation report (Plus 94 research, 2017)

The next question to consider was the question of mode captivity. How should the survey handle users who were captive to a mode on either the first- or last-mile trip? *Captive* refers to a type of user who if their current access/egress mode became infeasible, would stop using the Gautrain. For example, consider a student who walks to the station in the morning on their first-mile trip. Then the student takes the bus in the evening on their last-mile trip. If their bus stop is moved, causing a longer walk in the evening, the student might stop using the Gautrain out of concern for their safety and start carpooling to campus.

Captive users on the access side and the egress side had to be filtered out of the data. It was also important to have a solution to deal with people captive on either the access or egress side but free to choose on the other side. A way was needed to capture their choice data on the journey where they had a choice.

Captivity was dealt with as follows. First, it was assumed that if a person were captive to a first-or last-mile mode that they currently used, they would appear in the data as non-traders, only choosing their current mode for every scenario they were given. Based on this assumption, separating the first-mile or last-mile SP experiment allowed the identification of captives to one side of the journey. For example, a man who is only captive to his first-mile mode will be a non-trader in the first-mile SP experiment. However, he will then appear as a trader in the last-mile SP experiment. The key assumption in dealing with mode captivity was to assume that all captives show up as non-trader respondents.

It was decided to focus the survey on the morning peak period, which is the period in the day when the highest strain on first-/last-mile services (like parking or bus capacity) is experienced. If the first- or last-mile services are optimised for the morning peak, the rail system should manage other periods in the day. To capture the morning peak, the respondent was asked at the beginning of the survey to consider their last Gautrain trip, starting from their home. Asking that question removes the risk of getting data on the afternoon peak as well. Afternoon peak data would be quite different. For example, there would be a large share of people taking their private car for their last-mile journey in the afternoon peak period. Avoiding the afternoon peak narrows the scope of modes available to users on both sides of the trip. Based on boarding and alighting data from 2014, OR Tambo airport commuters were also excluded (by asking the purpose of their trip). A typical example of a morning peak period trip matrix is shown in Table 3-1 (07:00 – 08:00).

Table 3-1: Morning peak origin–destination matrix for the 07:00–08:00 peak (GMA, 2014)

		DESTINATION											
		Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield	ORTIA	Total	Ex Ortia
ORIGIN	Hatfield	0	7	102	94	19	411	171	259	10	23	1 096	**1 073**
	Pretoria	27	0	48	87	20	233	110	264	7	5	801	**796**
	Centurion	154	47	0	110	33	577	252	369	12	7	1 561	**1 554**
	Midrand	130	87	76	0	10	328	132	295	17	9	1 084	**1 075**
	Marlboro	43	49	47	28	0	80	23	44	17	18	349	**331**
	Sandton	78	59	58	23	4	0	41	106	44	117	530	**413**
	Rosebank	78	62	47	33	5	58	0	32	9	31	355	**324**
	Park	119	85	115	94	7	468	69	0	21	9	987	**978**
	Rhodesfield	34	17	34	51	28	650	168	73	0	0	1 055	**1 055**
	ORTIA	5	6	17	5	11	67	7	2	0	0	120	
	Total	668	419	544	525	137	2 872	973	1 444	137	219	7 938	
	Ex ORTIA	**663**	**413**	**527**	**520**	**126**	**2 805**	**966**	**1 442**	**137**			**7 599**

As can be seen from Table 3-1, the ORTIA (OR Tambo International Airport) trips make up a small percentage of the morning peak period. Also, OR Tambo commuters are outliers in terms of access preference to the station. Their sensitivities are likely to be different to those of the average commuter. They carry baggage, are more sensitive to punctuality concerns out of fear of missing their flight. Capturing their sensitivities would bias the estimated coefficients away from the commuters. Because of this, airport travellers were excluded from the survey.

The other issue that came up was what modes should be assumed to be available to each user to choose from. Every user of the Gautrain has a different set of feasible access/egress modes. For example, people in safe neighbourhoods, close enough to a Gautrain station, can walk to the station. However, other people are too far away to reach their closest station on foot, so a walking trip cannot be assumed to be a feasible travel mode for everyone. Furthermore, mode availability is tied to the experiment size. If a person can feasibly choose between a private car/Gautrain bus/E-hail/drop-off/walking, and even if each mode is described in terms of only two attributes and two levels for each, the experiment would be 256 choices per respondent. Also, the time the trip is made changes the available modes significantly. For example, in the morning peak period, the private car is a significant access mode, while in the evening peak period, it becomes a significant egress mode.

The solution to the question of mode availability was to use respondent data to narrow the scope of available modes for each user. The survey asked respondents what their current access/egress mode was, providing a list of all the alternative modes considered in this study for access/egress journeys and asking them to select which other modes were available to them. To reduce the scope of the survey, respondents were asked to select the mode they were most likely to switch to, given a choice, taking them to an SP choice experiment between their current mode and their most likely alternative. All other modes were then assumed to be unavailable, or not considered by the respondent. For example, if a respondent currently used a private car to reach the station and then said they would most likely switch to the Gautrain bus service, they were taken to a stated choice experiment between those two modes.

This process was repeated for the last-mile journey. Respondents could choose from a similar set of modes, but the private car was removed because it was a factor in the morning peak period and it was unlikely that someone would have a private car at their egress station. Each given access mode had a set of limited feasible alternatives with SP experiments. Table 3.2 shows each mode: the one a respondent could say they currently used and the set of alternatives they could choose from in the SP experiment.

Table 3-2: Possible current access modes with a list of feasible alternatives

Current access mode	Possible alternative access modes	Explanation and reasons
Walking	- Gautrain bus - Drop-off - E-hail	A passenger currently walking is unlikely to use their car (if available) and incur the cost of the park and ride facility at a Gautrain station (currently R20 per day). A Gautrain bus is an alternative, depending on the walking distance, or a drop-off. E-hailing is an option, especially in an "emergency", e.g. poor weather.
Car (access side)	- Gautrain bus - Drop-off - E-hailing	A respondent currently using a car to park and ride is unlikely to walk to the station. Other motorised modes are the most likely alternatives.
Gautrain bus	- Car (only on access side) - E-hailing - Drop-off	A passenger using a Gautrain bus is unlikely to walk. If a car is available, they may consider this option. E-hailing is an option, especially because it is becoming cheaper. If available, a shared lift to a station is likely.
Drop-off	- Car (only on access side) - Gautrain bus - E-hailing	Being dropped off or sharing a car to the station is likely to be replaced by the car option, or a Gautrain bus if one is available. E-hailing is an option, especially if lift-reliant passengers are stranded due to their lift being unavailable, i.e. an "emergency" option. A person who is currently dropped off will probably not walk.
E-hail	- Car - Gautrain bus - Drop-off	A person who E-hails may have a car to use, or someone could become available to give them a lift. They may be persuaded to switch to a bus, but if they are E-hailing, they are unlikely to walk.

The approach outlined in this section was regarded as an acceptable middle ground between the level of detail of the survey and how practically workable the survey was.

3.2.2 Choice of attributes and levels

This section details the attributes that were chosen for each model. It discusses the level assignment (showing the orthogonal coding) and deals with the values assigned to each attribute's level.

The safety level of the sidewalk was an unusual variable included in the study, although it does occur in the literature. Tilahun and Li (2015) developed a binomial logit model for whether a person would walk to their transit station. In their SP survey, they defined the level of crime on a scale from 1 to 5 with 1 being the safest and 5 representing the least safe level. Including it in the study made sense given the context; Chicago ranks in the top 20 most dangerous cities in America according to CBS News (CBS, 2019).

Initially, it was hoped that the present survey could capture the person's actual attribute value (e.g. trip time) and define the upper and lower levels in terms of a percentage increase and decrease of the real value. However, the platform used did not allow for this functionality (see 3.2.3). It was decided to look at average attribute values and give upper and lower limits in a range that seemed plausible. Each mode's attributes are detailed below.

Private car

- Parking fare – Currently, a Gautrain passenger pays R22 a day to park their car. For the survey, the parking fare was set at a lower-level value of R18 and an upper-level value of R30. This was because it is unlikely the GMA will want to lower parking cost (as it will incentivise more private car travel) but would explore the effect of increasing parking cost. Thus it is more important to explore what happens at higher levels than lower levels of parking cost.
- IVT (In-vehicle Travel Time) – The average speed for a private car was assumed to be about 40 km/h, which is consistent with congested traffic conditions during the morning peak. To get a reasonable upper and lower level for IVT, the assumption was made that the average Gautrain user would make a private car journey no shorter than 10 km and no longer than 20 km. This is consistent with the observation that the majority of car access users travel between 10 and 20km to a station, according to Gautrain's station surveys. Gautrain's market segmentation report found that 95% of respondents said their total journey (first mile + train trip + last mile) was less than 30km in total. This assumption resulted in a lower-level IVT value of 15 minutes and a higher-level value of 25 minutes, with a third mid-level of 20 minutes in the experiment using the Gautrain bus as an alternative mode.
- Trip cost – Average fuel consumption was assumed to be 8 litres per 100 km. This was based on a 2018 finding by the global fuel economy initiative that the South African passenger car average fuel economy was 6.3 litres per 100 km with a small increase in consumption as a precaution. At the 2019 fuel price of R15/litre, the experiment was given a lower-level value of R6 and a higher-level value of R8 given the assumed trip distances. Although the difference of R2 seems insignificant, it was important to keep these choices realistic.

Drop-off/Pickup

- IVT – The IVT values were calculated from the same assumptions made in the private car mode, except that drop-off had two levels in every experiment. The upper level was 25 minutes and the lower 15 minutes, the same upper and lower levels as the IVT of the private car. This was assumed because the trip characteristics of a private car trip and a drop-off/pickup trip are similar (traffic congestion levels, distances travelled and speeds). Hence the IVT values should not be changed in any significant manner.

- Cost – The upper level for drop-off was set at R30, and the lower level at R15. These values are higher than the trip cost for the private car mode as it needed to account for the time of the person driving the car, although not the full time-cost as the driver is typically a colleague or family member and not a professional driver. At these levels, the value of time of the driver is taken to be approximately R60/hr, which is reasonable. Early pilot testing indicated that under-valuing the trip cost of the drop-off/pickup mode resulted in its becoming a dominant mode (i.e. chosen every time), which is not realistic.

E-hail

- IVT – The upper level was made 25 minutes, and the lower level 15 minutes. This is again because the average IVT of an e-hail trip is probably quite similar to a drop-off/pickup trip or private car trip due to similar trip characteristics (see the section on drop-off IVT).
- Fare – The comparative prices for Uber and Bolt (the two biggest E-hailing apps in South Africa) appeared to range between R30 and R50 depending on the time of day, trip length etc. These figures were obtained by requesting rides of between 10 to 20 km to various Gautrain stations. To obtain a wider range of measurable behaviour, the lower level was decreased to R20 and the upper increased to R60.

Walking

- Walking time – The average walking speed was taken to be 4 km/h. This was from a study that indicated the average human walking speed ranges from about 4-6 km/h depending on the environment (Cronkelton, 2019). Gautrain users were assumed to take walking trips of no longer than 1,5 km to reach the station. Cervero’s (2001) work indicated Californians would walk no more than 1 km to a station. People in Africa tend to be willing to walk longer distances (Venter, 2020). A walking trip of 1,5 km delivered the upper-level walk time of 22 minutes. The lower level was set to a 1 km trip and worked out to be 15 minutes.
- Safety level of the walk – This variable describes the safety of the walking environment. The level of safety was described to the respondents in terms used in the study by Ryan and Frank (2009). (see Section 2.3). “Very secure” was described as an environment with many people around, bright lighting and a visible security presence. “Somewhat secure” was described as an environment with fewer people about, dim lighting and no visible security presence.

Gautrain bus

- Walking time – 4 km/hr was again used as the walking speed. However, it was assumed people would walk no more than 1 km to get to a bus stop, so the upper level was set to 10 minutes and the lower to 5 minutes. This was based on existing literature on the first-/last-mile walking behaviour of Gautrain users (Venter, 2020).
- Waiting time – By analysing the bus schedules available on the Gautrain website, a good estimate of the average headway in the morning peak could be determined. During the morning peak, the average headway over the peak period is about 7 minutes. Given this average, the maximum level of waiting time was set to 10 minutes, to take into account lower frequencies and irregular arrivals. The minimum was set to five minutes.
- Bus fare – Gautrain offers different kinds of bus services: it was found that the lowest fare was R6 and the highest fare was R15, depending on the time of day and size of the vehicle. These were used as the lowest and highest value respectively.
- IVT – Given the high traffic volumes of the morning peak period, buses will travel at the same speed as cars. Therefore, the upper level was kept at 25 minutes and the lower level at 15 minutes.
- Safety level of the walking – see explanation in the discussion of the walking mode.

After the selection of attributes and levels, each choice combination was coded orthogonally into the relevant SP experiment.

Table 3-3: Level coding for car vs drop-off choice experiment

Design	Car			Drop-off	
Choice scenario	IVT	Petrol cost	Parking fee	IVT	Carpool cost
1	1	1	–1	–1	–1
2	–1	–1	–1	–1	1
3	1	–1	–1	1	–1
4	–1	1	–1	1	1
5	–1	1	1	–1	–1
6	1	–1	1	–1	1
7	–1	–1	1	1	–1
8	1	1	1	1	1

Table 3-4: Level coding for walking vs drop-off/pickup choice experiment

Design	Walking		Drop-off	
Choice scenario	Walking time	Safety of walk	IVT	Carpool cost
1	–1	–1	–1	–1
2	1	–1	–1	1
3	1	–1	1	–1
4	–1	1	–1	–1
5	–1	–1	1	1
6	1	1	–1	1
7	1	1	1	–1
8	–1	1	1	1

Table 3-5: Level coding for the Car V E-hail choice experiment

Design		Car		E-hail	
Choice scenario	IVT	Petrol cost	Parking fee	E-hail fare	IVT
1	1	1	–1	–1	–1
2	–1	–1	–1	–1	1
3	1	–1	–1	1	–1
4	–1	1	–1	1	1
5	–1	1	1	–1	–1
6	1	–1	1	–1	1
7	–1	–1	1	1	–1
8	1	1	1	1	1

Table 3-6: Level coding for Gautrain bus vs pickup/drop-off choice experiment

Design	Drop-off/pickup		Gautrain bus				
Choice scenario	IVT	Carpool cost	Walking time	Waiting time	IVT	Fare	Safety of walk
1	–1	1	1	1	–1	–1	–1
2	1	1	–1	–1	–1	–1	1
3	1	–1	1	–1	–1	1	–1
4	–1	–1	–1	1	–1	1	1
5	1	–1	–1	1	1	–1	–1
6	–1	–1	1	–1	1	–1	1
7	–1	1	–1	–1	1	1	–1
8	1	1	1	1	1	1	1

Table 3-7: Level coding for drop-off/pickup vs E-hail choice experiment

Design	Drop-off/pickup		E-hail	
Choice scenario	Car pool cost	IVT	E-hail fare	IVT
1	–1	–1	–1	–1
2	1	–1	–1	1
3	1	–1	1	–1
4	–1	1	–1	–1
5	–1	–1	1	1
6	1	1	–1	1
7	1	1	1	–1
8	–1	1	1	1

Table 3-8: Level coding for Gautrain bus vs E-hail choice experiment

Design	Gautrain bus					E-hail	
Choice scenario	Walking time	Waiting time	IVT	Fare	Safety of walk	E-hail fare	IVT
1	–1	1	1	1	–1	–1	–1
2	1	1	–1	–1	–1	–1	1
3	1	–1	1	–1	–1	1	–1
4	–1	–1	–1	1	–1	1	1
5	1	–1	–1	1	1	–1	–1
6	–1	–1	1	–1	1	–1	1
7	–1	1	–1	–1	1	1	–1
8	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Table 3-9: Level coding for walking vs E-hail choice experiment

Design	Walking		E-hail	
Choice scenario	Walking time	Safety of walk	IVT	Fare
1	–1	–1	–1	–1
2	1	–1	–1	1
3	1	–1	1	–1
4	–1	1	–1	–1
5	–1	–1	1	1
6	1	1	–1	1
7	1	1	1	–1
8	–1	1	1	1

Table 3-10 Level coding for car vs Gautrain bus blocked experiment design

Design	Car			Bus					
Choice scenario	IVT	Travel cost	Parking fee	Walking time	Walking safety	IVT	Waiting time	Bus fare	Block
0	0	–1	–1	–1	–1	0	–1	–1	3
1	1	–1	–1	0	–1	1	0	0	1
3	1	0	0	0	–1	1	–1	0	1
4	0	0	–1	0	–1	0	–1	–1	3
5	–1	–1	–1	0	0	0	0	0	1
6	0	0	–1	–1	0	–1	–1	0	0
7	0	–1	0	0	0	–1	–1	–1	0
8	–1	–1	0	–1	0	0	–1	0	1
9	1	0	0	–1	–1	–1	0	0	3
10	–1	–1	0	–1	–1	1	0	–1	0
11	–1	0	–1	–1	0	1	0	–1	0
12	1	0	0	0	0	–1	0	–1	3
13	1	–1	–1	–1	–1	1	–1	–1	0
14	–1	–1	–1	0	–1	–1	0	0	3
15	–1	0	0	0	–1	–1	–1	0	3
16	1	0	–1	0	–1	1	–1	–1	0
17	0	–1	–1	0	0	1	0	0	3
18	1	0	–1	–1	0	0	–1	0	1
19	1	–1	0	0	0	0	–1	–1	1
20	0	–1	0	–1	0	1	–1	0	3
21	–1	0	0	–1	–1	0	0	0	0
22	0	–1	0	–1	–1	–1	0	–1	1
23	0	0	–1	–1	0	–1	0	–1	1
24	–1	0	0	0	0	0	0	–1	0
25	–1	–1	–1	–1	–1	–1	–1	–1	1
26	0	–1	–1	0	–1	0	0	0	0
27	0	0	0	0	–1	0	–1	0	0
28	–1	0	–1	0	–1	–1	–1	–1	1
29	1	–1	–1	0	0	–1	0	0	0
30	–1	0	–1	–1	0	1	–1	0	3
31	–1	–1	0	0	0	1	–1	–1	3
32	1	–1	0	–1	0	–1	–1	0	0
33	0	0	0	–1	–1	1	0	0	1
34	1	–1	0	–1	–1	0	0	–1	3
35	1	0	–1	–1	0	0	0	–1	3
36	0	0	0	0	0	1	0	–1	1

Given the work shown in tables 3-3 to 3-10, each scenario was prepared as shown in Figure 3-2.

	Car	Gautrain Bus
Bus walk time to stop (min)	-	5
Safety level on walk to bus stop	-	Very Secure
Bus wait time at stop (min)	-	10
Travel time car & bus (min)	25	20
Bus fare (Rand)	-	6
Car petrol cost to station (Rand)	10	-
Car parking fee (Rand)	18	-

Figure 3-2: Example of a choice scenario that a respondent could encounter in the 2019 passenger survey

3.2.3 Choice of survey instrument and the layout of the survey

When the design of the questionnaire was completed, the next step was to determine how to disseminate it to the public. The segmentation report yielded information on the media consumption habits of the different segments of the Gautrain market. Of the 661 people surveyed, 94% of respondents listed the internet as one of their modes of media consumption. It was the most frequently selected mode of available options. Conventionally, SP surveys in South Africa have been done by approaching people at bus stations, petrol stations and other commuter stop-off points. A common problem is survey fatigue (see 2.4.1). By putting the survey online, people would be able to take the survey at their own pace, so the risk of survey fatigue was greatly diminished.

It was therefore decided to launch the survey on the web, coordinating with the GMA to use their social media presence for advertising the survey.

Because no experienced web designer was available to put the website together from scratch, an online survey hosting tool, Survey Monkey, was used. The tool proved to be significantly challenging as it had never hosted a stated choice survey before. It was apparent that the technical team did not understand how the survey had to be structured to factor in a percentage variation on the respondent's actual value. The survey design was eventually completed after much experimentation and exploring the application of the Survey Monkey toolkit.

The layout of the survey (how it takes a respondent through the questionnaire) is shown in Figure 3-3.

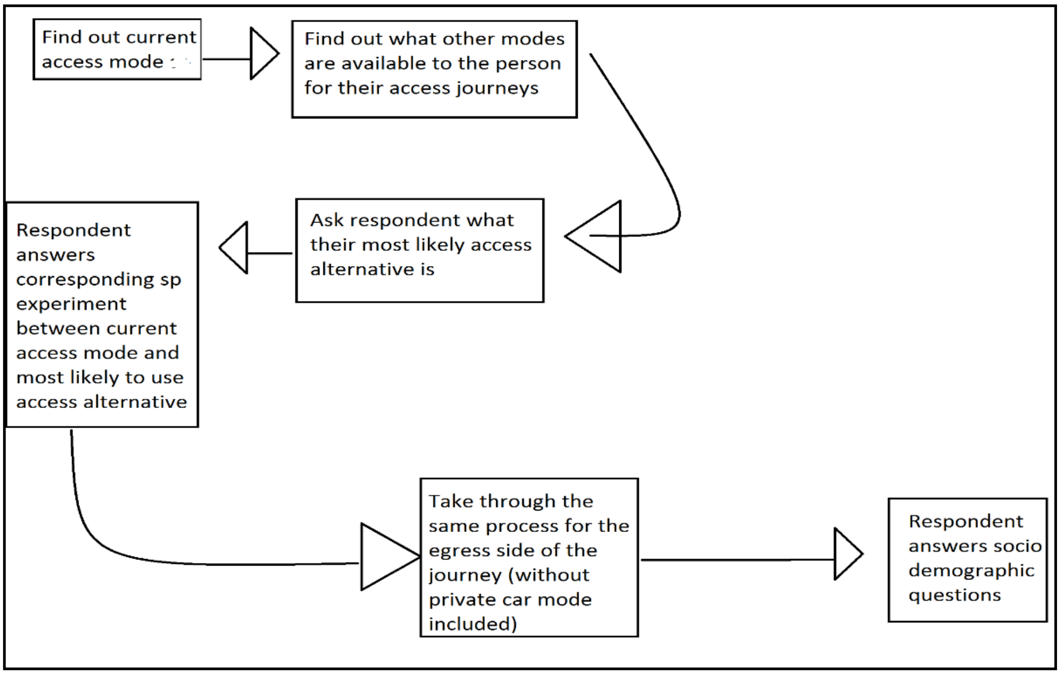

Figure 3-3 Layout of survey on Survey Monkey platform

3.3 SURVEY SAMPLE RESULTS AND BREAKDOWN

Section 3.3 outlines the results of the online survey. It discusses the total sample of respondents, how the sample was cleaned for analysis and how the non-traders in the sample were handled. It then breaks down the demographics of the sample and compares it to the demographics of the 2017 Gautrain market segmentation report.

3.3.1 Total sample responses, cleaning data and handling non-traders

In all, there were 375 responses during the month and a half that the survey was left online. About 20% of the data (78 responses) were incomplete due to poor internet connectivity. Another 15% of respondents were travellers on the OR Tambo line. These responses were removed from the sample. Of the remaining approximately 240 responses, 25% were complete non-traders. That is to say, of the two modes they had to choose from, they consistently chose the same option on the access or egress side of the survey and showed no willingness to trade off the attributes given in the experiment. Figures 3-4 and 3-5 show the mode share of the modes chosen by the non-traders.

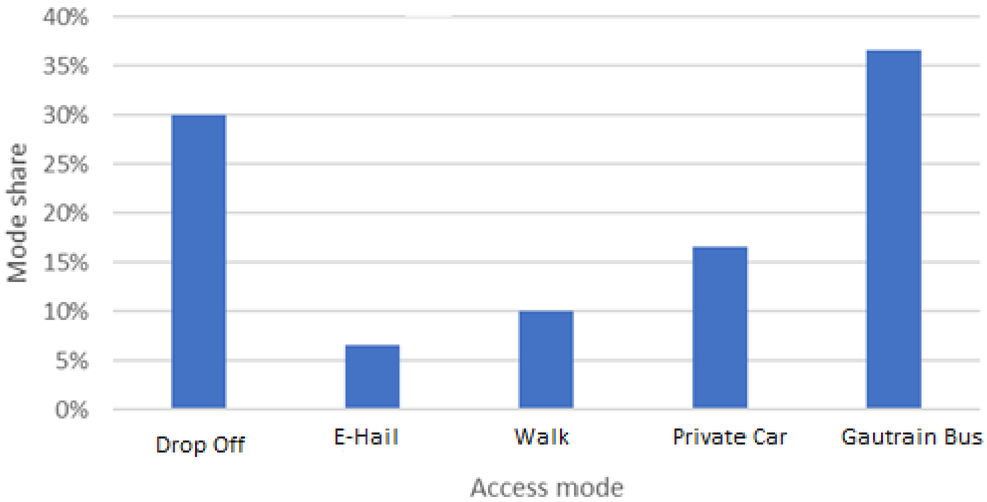

Figure 3-4: Mode split of access journey: non-traders

Dealing with non-traders in SP data is a contentious issue. Hess et al. (2010) outline the pitfalls that a high rate of non-traders presents to modellers. There are two possible explanations for non-trader behaviour. The first explanation is that the non-traders are acting rationally, seeking to maximise their utility. However, they value one alternative significantly higher than other available alternatives, to the extent that the value of the levels chosen in the survey does not prompt the respondent to change their mode of transport. The second explanation for this behaviour is that people are acting irrationally, hence not maximising utility, for example from the force of habit.

In the first case, the data where people are choosing rationally should be included in the model in principle. However, the data in cases where people choose irrationally should be discarded because they do not conform to the maximisation of utility principle. The main problem is that there is no way to distinguish between these two types of non-trader data.

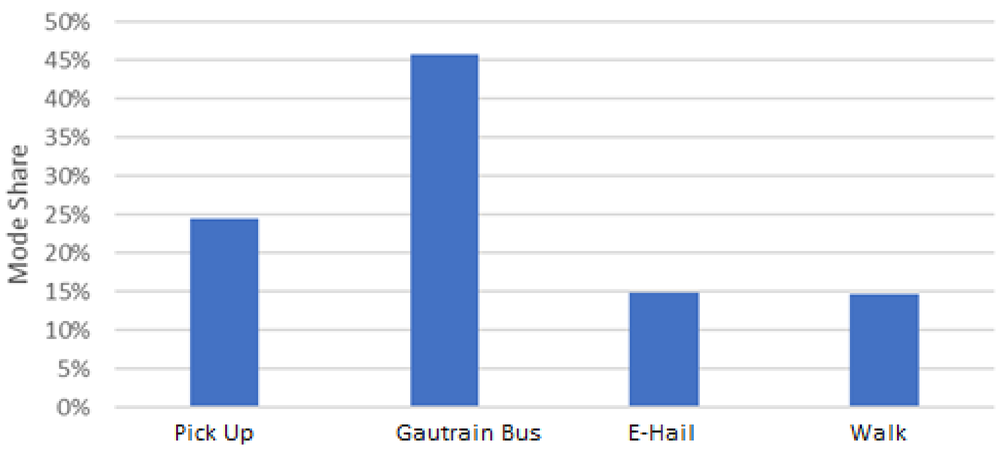

Figure 3-5: Mode split of egress journey: non-traders

Hess et al. (2010) further probed the dilemma by examining four sets of choice data from different parts of the world. They compared the effects that removing non-traders and traders had on models. There were cases where non-trading comprised a small percentage of the data (approximately 1% of respondents). Here, no significant effects on the model were recorded when removing non-traders from the sample.

In another study surveyed by Hess et al. (2010) there was a higher percentage of non-traders in south Yorkshire, removing non-traders had a substantial effect on the data. Removing car non-traders significantly decreased the value of time savings of car users. Including the non-traders significantly increased the value of the ASC in the model.

Hess et al. (2010) therefore proposed three approaches to dealing with non-trader data. The first approach was to leave the data in the sample. The significant risk of this approach is that it may include data where people are not following the utility maximisation assumption, which leads to distortion of the coefficients and the ASCs of the model. The second approach is to remove the non-trader data, which reduces the chance of inflation of the ASC and VOT figures. However, this risked making the sample non-representative of the population or removing data that is actually rational but shows non-trader behaviour. The third approach is to develop a method to distinguish between the two types of non-trader data. Currently, no such method exists.

Given the circumstances and the significant percentage of non-traders in this dissertation's sample, it was decided to estimate models with the non-trader data both included and excluded. Doing this allowed for determining how much distortion the non-traders had on the data set.

3.3.2 Socio-demographic breakdown and comparison to previous sample of Gautrain users

Section 3.3.2 presents the socio-demographic breakdown of the respondents of this survey. To determine how the survey sample compared to the average Gautrain population, each socio-demographic split was compared with the corresponding split from the Gautrain market segmentation report (Plus 94 research, 2017). Every socio-demographic question provided a "Prefer not to say" option, allowing respondents to opt out of the question. In every category, about 15% of the 182 traders selected the "prefer not to say" option. As these answers comprised such a small portion of the responses, they were simply ignored in the sample breakdown. Only the demographics of the traders were given and because they were the data source for the final models, the non-traders and incomplete responses were irrelevant.

Gender

There was a minor gender reversal in this survey's sample compared to the Gautrain market segmentation report, the sample being 55% female in comparison with the Gautrain market segmentation report's sample of 58%. This is shown below in Figure 3-6

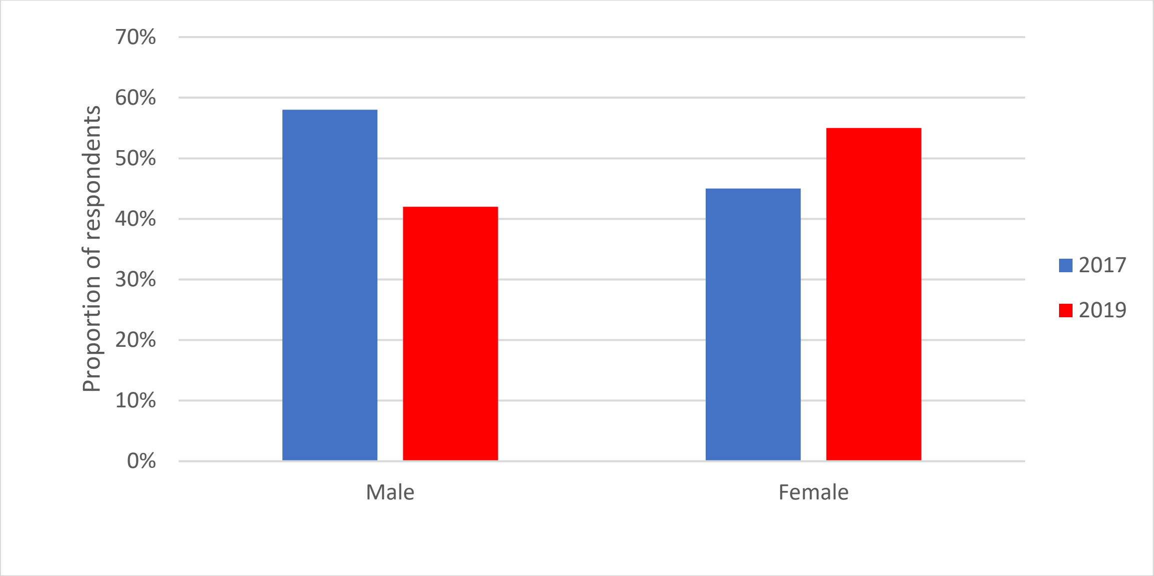

Figure 3-6 Gender of respondents from 2017 Gautrain market report compared to gender of respondents from 2019 SP survey

The reversal of the gender majority might best be explained by how the sample was collected. The GMA conducted face-to-face interviews throughout their stations. In the present study, the survey was advertised online through the Facebook and Twitter pages of the GMA. Men and women might have different social media habits which would lead to the reversal seen in this sample. However, the imbalance is still not significant enough to render the sample biased.

Age

The age breakdown of the two samples is quite similar. However, the Gautrain market segmentation report had a larger proportion of respondents in the 18–24 age range. The 2019 sample had a somewhat larger proportion in the 35–44 age range. However, both samples had a mean and median age in the mid-twenties. The results are shown in Figure 3.7.

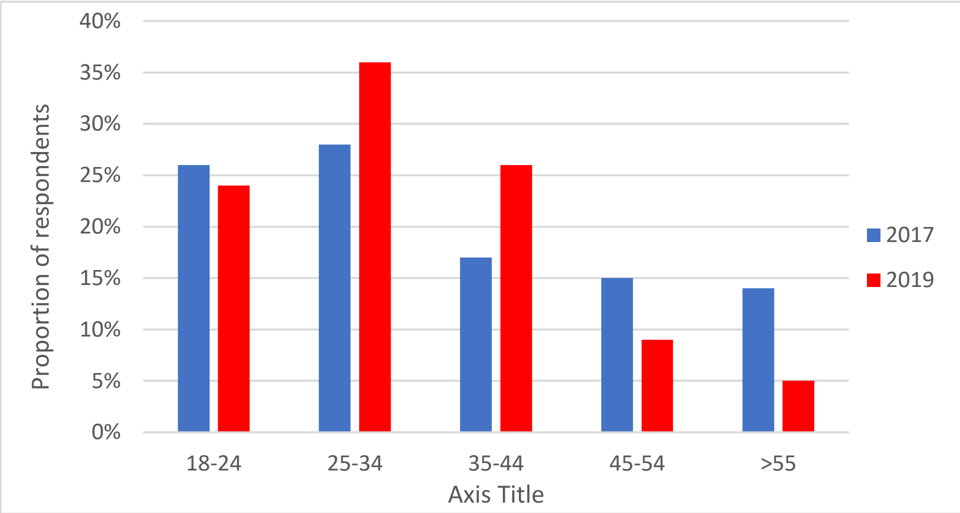

Figure 3-7 Age of respondents from 2017 Gautrain market segmentation report compared to age of respondents from 2019 SP survey

The significant difference in the 2019 survey seems to be that the sample was skewed toward younger audiences. There was a significant drop in the number of respondents of 45–54 years old and older than 55 years. This could be for two reasons, the first being that older people used social media less and were less likely to see the advertisement of the survey. The second issue could be that older people tended to be wealthier and that a voucher of R50 might have been too low an incentive for them to take the survey.

Employment status

As Figure 3-8 shows, the employment status of the 2019 sample was similar to that of 2017. The main difference of the 2019 sample was that it had a higher percentage of part-time employed persons and scholars; however, full-time employment still dominated the sample demographics.

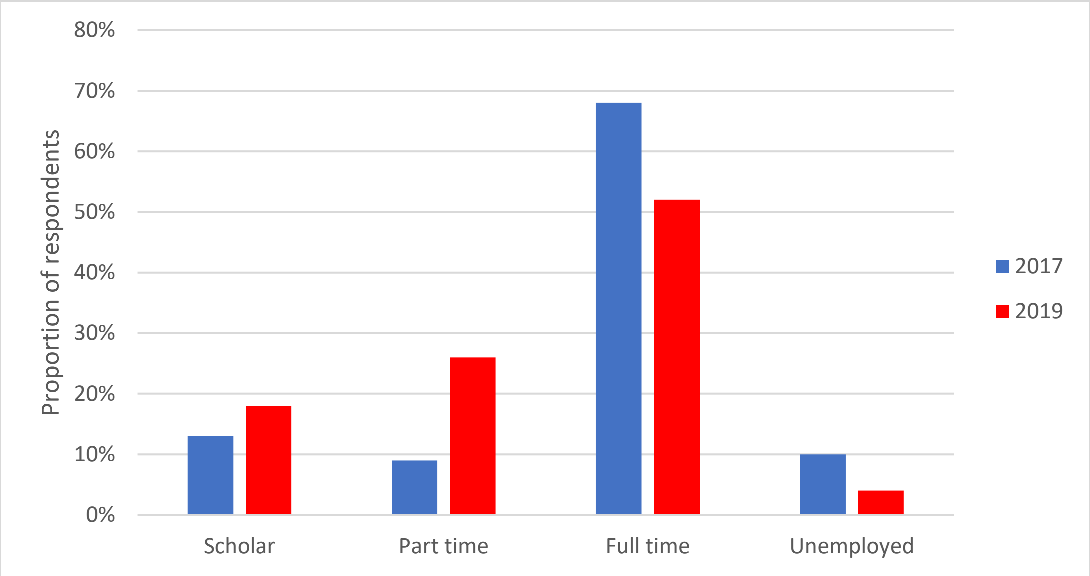

Figure 3-8 Employment status of respondents from the 2017 Gautrain market segmentation report compared to the employment status of respondents from the 2019 SP survey.

The sample from 2019 has some differences to the 2017 sample but the differences are not significant. If the sample were representative, the increase in part time employment and decrease in full time employment might have been due to the economic downturn in South Africa over the last two years. Moreover, the increase in *scholar* as an occupation might be due to the bias toward younger age ranges discussed earlier.

Income of respondents

Both the 2019 and 2017 surveys similarly asked for personal income, and the 2019 sample indicates that there has been a decrease in the number of people with a personal income of less than R10 000 per month. The income patterns were otherwise very similar, as shown in Figure 3-9. However, the 2017 report sample showed that household income was significantly higher than personal income. This goes some way to explaining some of the findings in the 2019 study's model (discussed in Chapter 4). Figure 3-9 shows the household income breakdown of the 2017 report.

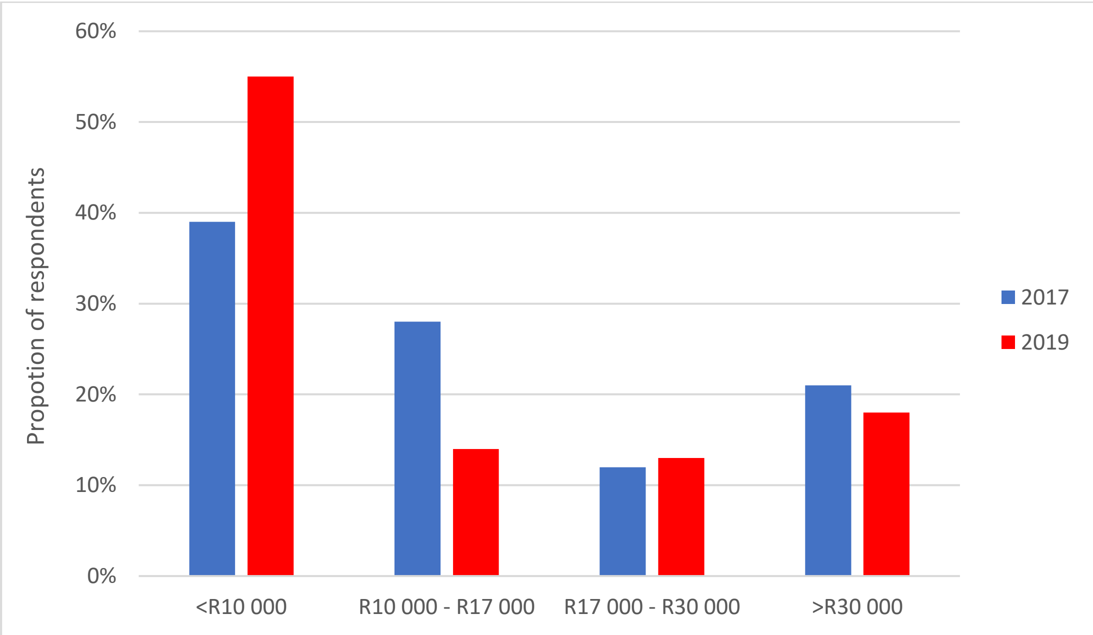

Figure 3-9 Income of respondents from the 2017 Gautrain market segmentation report compared to the income of respondents from the 2019 SP survey

The 2019 survey shows a sharp decrease of respondents in the < R10 000 income range, from 55% to 39%, and an increase of respondents in the R10 000 to R17 000 range, from 15% to 28%. Otherwise, the latter categories of income have not changed much between the two reports.

Current access mode

There have been some significant changes in access behaviour since the 2017 report. One of the major 2017 modes was in relation to taxis. It was thought that E-hail had largely supplanted this in the last two years. To test this assumption, the survey asked the question "How often do you use Metered Taxis?" The breakdown of the responses is shown in Figure 3-10.

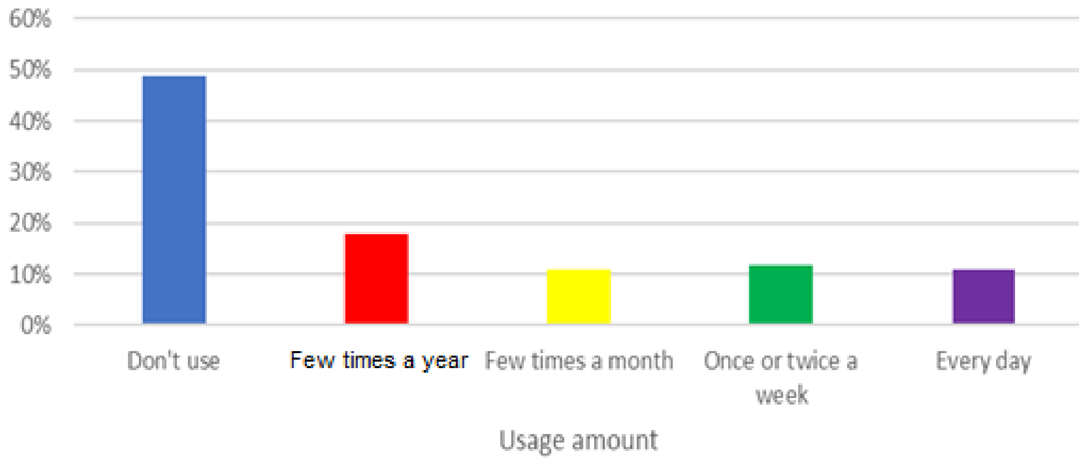

Figure 3-10 Metered taxi use in 2019 among Gautrain users

As shown, about 90% of respondents do not use a conventional metered taxi as a regular mode. This could best be explained by the rise of E-hailing in South Africa over the last two years. It should be mentioned that the Gautrain market segmentation report does not distinguish between the private car and drop-off/pickup modes. They had been lumped together in the data. In 2017, respondents were asked for their most feasible access mode; in the current study, respondents were asked for their most recent access mode. The access modes are displayed in Figure 3-11 and Figure 3-12.

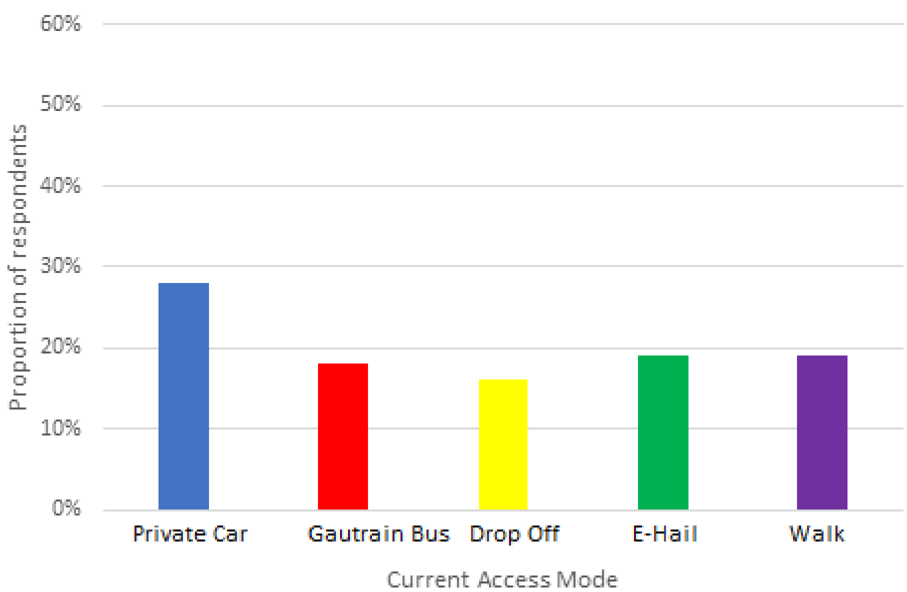

Figure 3-11 Access mode of trader respondents in 2019 survey

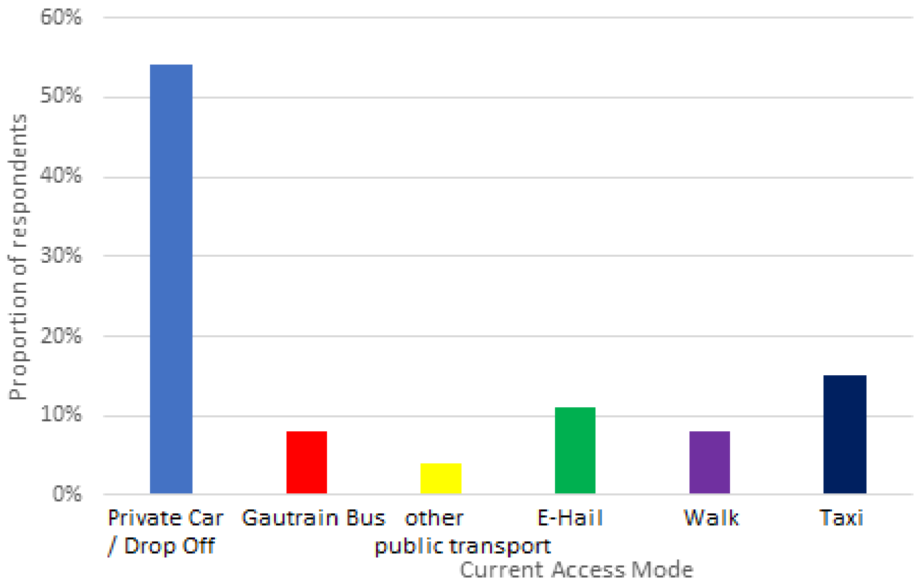

Figure 3-12 Access modes: Gautrain market segmentation report sample (2017)

The graphs show that commuters have migrated from taxis and other private transport to E-hail, Gautrain bus, private car/drop-off, and walking. The increase in walkers might be because of the bias toward younger Gautrain riders in the 2019 sample. Older people may be less likely to walk to the station than younger people.

Current egress mode

Figures 3-13 and 3-14 show the currently used egress mode share of the 2019 survey and then the 2017 report.

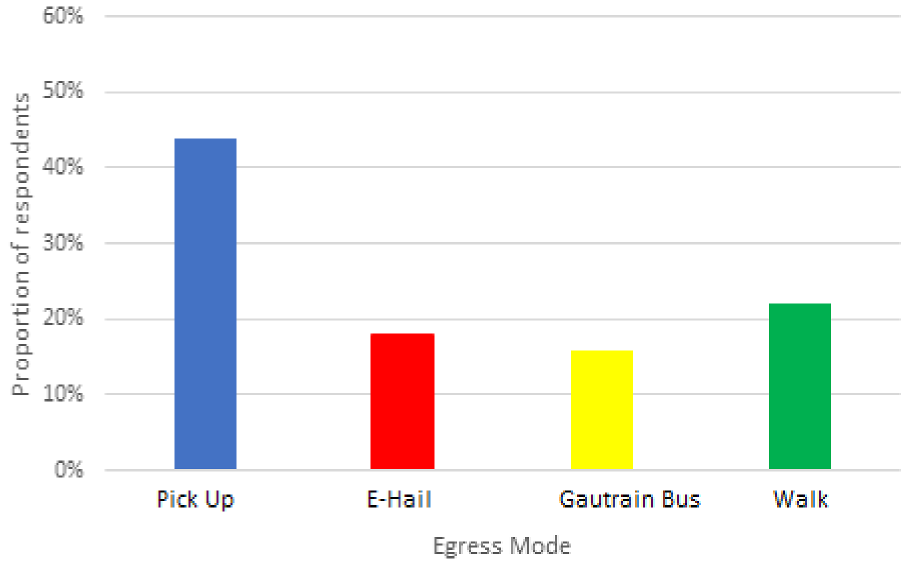

Figure 3-13 Egress mode of trader respondents in 2019 survey

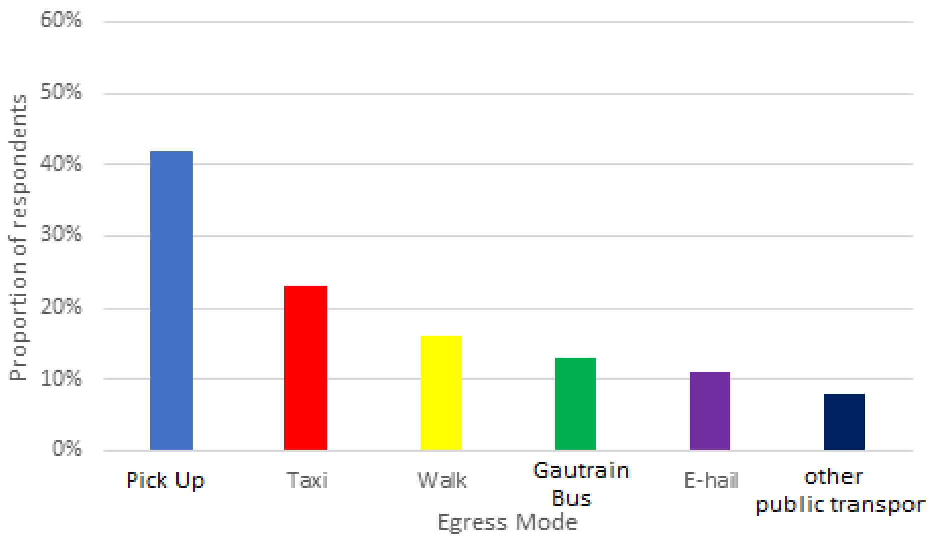

Figure 3-14 Egress modes of Gautrain market segmentation report (2017)

The same increase in a walking sample can be seen. E-hail also increases in mode share. Some modes like being picked up remain relatively unchanged.

Conclusions

In conclusion, the sample was biased in some categories but not others. The gender change is significant but not extreme enough to sharply bias the results. The 2019 sample is biased towards younger riders in the age category. It is also biased toward the lower middle-income range (reducing the low-income user share in the sample). The current access/egress modes of the sample remain relatively unbiased. There is a small increase in walking mode share, probably due to the age bias. Given the information on metered taxi use, and assuming the sample is representative, the increase in modes like E-hail probably reflects real changes in transport preferences since the GMA survey was carried out.

In general, the level of change between the two surveys suggests that the 2019 sample is representative within a reasonable margin of error of the 2017 survey sample. The most significant bias is to reduce the percentage of older people. As mentioned previously, this has had some effect on the other categories in the sample.

4 MODELLING RESULTS

This section describes the models developed with the data and their results. This chapter is divided into three parts. The first section presents the models estimated on the access SP data alone. The second part outlines models estimated on the egress SP data. The last part presents models estimated on the combined access and egress data. The abbreviations used to describe the modelling results and utility equations throughout this chapter are defined in Table 4-1.

Table 4-1: Abbreviations used in modelling work.

Abbreviation	Meaning
Pc	Private Car
DO/PU	Drop-off/Pickup
Gb	Gautrain bus
EH	E-hail
Wk	Walking
IVT	In-vehicle time
Wkt	Walking time
Wtt	Waiting time
Pk	Parking cost
f	Trip cost
Gen	Gender
Sf	Safety level

It should be noted that no interaction effects were estimated for any of the models because the SP experiments used the fractional factorial design method and blocking in the case of the private car and Gautrain bus experiment.

4.1 ACCESS MODELS

This section starts by showing the MNL models estimated for the access data. It then presents two different nesting structures estimated on the data.

4.1.1 Access MNL

The first model run was the access SP data as a standard MNL. The utility equations were specified below. Then the non-traders from the access SP experiment are included in the data, and the model is re-estimated to understand the effect of non-traders on the model.

Model without non-traders

U(do)= b11*gen + b5*ivt + b6*f

Equation 4-1 – Utility of DO alternative

U(pc)= b0 + b5*ivt + b6*f + b7*pk

Equation 4-2 – Utility of PC alternative

U(gb)= b1 + b3*wkt + b4*wtt + b5*ivt + b6*f + b8*sf

Equation 4-3 Utility of GB alternative

U(eh)= b9 + b5*ivt + b6*f

Equation 4-4 Utility of EH alternative

U(wk)= b2 + b11*gen + b3*wkt + b8*sf

Equation 4-5 Utility of Wk alternative

The results of this model run are shown on the next page in Table 4-2. The significance of the attributes as a group is shown by where they sit on a chi-square distribution. This probability is given in the column “probability that the model as a whole is not significant”.

Table 4-2 Best-fitting MNL for Access only data without non-traders

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
Gen (B11)	-0.334	0.160	-2.09	0.036
IVT (B5)	-0.021	0.012	-1.77	0.076
F (B6)	-0.032	0.006	-5.64	0.000
PC ASC (B0)	0.284	0.402	0.71	0.480
Pk (B7)	-0.038	0.015	-2.46	0.014
Gb ASC (B1)	-0.171	0.309	-0.55	0.580
Wkt (B3)	-0.046	0.021	-2.18	0.029
Wtt (B4)	-0.007	0.031	-0.23	0.816
Sf (B8)	0.571	0.136	4.20	0.000
EH ASC (B9)	-0.084	0.140	-0.60	0.546
Wk ASC (B2)	-0.021	0.491	-0.04	0.965
Final LL function		-840.82		
AIC (Akaike information criterion)		1704		
Chi-Square Test result		2206.30		
P Score (chi distribution)		0.00		

The model yielded good results. As this MNL served as the baseline for other models, the LL and AIC are simply baselines with which to measure the quality of other models. However, the chi-square results show clearly that the coefficients as a group are significant (see chi-square test result and P Score). The ASCs are all very insignificant in their z-scores (the equivalent of a t-test). This result shows that most of the significant utility attributes were captured for each access mode. However, this result is mainly due to the removal of the non-trader data from the sample.

All the coefficient signs are rational. Safety is positive since it is equal to one. All the categories of time have negative coefficients, as do all the financial variables, which makes sense as they are all disutilities. Also, their z values (Wald value, equivalent to a t-test) are all significant within 1% or 5% or 10% intervals.

An interesting finding is a very high value of IVT savings in the South African context (R40/hr). Other forms of public transport have values of about R15/hr or less (Hayes & Venter, 2017). This difference is likely due to the very high income of Gautrain users relative to the rest of the population. Waiting time is entirely insignificant in the model, whereas walking time and IVT are highly significant; this will be discussed later. The negative gender coefficient does make sense, and it shows women are more inclined towards private modes of transport, as they might feel less comfortable travelling in groups of people. In addition, that effect may have been captured in the safety variable.

Table 4-3: Access MNL variance/covariance matrix

	B11	B5	B6	B0	B7	B1	B3	B4	B8	B9	B2
B11	1.00	0.00	0.01	0.19	0.00	0.24	0.03	0.00	-0.01	0.51	0.00
B5	0.00	1.00	0.02	-0.05	0.06	0.02	0.01	-0.03	-0.01	0.00	0.16
B6	0.01	0.02	1.00	0.18	0.02	0.19	0.02	0.01	-0.03	-0.32	0.24
B0	0.19	-0.05	0.18	1.00	-0.92	0.12	0.03	-0.01	0.00	0.13	0.03
B7	0.00	0.06	0.02	-0.92	1.00	0.00	0.00	0.01	-0.01	0.00	0.02
B1	0.24	0.02	0.19	0.12	0.00	1.00	-0.45	-0.75	-0.15	0.15	0.52
B3	0.03	0.01	0.02	0.03	0.00	-0.45	1.00	0.01	-0.11	0.05	-0.89
B4	0.00	-0.03	0.01	-0.01	0.01	-0.75	0.01	1.00	-0.02	0.00	-0.01
B8	-0.01	-0.01	-0.03	0.00	-0.01	-0.15	-0.11	-0.02	1.00	0.01	-0.03
B9	0.51	0.00	-0.32	0.13	0.00	0.15	0.05	0.00	0.01	1.00	-0.07
B2	0.00	0.16	0.24	0.03	0.02	0.52	-0.89	-0.01	-0.03	-0.07	1.00

The variance/covariance matrix (Table 4-3) reasonably fits the IIA assumption. It does have high correlations between some of the coefficients. However, these correlations involve the insignificant ASCs, so it is not a severe detriment to the model.

Model with non-traders

To estimate the effect of the non-traders on the sample data, the same model was estimated on the sample. The only difference was the inclusion of non-trader choice data. The results of the model are shown below in Table 4-4.

Table 4-4: Access SP data MNL with non-traders included

Parameter	Estimated coefficient	Standard error	T-Test result	Wald value (probability parameter is not significant)
IVT (B5)	-0.016	0.011	-1.46	0.145
F (B6)	-0.029	0.005	-5.61	0.000
Gen (B11)	-0.376	0.147	-2.56	0.011
PC ASC (B0)	0.238	0.359	0.66	0.507
Pk (B7)	-0.037	0.014	-2.68	0.007
GB ASC (B1)	-0.164	0.281	-0.58	0.560
Wkt (B3)	-0.027	0.019	-1.44	0.151
Wtt (B4)	-0.002	0.028	-0.09	0.930
Sf (B8)	0.533	0.123	4.33	0.000
EH ASC (B9)	-0.248	0.134	-1.87	0.062
WK ASC (B2)	-0.256	0.440	-0.58	0.560
Final LL function		-1018.03		
AIC (Akaike information criterion)		2058.10		
Chi-square Test result		2676.77		
P Score (chi distribution)		0.000		

The results of the non-trader model are a useful demonstration of the problem with non-trader data. The effects on the model are mostly negative but there are individual improvements. Overall, the LL decreased by about 200 points. This indicates the model will make inaccurate predictions more often. The model was fed additional choice data and became less able to make an accurate prediction. There are two possible explanations for the drop in predictive power. Firstly, it could be that the choice behaviour that underlies the data does not conform to the utility-maximising assumption that these models are based on. Alternatively, there may be attributes that are considered when making this choice that have not been added into the SP experiments, for example seat availability on a bus. However, given that the attributes common to these kinds of studies have all been included, the first explanation is more plausible.

In discussing the non-trader behaviour in the current study, it is reasonable to assume that a decrease in the magnitude of the ASC of a particular mode indicates that its non-trader data is still consistent with the utility maximising assumption. The reason for this is if a particular attribute drives non-trader behaviour, then the attribute causing the non-trader behaviour will become more significant in the model. For example, consider a choice between two modes where one mode is always more expensive than the other, but the time savings of the expensive mode vary. If a respondent always choses the cheaper mode in that scenario, then the model would interpret that behaviour as a very high sensitivity to trip cost and increase the significance, and magnitude of the trip cost sensitivity coefficient. This kind of result would increase the proportion of known utility relative to unknown utility of a particular mode. This would have the effect of decreasing the magnitude of the ASC of the mode. Conversely, the non-trader data of a mode that makes the ASC magnitude increase suggests the data is inconsistent with utility maximising behaviour.

The significance of E-hail and walking ASCs have both increased; E-hail even comes close to 1.96 on the t-test score. The magnitude and significance of both coefficients have increased and by extension, the amount of unquantifiable utility has increased. Given these effects, it seems plausible that the E-hail and walking non-trader data were irrational (non-utility maximising).

However, one sees the opposite effect in the ASCs for private car and Gautrain bus respectively. Both coefficients become smaller and less significant. This shows there is less unquantifiable or irrational behaviour in the alternatives, which implies that non-trader data for these modes may well be rational (utility-maximising). However, the overall effect of the non-traders in the model is to degrade it. This is because the amount by which car and bus improve is significantly smaller than the amount by which E-hail and walking decrease.

The non-trader effect on the magnitude of attribute coefficients does not reveal any noticeable pattern and most of the changes in magnitude are quite small. However, except for parking cost and safety, all the attributes had their t-test score decrease significantly.

4.1.2 Nested access models

Section 4.1.2 covers three areas. First, it presents the different nesting structures attempted. Each structure's results are summarised along with the reason they were rejected in favour of the best-fitting structure. Then the section discusses the best-fitting nested logit model on the access SP data. Finally, a version of the best-fitting nested model structure with non-traders in the sample is presented and discussed.

The best-fitting nesting structure is defined as the one which shows the most significant reduction in LL and has the most significant nesting parameters and the lowest magnitude of nesting parameters. The utility equations used in this section are the same ones used in 4.1.1.

Summary of rejected nesting structures

The alternative structures rejected in favour of the best-fitting model are shown in figures 4.1–4.3. At the first level of each figure is the reason why the modes in the nest were grouped. It was assumed that modes with overlapping characteristics would correlate with the unknown utility part of the model.

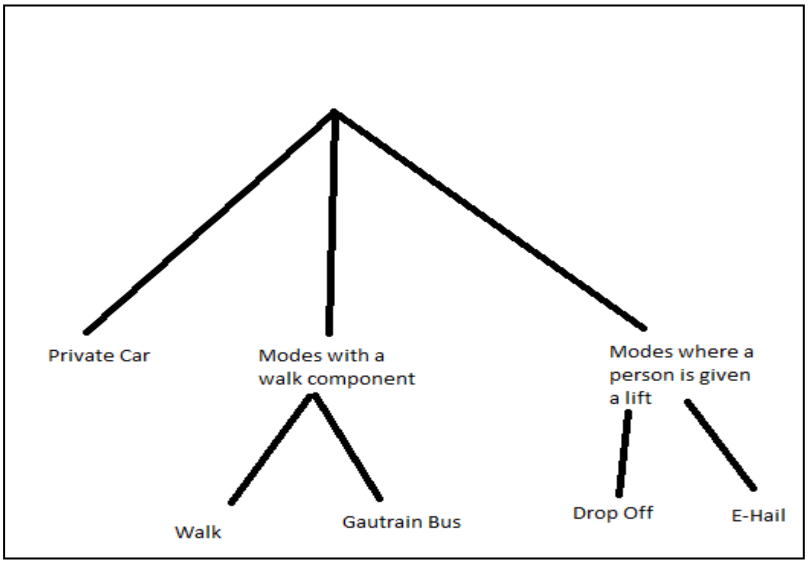

Figure 4-1 Alternative nesting structure 1

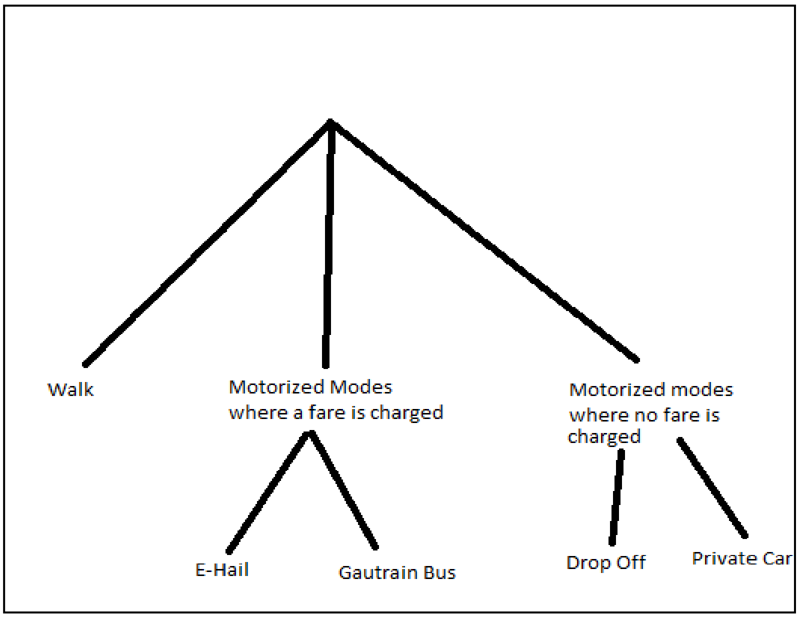

Figure 4-2 Alternative nesting structure 2

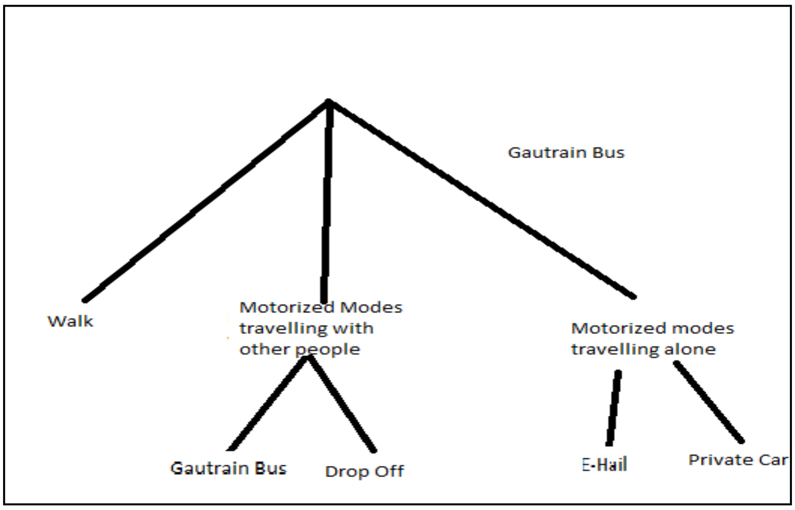

Figure 4-3 Alternative nesting structure 3

It should be noted in Figure 4-4 most people take an E-hail alone in the morning commute. It was supposed that both private car and E-hail involve travelling in an isolated space, on your own, away from any number of people. This comfort of being away from a crowd might have some correlation in the unknown utility. Table 4-5 summarises the results of each alternative structure and the best-fit model.

Table 4-5: Summary of nesting structure results for access SP data

Model structure	Log likelihood	Nesting parameters	Reason for rejection
Best-fit model (Figure 4.4)	-838.21	(PC, EH)-0.87 (GB, WK)-0.67 (DO)-0.40	n/a
Alternate Structure 1 (Figure 4.1)	-838.47	(WK,GB)-1.16 (DO, EH)- 1.54 (PC)-3.28	Lower LL than best-fit model and nesting parameters >1
Alternate Structure 2 (Figure 4.2)	-840.50	(WK)-0.71 (GB, EH)- 0.77 (PC, DO)-0.88	Lower LL than best-fit model
Alternate Structure 3 (Figure 4.3)	-840.20	(WK)-0.96 (GB, DO)- 1.14 (PC, EH)-1.52	Lower LL than best-fit model and nesting parameters >1

Best-fitting nested model structure

The nesting structure for this model is shown below in Figure 4-4. The best model nested private car and E-hail together. This correlation could be because both modes have an element of privacy to them (morning E-hails would tend to be solo trips). E-hail does not involve trying to find a lift from a friend or colleague, and it can be organised quickly and conveniently. That privacy element might allow for some correlation between the two alternatives. Gautrain bus and walking were correlated because the walking element overlapped in the modes (along with the safety element). Drop-off was left in its own nest.

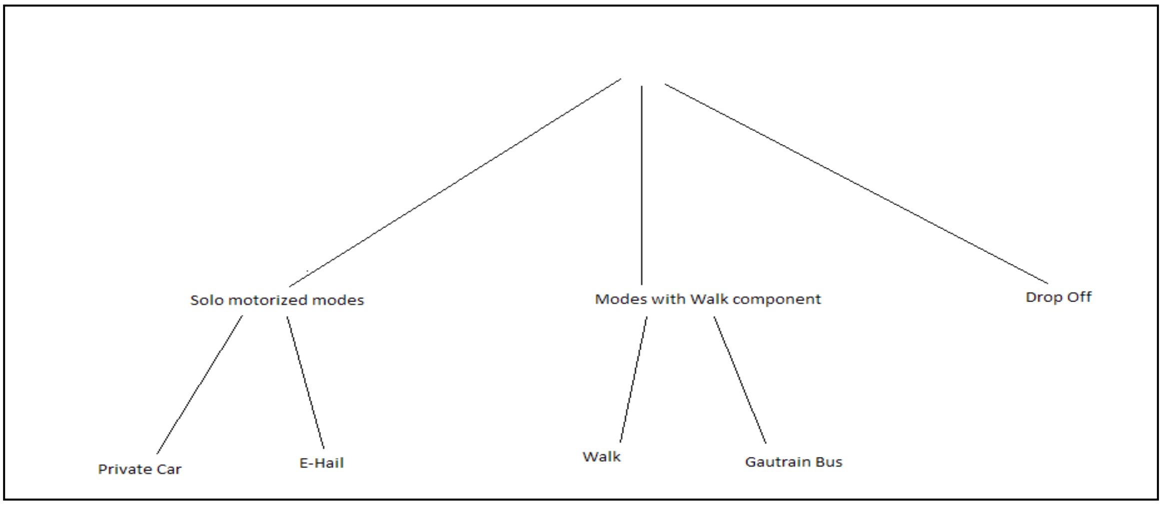

Figure 4-4 The nesting structure of the best-fitting nested model on the access SP data.

The utility equations are unchanged. The results are in Table 4-6.

Table 4-6: Best-fitting nested logit model results for access only data

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Probability that coefficient is not significant
Gen (B11)	-0.530	0.289	-1.83	0.069
IVT (B5)	-0.031	0.017	-1.79	0.074
F (B6)	-0.045	0.012	-3.89	0.000
PC ASC (B0)	1.081	0.622	1.74	0.082
Pk (B7)	-0.0420	0.020	-2.12	0.034
Gb ASC (B1)	0.425	0.659	0.65	0.519
Wkt (B3)	-0.058	0.028	-1.99	0.046
Wtt (B4)	-0.004	0.020	-0.10	0.921
Sf (B8)	0.679	0.194	3.51	0.001
EH ASC (B9)	0.884	0.524	1.69	0.092
Wk ASC (B2)	0.630	0.762	0.83	0.410
Lambda (PC,EH)	0.873	0.289	3.10	0.002
Lambda (GB,WK)	0.675	0.251	2.69	0.071
Lambda (DO)	0.402	0.210	1.91	0.056
Final LL function	-838.22			
AIC (Akaike information criterion)	1704.00			
Chi-squared test result	2446.50			
Probability that model is not significant	0.000			

Compared to the access only MNL, this model is not a significant improvement. All the nesting parameters are below 1.0 and have significant t-test values, which means that the correlations in the unknown utilities are significant. The LL is about two points stronger than the access MNL, indicating an improvement in predictive power. The ASCs are more significant in this model compared to the MNL non-trader model All the coefficients are similar to the MNL (in sign and significance), including the peculiarity over waiting time. The value of time is also about R40/hr (b5/b6). The ratio of wkt/IVT is about 1.8, meaning that every one minute of IVT is worth two minutes of walking time. This makes sense as people tend to dislike walking relative to being in a vehicle.

Table 4-7 outlines the correlation matrix for the access NL.

Table 4-7: Correlation matrix of access NL

	B11	B5	B6	B0	B7	B1	B3	B4	B8	B9	B2	S_MT	WK_M	DO
B11	1.00	0.12	0.12	-0.01	0.05	0.06	0.13	-0.02	-0.10	-0.02	-0.15	0.10	0.20	0.12
B5	0.12	1.00	0.23	-0.31	0.12	0.02	0.17	-0.07	-0.22	-0.35	0.05	0.20	0.52	0.07
B6	0.12	0.23	1.00	-0.29	0.39	-0.21	0.17	-0.04	-0.18	-0.41	-0.13	0.74	0.42	0.46
B0	-0.01	-0.31	-0.29	1.00	-0.67	0.51	-0.03	0.03	0.09	0.68	0.37	-0.19	-0.23	-0.66
B7	0.05	0.12	0.39	-0.67	1.00	-0.09	0.06	-0.01	-0.07	-0.03	-0.05	0.51	0.15	0.19
B1	0.06	0.02	-0.21	0.51	-0.09	1.00	-0.14	-0.48	-0.15	0.61	0.67	-0.16	0.15	-0.72
B3	0.13	0.17	0.17	-0.03	0.06	-0.14	1.00	-0.03	-0.27	-0.06	-0.68	0.14	0.38	0.05
B4	-0.02	-0.07	-0.04	0.03	-0.01	-0.48	-0.03	1.00	0.01	0.04	0.00	-0.03	-0.09	-0.03
B8	-0.10	-0.22	-0.18	0.09	-0.07	-0.15	-0.27	0.01	1.00	0.12	0.01	-0.14	-0.41	-0.08
B9	-0.02	-0.35	-0.41	0.68	-0.03	0.61	-0.06	0.04	0.12	1.00	0.45	-0.05	-0.31	-0.81
B2	-0.15	0.05	-0.13	0.37	-0.05	0.67	-0.68	0.00	0.01	0.45	1.00	-0.10	-0.05	-0.59
S_MT	0.10	0.20	0.74	-0.19	0.51	-0.16	0.14	-0.03	-0.14	-0.05	-0.10	1.00	0.35	0.37
WK_M	0.20	0.52	0.42	-0.23	0.15	0.15	0.38	-0.09	-0.41	-0.31	-0.05	0.35	1.00	0.23
DO	0.12	0.07	0.46	-0.66	0.19	-0.72	0.05	-0.03	-0.08	-0.81	-0.59	0.37	0.23	1.00

Again, most of these correlations are very low, which is good. The high correlations involve either the insignificant ASCs or the nesting parameters. Correlations between the ASCs indicate that there is an unknown utility that overlaps with both ASCs. Similar correlations between walking time and the walking ASC occurred in the access model.

Best-fitting nested structure with non-traders included in the sample

The sample of access data with non-traders used in section 4.1.1 was re-used here but it was modelled with the best-fitting nested logit model structure from earlier in this section. The results of this nesting structure with non-traders included are shown below in Table 4-8.

Table 4-8 Best-fitting nested logit model on access SP data with non-traders

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Probability that coefficient is not significant
IVT (B5)	-0.008	0.016	-0.48	0.628
F (B6)	-0.041	0.009	-4.31	0.000
Gen (B11)	-0.751	0.283	-2.65	0.008
PC ASC (B0)	0.890	0.529	1.68	0.093
Pk (B7)	-0.043	0.019	-2.31	0.021
Gb ASC (B1)	0.538	0.397	1.35	0.176
Wkt (B3)	-0.032	0.021	-1.54	0.124
Wtt (B4)	-0.003	0.030	-0.09	0.926
Sf (B8)	0.545	0.161	3.39	0.001
EH ASC (B9)	0.464	0.458	1.01	0.311
Wk ASC (B2)	0.588	0.550	10.07	0.285
Lambda (PC,EH)	1.035	0.283	2.75	0.005
Lambda (GB,WK)	1.235	0.315	2.45	0.003
Lambda (DO)	1.159	0.315	1.71	0.003
Final LL function		-1014.21		
AIC (Akaike information criterion)		2056.40		
Chi-squared test result		3024.42		
Probability that model is not significant		0.000		

The behaviour of the ASCs of the MNL nested model including non-traders was similar to that of the ASCs in the version without non-traders.

The more interesting point here is that all the nesting parameters increased in magnitude and decreased in significance. All the nesting parameters exceeded 1 in magnitude. This means in practice if using the non-trader data, the nesting structure should be discarded and the MNL accepted.

This result was obtained for every nesting structure attempted with the sample with non-traders included. The most likely explanation is that the behaviour that underlies the non-trader data does not conform to the utility-maximising assumption. The non-trader data therefore contradicts the assumptions which support the nested model structure. Therefore, the nested structure should be discarded in favour of the MNL.

4.2 EGRESS MODELS

Section 4.2 outlines the models developed on egress data. It is presented in the same format as the section of the access SP data. The baseline MNL is shown and compared with the MNL estimated on the sample with non-traders added. The best-fitting nested logit model is presented and compared to the same structure with non-traders added to the data. The alternative nesting structures tried are shown and their key results summarised. Finally, the results of a cross-nesting structure are shown.

4.2.1 Egress MNL

This section shows the MNLs with and without non-traders included.

MNL without non-traders

As with the access data, the first model run on the egress data was a standard MNL. This model was estimated to establish a baseline LL and AIC from which to measure other parameters. The results and utility equations are shown below.

U(gb) = b0 + b3*wkt + b4*wtt + b5*ivt + b6*f + b8*sf

Equation 4-6 Utility of GB alternative

U(eh) = b9 + b5*ivt + b6*f

Equation 4-7 Utility of EH alternative

U(pu) = b5*ivt + b6*f + b11*gen

Equation 4-8 Utility of pickup alternative

U(wk) = b2 + b3*wkt + b8*sf +b11*gen

Equation 4-9 Utility of Wk alternative

Table 4-9: Egress only data best-fitting MNL

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
GB (B0)	-0.004	0.283	-0.01	0.998
WKT (B3)	-0.058	0.019	-3.07	0.002
WTT (B4)	-0.017	0.029	-0.59	0.554
IVT (B5)	-0.039	0.012	-3.20	0.001
F (B6)	-0.026	0.006	-4.31	0.000
SF (B8)	0.245	0.125	1.95	0.051
EH ASC (B9)	-0.057	0.028	-1.99	0.046
Gen (B11)	-0.007	0.150	-0.05	0.964
WK (B2)	0.195	0.449	0.43	0.668
Final LL function	-747.91			
AIC (Akaike information criterion)	1511.80			
Chi-squared test result	1462.25			
P Score (chi distribution)	0.000			

There are about 40 fewer observations in the egress non-trader only data. This reduction is because there were fewer egress traders than access traders in the sample. The chi-square test result shows that the model is statistically significant. The ASC for E-hail is statistically significant and negative, indicating people would rather be picked up than catch an E-hail to their destination, which corresponds to the focus group finding of E-hail as a backup, not a primary, mode. The other ASCs show an insignificant, but positive coefficient, indicating a slight preference for that mode relative to being picked up.

The results of Table 4-10 show the model captures trade-offs between service attributes well. IVT, safety, fare, and walking time are all significant with correct sign coefficients. Gender was only slightly significant. This decrease in significance is because the time constraints of the egress journey (the desire to make it to work on time) start to override gender-specific preferences. The ratio of wkt/IVT is again about 1.5. This result is similar to the findings on the disutility of walk time on the access trip (see 4.1.2). The more time spent in a vehicle, the more unattractive the option will be. One of the interesting findings of the egress model is the large increase in the VOT among the Gautrain trader population relative to their access journey. The VOT of the egress model is double the VOT of the access model (b5/b6)= R88.56/hr. This result is likely an indication that, on the second leg of the journey, the person is more anxious to reach their destination, which would be the case if the arrival time at the destination is fixed and more important than the departure time. Thus passengers become more sensitive to the speed of the egress mode option than that of the access mode.

The correlation matrix of the egress MNL shows similar correlation patterns to the access MNL. However, there are high correlations between ASC's with low t-test scores as seen in Table 4-10.

Table 4-10: Correlation matrix for the egress MNL

	B0	B3	B4	B5	B6	B8	B9	B11	B2
B0	1.00	-0.44	-0.78	-0.03	0.20	-0.17	0.03	0.01	0.52
B3	-0.44	1.00	0.00	0.06	0.03	-0.09	0.05	-0.04	-0.88
B4	-0.78	0.00	1.00	0.03	0.00	-0.01	0.01	0.08	0.00
B5	-0.03	0.06	0.03	1.00	0.07	-0.02	0.01	0.09	0.16
B6	0.20	0.03	0.00	0.07	1.00	-0.01	-0.38	0.1	0.28
B8	-0.17	-0.09	-0.01	-0.02	-0.01	1.00	0.00	-0.01	-0.07
B9	0.03	0.05	0.01	0.01	-0.38	0.00	1.00	0.00	-0.09
B11	0.01	-0.04	0.08	0.09	0.1	-0.01	0.00	1.00	-0.17
B2	0.52	-0.88	0.00	0.16	0.28	-0.07	-0.09	0.12	1.00

As shown the IIA/IID, assumptions mostly hold but for a few exceptions. The insignificant Gautrain bus ASC correlates with walking time and waiting time, possibly for the same reasons touched on in the access MNL. Walking time has an almost perfect correlation with the walking ASC. The Gautrain bus correlates with walking time.

MNL with non-traders

Given the same utility equations, the results of the MNL model with non-traders included are shown in Table 4-11.

Table 4-11: Egress MNL with non-traders included

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
GB (B0)	.0171	.261	0.07	.948
WKT (B3)	-.044	.017	-2.65	.008
WTT (B4)	-.017	.026	-0.67	.503
IVT (B5)	-.029	.010	-2.77	.006
F (B6)	-.022	.005	-4.09	.000
SF (B8)	.206	.110	1.87	.062
EH ASC (B9)	-.481	.123	-3.92	.001
Gen (B11)	.010	.137	0.73	.468
WK (B2)	.182	.401	0.45	.650
Final LL function	-953.15			
AIC (Akaike information criterion)	1924.30			
Chi-squared test result	1886.63			
P Score (chi distribution)	0.000			

The results of the addition of non-traders are significantly worse than in the case of the access MNL. The LL drops 200 points. Also, there is no decrease in the significance of any ASC by even a small amount. This indicates that all the non-trader egress data are probably an irrational behaviour type and should not be included in the models.

4.2.2 Egress nested logit model

As in the discussion of the access data, Section 4.2.2 first considers the alternate nesting structures before presenting the best-fit model with non-traders and the best-fit model without non-traders. The utility equations used remain unchanged from the egress MNL in 4.2.1.

Rejected nesting structures

Figures 4-5 and 4-6 show the nested structures tested on the egress SP data without non-traders included.

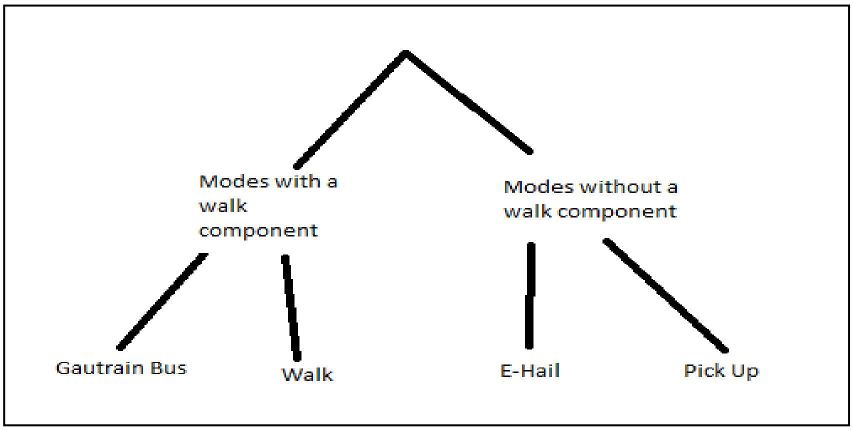

Figure 4-5 Alternate nesting structure 1

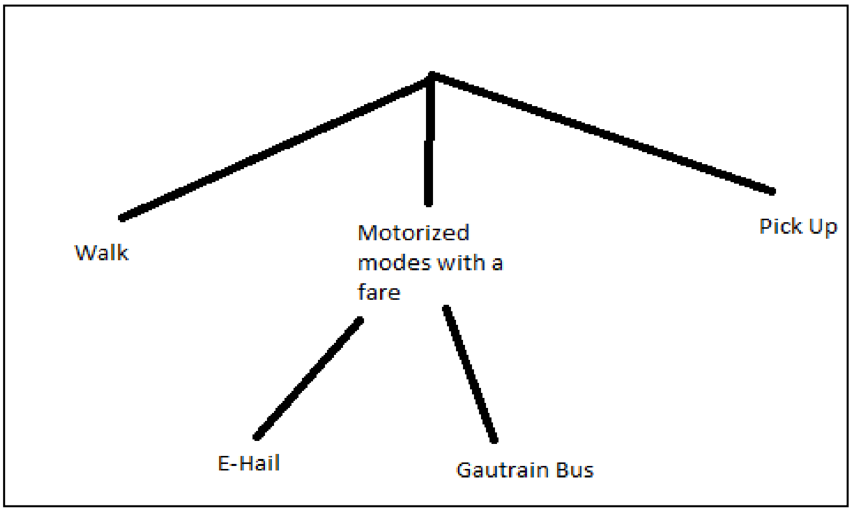

Figure 4-6 Alternate nesting structure 2

The key results are summarised in Table 4-12. Given the results below, use of the lowest magnitude LL as the criterion for the best fitting model structure was insufficient. Alternate Nesting structure 2's nesting parameter for the walk nest was of such high magnitude that walk would never have been chosen in the model. Given that it is unlikely that walk will never be chosen as an egress mode Alternate Nesting structure 2 was discarded. The choice of the best-fitting structure then came down to the Lowest LL and lowest magnitude of nesting parameters.

Table 4-12: Nested model structure summary for egress SP data

Model structure	Log likelihood	Nesting parameters	Reason for rejection
Best-fit model	-745.35	Lambda (eh,gb)-0.56 Lambda (pu, wk)-0.56	n/a
Alternate Structure 1	-747.07	Lambda (gb,wk)-1.57 Lambda (eh, pu)-1.00	Lower LL than best-fit model and nesting parameters >=1
Alternate Structure 2	-741.27	(WK)-50.2 (GB, EH)- 0.85 (PU)-0.65	Though it has a slightly lower LL the nesting coefficient for walking is so high that walking will virtually never be chosen in the model which is unrealistic

Best-fitting nested logit model without non-traders

The best-fitting nested model on the egress data divided the modes into two nests. It groups the private modes (pickup, walking) in one nest and the public modes (E-hail, Gautrain bus). The structure is illustrated in Figure 4-7.

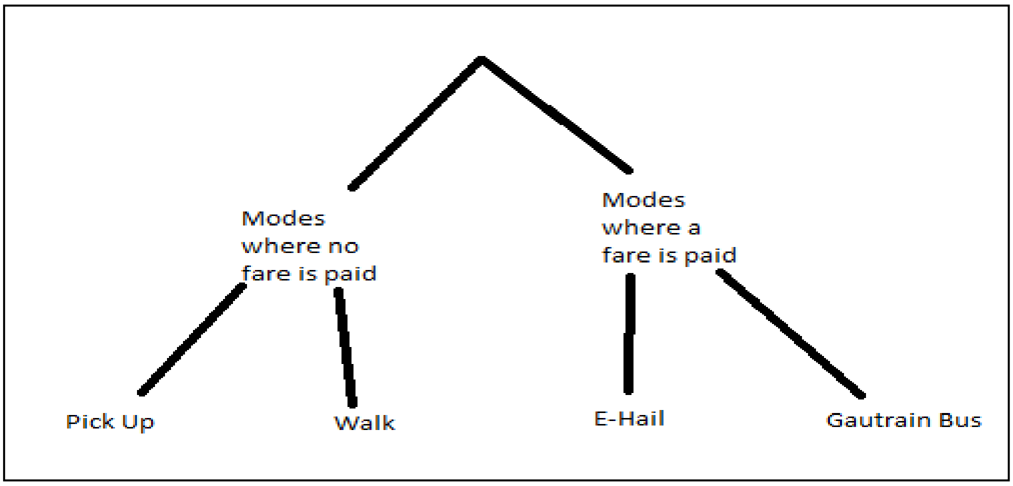

Figure 4-7 Best-fitting nested logit structure

The structure in Figure 4-7 showed a reasonable improvement over the baseline MNL and outdid all other nesting structures. There is a question over whether E-hail is a public or private mode; it is often a trip made privately but driven by someone else. In the access nesting structure, E-hail nested best with a private car. Yet with an egress nesting structure, the Gautrain bus made the best nesting pair with E-hail. The results of the nested model are shown in Table 4-13.

Table 4-13: Best-fitting nested logit model on the egress data

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
GB (B0)	-0.004	0.724	-0.01	0.998
Gen_ (B10)	-0.224	0.213	-1.05	0.293
WKT (B3)	-0.101	0.035	-2.89	0.004
WTT (B4)	-0.019	0.041	-0.29	0.771
IVT (B5)	-0.055	0.017	-3.14	0.002
F (B6)	-0.040	0.016	-3.49	0.001
SF (B8)	0.367	0.180	2.04	0.042
EH ASC (B9)	-0.500	0.650	-0.77	0.442
WK ASC (B2)	0.466	0.668	0.70	0.485
Lambda (eh,gb)	0.558	0.165	3.37	0.001
Lambda (pu, wk)	0.558	0.229	2.47	0.015
Final LL function		-745.35		
AIC (Akaike information criterion)		1514.70		
Chi-squared test result		1636.80		
P Score (chi distribution)		0.000		

The best-fit model structure with non-traders included is shown in Table 4-14.

Table 4-14: Best-fit nesting structure with non-traders included

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
GB (B0)	-0.508	0.670	-0.76	0.448
WKT (B3)	-0.069	0.026	-2.62	0.009
WTT (B4)	-0.014	0.034	-0.42	0.672
IVT (B5)	-0.035	0.013	-2.67	0.008
F (B6)	-0.031	0.009	-3.36	0.001
SF (B8)	0.275	0.142	1.92	0.054
EH ASC (B9)	-1.081	0.618	-1.75	0.080
Gen (B10)	0.1082	0.180	0.60	0.549
WK (B2)	0.292	0.509	0.57	0.566
Lambda (eh,gb)	0.606	0.170	3.58	0.001
Lambda (pu, wk)	0.781	0.306	2.55	0.011
Final LL function		-953.15		
AIC (Akaike information criterion)		1924.30		
Chi-squared test result		1886.63		
P Score (chi distribution)		0.000		

These results show the same trends as when the non-traders were added to the access data nested model. The nesting parameters increased, though in this case not past the threshold of 1, and the ASCs on the whole increased in significance. Given the lower LL, lower t-test scores and higher magnitude nesting parameters the model without non-traders in the sample performs significantly better than the version of the model with non-traders in the sample.

4.3 MODELLING WITH SOCIO-DEMOGRAPHIC PARAMETERS

This section shows the estimation of the socio-demographic parameters of income and age. The best-fitting nested model structures for access and egress data were selected, and income and age variables added to the utility functions of each of the modes in turn, and the results recorded. Only the models with the best results are reported here.

Age and income were added using effects coding. Given that there are three categories of each variable, two parameters are added to each model. The coding is outlined below in Table 4-15 and 4-16.

The effects coding should be interpreted in the following way. If Age 1 has a significant positive β coefficient it means that older people, relative to younger people, chose the modes where age 1 is included in the utility equation more often. If the β coefficient is negative and significant it means that older people, relative to younger people, chose the modes where age 1 is a variable less often. Age 2 results should be interpreted in a similar manner however age 2 compares the choices of middle-aged people relative to younger people. If the β coefficient for both age variables is insignificant it indicates there is no significant difference between the choices of the two demographic groups.

Income variables work in a similar manner to the age variables. Income 1 compares the choice frequencies of high-income people to those low-income people. Income 2 compares the choice frequencies of middle-income people to those of low-income people.

Table 4-15: Effects coding values for the Age parameter

Variable Levels	Age1	Age2
Older (> 34 years)	1	0
Middle (25-34 years)	0	1
Younger (< 25 years)	-1	-1

Table 4-16: Income effects coding for the income parameter

Variable Levels	Inc1	Inc2
High (> R30k/month)	1	0
Middle (R10k-R30k/month)	0	1
Low (< R10k/month)	-1	-1

The access model and then the egress model are discussed in the following section.

4.3.1 Access model

The expanded utility equations in the access model are as follows. The model structure is found in 4.1.2.

U(do)= b5*ivt + b6*f + b11*gen

Equation 4-10 Drop-off utility

U(pc)= b0 + b5*ivt + b6*f + b7*pk + b12*inc1+ b13*inc2

Equation 4-11 Private car

U(gb)= b1 + b3*wkt + b4*wtt + b5*ivt + b6*f + b8*sf

Equation 4-12 Gautrain bus

U(eh)= b9 + b5*ivt + b6*f + b12*inc1+ b13*inc2

Equation 4-13 E-hail utility

U(wk)= b2 + b3*wkt + b8*sf + b11*gen + b14*age1+ b15*age2

Equation 4-14 Walking utility

The model results are shown below in Table 4-17.

Table 4-17: Access nested model with income and age

Attribute (coefficient code)	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
IVT (B5)	-0.034	0.017	-1.92	0.054
F (B6)	-.0460	0.012	-3.96	0.001
Gen (B11)	-0.467	0.294	-1.59	0.111
Pc asc(B0)	1.164	0.646	1.80	0.072
Pk (B7)	-0.042	0.020	-2.12	0.034
Inc 1(B12)	0.307	0.172	1.79	0.073
Inc 2 (B13)	-0.069	0.130	-0.52	0.600
Gb asc (B1)	0.430	0.752	0.57	0.568
Wkt (B3)	-0.060	0.030	-2.01	0.045
Wtt (B4)	-0.003	0.045	-0.07	0.947
Sf (B8)	0.709	0.200	3.55	0.004
Eh asc (B9)	1.009	0.550	1.84	0.067
Wk asc (B2)	0.633	0.839	0.76	0.449
Age 1 (B14)	0.443	0.301	1.47	0.141
Age 2 (B15)	0.093	0.226	0.41	0.682
Lambda (PC,EH)	0.863	0.276	3.13	0.002
Lambda (GB,WK)	0.602	0.222	2.71	0.007
Lambda (DO)	0.400	0.214	1.87	0.062
Final LL function	-833.73811			
AIC (Akaike information criterion)	1704			
Chi-squared test result	2455.46344			

This model shows a significant improvement from the best-fit nested logit model. The LL has increased five points and the other results discussed previously remain. The addition of income and age has interesting results. Older people (>34) tend to prefer walking, unlike young people. This could be for health reasons or different lifestyle habits between older and younger generations. This result has not been reported in the literature before and warrants further research as to the peculiar habits relating to walking in South African populations. The middle age group coefficient is not statistically significant.

The income coefficients display intuitively correct results. Richer people are significantly more likely to prefer E-hail and private car use than poorer people. The high income of rich people means they would be more amenable to expensive modes than low income people. The results for middle-income people show insignificant t-test coefficients.

4.3.2 Egress model

The same process is followed for the egress data. The structure used was adapted from 4.2.2, and the utility equations of the final model are shown below.

U(gb) = b0 + b3*wkt + b4*wtt + b5*ivt + b6*f + b8*sf + b13*inc1 + b14*inc2

Equation 4-15 utility of Gautrain bus

U(eh) = b9 + b5*ivt + b6*f + b11*age1 + b12*age2

Equation 4-16 utility of E-hail

U(pu) = b5*ivt + b6*f + b10*gen

Equation 4-17 utility of pickup

U(wk) = b2 + b3*wkt + b8*sf + b10*gen

Equation 4-18 utility of walking

The results of the model run are shown below in Table 4-18.

Table 4-18: Egress model including age and income.

Attribute (coefficient code	Coefficient	Standard error	T-test score	Wald value (probability parameter is not significant)
Gbs_asc (B0)	0.223	0.281	0.80	0.421
Wkt (B3)	-0.039	0.018	-2.05	0.041
Wtt (B4)	-0.001	0.019	-0.36	0.722
IVT (B5)	-0.032	0.011	-2.80	0.005
F (B6)	-0.022	0.006	-3.52	0.004
Sf (B8)	0.136	0.103	1.32	0.186
Inc1 (B13)	-0.185	0.115	-1.60	0.110
Inc2 (B14)	0.020	0.074	0.28	0.782
EH_asc (B9)	-0.343	0.134	-2.56	0.011
Age1 (B11)	0.032	0.128	0.25	0.803
Age2 (B12)	-0.289	0.122	-2.36	0.018
Gen (B10)	0.004	0.141	0.03	0.976
Wk (B2)	0.268	0.315	0.85	0.396
Lambda (eh,gb)	1.767	0.690	2.56	0.011
Lambda (pu, wk)	1.179	0.340	3.46	0.001
Final LL function	-741.30			
AIC (Akaike information criterion)	1512.60			
Chi-squared test result	1644.87			
P Score (chi distribution)	0.000			

Compared to the access model, this model shows less of an improvement over the original egress model. Only one of the four socio-demographic variables are significant at the 95% confidence level, suggesting that egress behaviour is more uniform across socio-demographic groups, than access behaviour. In addition, the addition of the new variables causes the nesting parameters to rise above 1, and therefore make this effectively an MNL.

Regarding the age data, Age1 shows no significant difference between older people and younger people in mode preference. Older people are not significantly more likely to prefer E-hail than younger people. This might be because, as shown before, people become more time-sensitive on the egress side of their journey and thus younger people would be more willing to pay the extra cost for an expensive mode like E-hail. This would reduce the difference between them and the older people in the survey. However, the significant negative coefficient for age 2 indicates middle-aged people strongly dislike E-hail compared to younger people.

Income 1 presents an interesting result. The coefficient indicates Gautrain's wealthier people are less inclined to use the Gautrain bus than poorer people. If this is true it is likely because they have more spending power, and therefore are not concerned by the higher cost of more comfortable modes like private car and E-hail. However, there is a degree of uncertainty to this observation because Income 1 has a t-test score of 1.6 which is both smaller than 1.96 and even smaller than the t-score of Income 1 in the access model (1.79). While a t-score of 1.6 is on the lower end of valid results it still a significant score on a t-test. Also, it seems intuitively reasonable for high income people to be less inclined to use a mode where the main advantage is its low cost. Given these mitigating factors the result seems worth noting. Though further research might be needed to confirm this finding.

Income 2 is insignificant in the access and egress models shown in this section. This indicates that the income variable probably needed different boundaries. The lack of significant difference between middle income people and low-income people indicates the two probably should have been combined into one category.

4.4 FINAL DISCUSSION

This section offers an analysis of the modelling results. It deals with access behaviour, egress behaviour, effects of non-trader data and finally the comparison between access and egress behaviour seen in the models.

4.4.1 Access behaviour

The best-fitting access models (without non-traders) show findings that advances our understanding of access mode choice behaviour amongst rail commuters. Gender influences decision-making. Women are less inclined to walk or use drop-off modes. All the attribute coefficients have rational signs. All the trip times have negative coefficients, all the costs have negative coefficients (indicating they are both disutilities), and high safety has a positive coefficient (indicating it adds to utility). These signs all make sense. The ASCs only become significant when non-trader data is added, as would be expected since non-traders by definition do not respond to trip attributes.

The value of IVT works out to about R40/hr which is significantly higher than the average South African public transport user's value of time (Hayes & Venter, 2017). Interestingly, waiting time is not statistically significant (waiting time t-test much lower than 1.96). The value of walking time is almost double the value of IVT (R85/hr). This seems reasonable and in line with results reported in the literature, as people tend to dislike walking during a commute, partly due to safety and comfort preferences. Waiting time for an access mode is insignificant, suggesting that Gautrain users are not sensitive to waiting for buses. The model also shows that parking is a statistically significant factor. Parking cost coefficients are not as sensitive as IVT or walk time. This finding shows that adjusting the parking fare could have an effect, but it is not as effective a strategy as reducing walk times or IVT.

The findings around the best-fitting model structure are credible. The best-fitting nesting structure seems to indicate that people perceive E-hail as like private car, and that the walking component of the Gautrain bus and walking modes seems to cause them to be correlated together in the data. This provides new evidence regarding the structure of correlations between different access modes and how they are perceived that has not been reported in the literature before.

4.4.2 Egress behaviour

The egress models (without non-traders) are similar to the access models in many respects, but also provided some unexpected results. The first noticeable difference between the access and egress models are the estimated ASCs for each of the modes for the models without non-traders.

Nearly all the ASCs in the egress models are insignificant except for E-Hail which is negative and statistically significant. This indicates that respondents significantly prefer being picked up from the station to E-hailing, whereas there was not one statistically significant ASC in the access models. The other major difference is the value of IVT is R81/hr, nearly double the value of IVT in access models (the reasons for this are discussed in Section 4.5.4). The final significant difference between the access and egress models is that the gender coefficient was insignificant in this model. This could indicate that people are more anxious to reach their destination on time on the egress trip. That anxiety might erode preferences that are unique to a particular gender.

There are also some similarities between the two types of choices. Of the three different kinds of time, walking time is the most valuable to people (R134/hr), followed by IVT at (R81/hr) then waiting time (R41/hr). In the access models, the same order is observed; walking time is the most valuable followed by IVT and finally waiting time. Waiting time is also insignificant (in terms of t-score) in both the access and the egress models. High levels of safety are also a significant positive utility in the model, showing that safety concerns are present on the whole of the journey.

The best-fitting nested structures findings are interesting. E-hail shows a strong correlation with the Gautrain bus and pickup correlates strongly with walking. This suggests the perceptual correlation patterns of passengers on the egress journey is different. E-hail and the Gautrain bus are the only modes where a fee is paid. Pickup and walking are the modes where no fee is paid. These show the highest levels of unknown utility correlation. The change in best-fitting nested structure is a notable difference to the access models.

4.4.3 The effect of non-traders

The effects of including the non-traders are the same for both access and egress data. These effects include a significant reduction in the LL of the models. The ASCs mostly show significantly increased t-score values (increasing statistical significance) and increased magnitudes. The signs of the ASCs are unchanged from the models without non-traders.

While the non-trader data generally decreases the t-scores of individual attributes it does not change the signs of attributes or the relative magnitudes of attributes relative to each other. For instance, the value of IVT of the access models is still about half the value of IVT of the egress models. Walking time is the most valuable of the three times, followed by IVT and then waiting time, which always remains statistically insignificant.

The other effect of the non-trader data is that it increases the magnitude of nesting parameters; they often equal more than 1, collapsing the nesting structure. This trend suggests that non-trader data reduces the correlation between the unknown utilities.

4.4.4 Differences and similarities between access and egress behaviour

In general, it is concluded that the models for access trips and egress trips are similar. Safety, fare, IVT and walking time are all highly significant parameters. Waiting time proved to be insignificant on both trips. Walking time is the most sensitive of the three times, followed by IVT and then walking time. This applies to both access and egress behaviour, although people are more sensitive to time penalties on the egress journey.

The major difference between the two trips is the value of IVT. The value of IVT on access/egress to Gautrain is more significant than on any other form of public transport in South Africa, likely due to the users' much higher income relative to South Africa as a whole. However, the VOT for the egress journey is about double the value of the access journey as shown in Table 4-19. The higher value of both walking time and IVT for the egress models could indicate that people become more anxious on the last-mile journey of their trip. They are more aware of the possibility of being late and this probably makes people anxious enough to be willing to pay significantly higher fares to shorten their egress trip.

Table 4-19: Value of Time summary

	VOT (R/hr)
Access MNL	40
Access NL	41
Egress MNL	90
Egress NL	81

There are several explanations for why waiting time is not a significant variable in this data. The first is that people may not wait at bus stops any more. One can now monitor where Gautrain buses are on the Gautrain mobile app. Thus, a rider waits in their home until the bus is close by and takes the bus when it suits them. Secondly, the habitual nature of these trips (e.g., using the bus as a morning access trip) means people might work out roughly when the bus arrives at their stop within a margin of error of five to ten minutes. Thus, they know when to leave to minimise their waiting time, eliminating it as a significant issue.

5 SCENARIO TESTING

With the models finalised, the question is then: What is the value of these models? There have been reports in the South African media indicating the Gautrain rail network will be expanded (Business Insider, 2020). The expansion means new stations will be constructed. This means the GMA will need to forecast parking demand and bus demand to plan routes in the various station catchment areas. Given the problems outlined in Section 1 and Section 3.2.1, it would be of some benefit to the GMA if they could do the following:

- Improve the occupancy rate of the Gautrain bus service.
- Reduce the demand for parking in the Gautrain stations.

For these two aims, the GMA has control over several variables:

- Bus waiting time – Not relevant due to low significance but could be decreased by reducing headway between buses on a route.
- Bus walking time – Significant. It can be decreased by building more bus stops and operating more routes. Its mean elasticity shows it influences bus mode share, but the main effect on ridership is IVT.
- Bus IVT – Significant. It can be reduced by building priority bus lanes (or securing the use of the BRT system in Pretoria or Johannesburg) This attribute has the most significant effect on bus mode share.
- Bus cost – Significant, but less strong than the time parameters and likely to reduce the profitability of the Gautrain bus system. It was not tested.
- Car parking cost – Significant, and can be changed by the GMA at any time. It has a noticeable effect on parking ridership.

This chapter shows that the models that have been developed can give an indication of the effect that the variables within the control of the GMA have on bus occupancy and parking demand.

To this end several exercises are carried out in this chapter. First, the mean elasticities of the variables within the GMA's control are calculated. For the access journey, the MNL is used for the sample with non-traders. The best fitting NL model is used for the sample without non-traders. For the egress journey, the nested models are used for each sample. This is because the access nested logit model collapses into an MNL when the non-traders are added (the nesting parameters all increase past 1). Then a revenue-neutral example of this planning is explored.

5.1 MEAN ELASTICITIES

The following section calculates point elasticities for both access and egress journeys of the samples with and without non-traders. To carry this out it was assumed all the attributes are averaged to their mean values (except for gender, which is equal to female as they were the majority respondents). The values assumed for each variable are summarised in Table 5-1.

Table 5-1: Values assumed to calculate mean elasticity of important variables.

Private car	
IVT (min)	20
Pk (R)	24
F (R)	7
Drop-off	
IVT (min)	20
F (R)	22.50
Gen	1
E-hail	
IVT (min)	20
F (R)	40
Walking	
Sf	0.5
Wkt (min)	18
Gautrain bus	
Wkt (min)	7.5
IVT (min)	20
SF	0.5
F (R)	10.5
Gen	1

The value of the controllable, significant variables in Table 5-1 are changed by +10% and –10% respectively. Then the change in either bus ridership (for bus variables) or car ridership (for parking cost) is recorded for both trader and non-trader samples. These are shown in Tables 5-2 to 5-5.

5.1.1 Access elasticities

Tables 5-2 and 5-3 show the effect the non-traders have in reducing the effectiveness of each variable. The removal of the non-traders from the data sample increases the elasticities of each variable significantly.

Table 5-2: Mean elasticities for the access model with non-traders

	Change in ridership	
	+10%	-10%
Parking cost	-2.0%	2.2%
Bus cost	-0.1%	0.1%
Bus IVT	-1.6%	0.8%
Walking time	-0.1%	0.04%

Table 5-3: Mean elasticities for the access models without non-traders

	Change in ridership	
	+10%	–10%
Parking cost	-2.1%	2.2%
Bus cost	-0.7%	0.8%
Bus IVT	-1.9%	2.1%
Walking time	-0.7%	0.80%

5.1.2 Egress elasticities

Tables 5-4 and 5-5 show the change in ridership of the Gautrain bus for the change in value of specific variables in the egress models. The non-traders affect the elasticity of a variable in the same way the effect the elasticities of variables in the access model. It is worth noting that the elasticities of the bus variables are higher than the elasticities of the same variables in the access data. This could be because there are fewer modes to choose from the effect of each variable is more significant.

Table 5-4: Mean elasticities for the egress model without non-traders

	Change in ridership	
	+10%	–10%
Bus cost	-1.1%	1.1%
Bus IVT	-5.6%	2.8%
Walking time	-2.0%	2.0%

Table 5-5: Mean elasticities for the egress model with non-traders

	Change in ridership	
	+10%	-10%
Bus cost	-0.7%	0.8%
Bus IVT	-1.6%	1.7%
Walking time	-1.1%	1.2%

5.2 HYPOTHETICAL SCENARIOS

For testing the usefulness of this work, two hypothetical scenarios are examined. One scenario uses increased parking revenue to reduce mean walking time for the Gautrain bus users. The other scenario uses increased parking revenue to reduce mean IVT for Gautrain bus users.

5.2.1 Walking time scenario

Consider a new station built for the expansion of the Gautrain service. In the morning peak hour, a total of 6 400 morning commuters come through the station with a 75:25 split between access and egress traffic. The percentage of non-traders on both sides of the journey is assumed to be 25% (see Section 3.3.1). Consider the best-fitting non-trader models, assuming the variable values are equal to those in Table 5-1. The starting mode splits of the trader sample are shown in table 5-6 and table 5-7.

Table 5-6: Starting egress mode split

Pickup	Walking	Gautrain bus	E_Hail
22.7%	39.3%	28.4%	9.5%

Table 5-7: Starting access Mode split

Drop-off	Walking	Gautrain bus	E_Hail	Private car
13.3%	30.7%	14.89%	13.0%	28.0%

Considering the problems mentioned in Chapter 1, it would be worth seeing how much extra revenue could be generated from increased parking fares and then exploring what might happen if that money was used to fund more bus routes (reduce walking time) or add more buses to reduce bus headway.

Currently, the relevant access mode share is as follows.

Gautrain bus = 536 passengers

Private Car = 1 008 passengers

The relevant egress mode share is as follows.

Gautrain bus = 114 passengers

Currently, the parking fare for a train rider is R21 per day. This results in a current parking revenue of R21 168 per day. If the parking fare is increased to R30 the mode share of private cars goes down to 22.9%. However, given the increase in price the parking revenue increases to R24 732 per day. Therefore a total of (1 008–824)=184 parking bays are saved and a total of (24 732–21 168)= R3 564 per day gain in revenue is accrued.

Before adjusting parking revenue, the Gautrain bus carries 536 passengers at R15 per ticket. This level of ridership gives a total revenue of R8 040 from bus fare per day. Assume the new station has two main bus routes of a total distance of 10 km with six buses each. With the increase in revenue of R3 564 per day with the higher parking fees, the GMA can improve conditions for the Gautrain bus service.

To improve the conditions of the Gautrain bus mode, the GMA has two options. The first thing they could do is add a new route to reduce walking time. The second would be to improve the speed of the buses; however, they do not have much control over this factor due to the congested nature of the road network in Gauteng and the high cost of priority lane infrastructure (this will be explored later). Aropet (2019) indicates adding a bus to a route costs the amounts shown in table 5-8.

Table 5-8: Assumptions on Bus cost

Vehicle operating cost	R13.06 per vkm/day
Maintenance cost	R16.65/day
Driver cost	R712/day

Assuming this new route is about 15 km long and a bus does ten trips per day the total cost is about (13.06x15x10) +712 + 16.65 = R1 959 per day. Assuming the new route does not add new passengers to the Gautrain (a conservative approach), running two buses on this route costs R3 918. When parking revenue is added to passenger revenue the costs of running the extra bus route will create a daily loss of R400 for the GMA (R3 564–R3 918). However, extending the area covered by the bus routes around the station will have the effect of reducing the average walking time to the bus stop. Say, for example, it reduces average walking time from 7.5minutes to 5minutes. When this change is combined with the increase in parking fare the Gautrain bus mode share increases to 18,67% of passengers. This adds an extra 136 passengers per day for the Gautrain bus, yielding an extra revenue of R15x136=R2 040 per day. These two changes cause a net profit of about R1 600 per day (R3 564+2040)–R3 918). This extra revenue could help secure deals to create exclusive lanes for Gautrain buses to reduce IVT levels in their buses. This reduction in walking time also affects the egress journey of the morning peak period. This increases the bus ridership on the egress trip by 27 passengers. Given the same ticket price, the station would generate extra revenue of (27xR15) R405 per morning peak period, increasing the profit per day to R2 005.

5.2.2 IVT adjustment scenario

As mentioned previously, the other viable strategy for the GMA to increase bus ridership is to reduce mean IVT. This can be done by allowing the Gautrain buses to drive outside the main traffic stream by way of priority lane infrastructure (similar to a BRT system). Del Mistro and Aucamp (2000) provide some average costing figures for this kind of infrastructure as shown below in Table 5-9. These values have been adjusted for inflation to the rand value in 2020.

Table 5-9: Average costs of BRT priority lane infrastructure

Variable (Unit)	Value
Cost of way (Rm/lane-km)	1.970
Land cost – CBD/Commercial (Rm/lane-km)	1.649
Land cost – Outer section (Rm/lane-km)	0.434
Land cost – Residential (Rm/lane-km)	0.198
Minimum cost of station/stop (Rm)	0.4
Life of terminals (years)	20

For this exercise, consider the station described in Section 5.2.1. The costs of bus stops are assumed to be R0.1 million (as Gautrain is not operating a BRT and does not need the more sophisticated stops that are part of a BRT system). Assume that no new routes are added, each route has three stops currently, and the routes run in CBD areas only. These assumptions result in the following total for upgrading a single route to a priority lane infrastructure route:

(0.1x3)+(1.970+1.649)x10 = R36.5 million per route.

If the GMA upgrades both routes, it would incur a capital outlay of R73 million. If it increased the parking cost to R30 and the average bus IVT drops to 15minutes due to the upgrades, private car use would drop to 22.3%, a loss of 5.7% of mode share. Conversely, the Gautrain bus rises to 18.91% of mode share, a gain of 4.02%.

The parking revenue is now (1 008x(22.3/100)x30)= R24 084, resulting in a gain in revenue of R2 916. There is a higher number of bus passengers (3 600x0.1891= 680) than the lower walking time scenario, resulting in an increase of 144 passengers per day on the access side. This increases bus fare revenue by (144x15)= R2 160 per day.

Assuming 260 workdays in the average year, the increase in bus and parking revenue results in an extra (2 160+2 916)x260 = R1,3 million extra revenue per year. Even if the egress revenue took the revenue to over R2 million per year it would still take 36 years (longer than the concessionaire has legal rights to the system) to pay it off. In this case, the rational course of action would be to add an extra route or two to reduce the average walking time for passengers, while increasing the parking revenue at the same time.

6 CONCLUSIONS AND RECOMMENDATIONS

This chapter summarises the findings of the study and provides recommendations going forward.

6.1 CONCLUSIONS

This section deals with conclusions from the access research, the egress research, the socio-demographic research and the differences/similarities between access and egress research.

6.1.1 Access conclusions

The access station findings have implications for station management in the future. The models show consistently that walking time is the most sensitive attribute of the various access modes. Section 5.2 showed that a reduction of average walking time by 33 % could increase bus ridership by over 20%. The value of IVT is higher than average public transport systems in South Africa and waiting time is not a significant factor in the choice of access mode. The findings on sensitivity to parking cost show that increasing the parking fare could reduce park and ride usage by close to 25% at a station.

6.1.2 Egress conclusions

The egress finding also provides a reliable understanding of what motivates people on the egress journey. Walk time also has a significant effect on Gautrain bus ridership. The value of IVT is even higher than the value of IVT in the access mode (the same goes for walking time and waiting time). This could have implications for how the GMA structure their bus service fare scheme. Gender has no noticeable effect on the egress choice (unlike the access choice).

6.1.3 Socio-demographic effects

The socio-demographic effect on first-/last-mile behaviour is interesting. Gender has a significant effect on first-mile behaviour, yet no significant effect on last-mile behaviour. The effects of age on first-mile behaviour (old people preferring to walking over younger people) are quite surprising, yet the effect of income seems predictable (richer people preferring E-hail and private car, in contrast with poorer people).

The egress model shows some different effects but nothing that directly contradicts the first-mile behaviour. Wealthy people are much less likely to use the Gautrain bus (which supports the finding from first-mile data that they prefer private modes like E-hail and private car).

Middle-aged people disfavour E-hail relative to younger people and older people, although this effect was not significant.

6.1.4 Differences and similarities between access and egress behaviour

There are several similarities between first-mile behaviour and last-mile behaviour observed during the investigations. Walking time is the most sensitive attribute for both journeys followed by IVT and finally waiting time. The safety of the walking environment is a large positive utility for both journeys. Also, waiting time is statistically insignificant for both journeys due to the reasons mentioned in 4.5.4.

There are also some notable differences. The main difference is in the value of IVT; the value of egress IVT is nearly double the value of access IVT. This has some very practical implications for service planning in station design. For example, the GMA could charge significantly higher fares for trips from the station to final destinations. Walking time is also significantly more valuable to people on the egress side of their journey. This means that aiming to reduce the average walking time to bus stops is a viable strategy to increase the ridership on the Gautrain bus.

The best-fitting nesting structures also imply quite different correlations in the unknown utility components of the modes. This suggests that a real difference exists in the perception of modes between first- and last-mile trips, although the reasons for this cannot be determined with the current data. A common factor between the two is the way non-trader data increases the magnitude of nesting parameters for both types of trips.

In conclusion, a significant finding from this study is that people’s behaviour changes between their access and egress journeys. VOT sensitivity almost doubles from access journey to egress journey. Other attributes sensitivities do not change significantly. It shows people become much more anxious about making it to their destination once they are on their egress journeys.

The non-traders affect the reliability of the models. The effect of the non-traders does not render the models insignificant, but the effect of trip attributes on mode share decreases.

6.2 RECOMMENDATIONS FOR FURTHER RESEARCH

For future research, the following points should be considered.

1. First, increase the size of the offered voucher to the value of, say, R200. This might encourage a bigger sample size and more high-income users.
2. Secondly, increase the length of the advertising campaign. If posters were put up in the station and flyers were handed out with the corresponding website link, a bigger sample might be attracted.
3. Thirdly, a different web platform should be used. Several problems were encountered with Survey Monkey. It is not advised to use the platform again for an SP survey.
4. With a big enough sample, more complex structures like cross-nested models and mixed logit models should be estimated to get a higher resolution understanding of people's behaviour.
5. Research should be done on the different ways the Gautrain buses could be separated from the bulk of the traffic, by building priority infrastructure or by striking a deal with BRT operators. The models built here could estimate the effect of the reduction of IVT on the Gautrain bus ridership.
6. Research should also be done on the possibility of reducing walking time to increase the Gautrain bus ridership, perhaps by putting more stops along the route or by extending the routes the reduction in walking time could increase ridership.
7. The other possibility to reduce parking demand is to investigate a subsidy deal with E-hail companies such as Uber or Bolt. Private car users had some correlation with E-hail users in the modelling exercise. This suggests that a drop in E-hail prices to the station might lower the demand for parking. These models could assist in the feasibility study of such a proposal.
8. Whatever research is undertaken should estimate the models with and without non-traders to understand how much the model changes with the non-traders included. This seems the best practice until it can be determined if someone is a non-trader for utility-maximising reasons.
9. Try different combinations of the number of categories per variable and their limits on the socio-demographic parameters estimated in any model.

7 REFERENCES

Agrawal, A., Schlossberg, M. & Irvin, K. (2008). How far, by which route and why? A spatial analysis of pedestrian preference. *Journal of urban design, 13*(1), 81-98.

Ashford, N. & Benchemam, M. (1990). Passengers' Choice of Airport: An Application of the Multinomial Logit Model *Transportation Research Record 1147*

Aropet, R. (2019). The Feasibility of Public Aided Transit at the University of Pretoria

Business Insider (2020). Gautrain still plans 18 more stations - but money is tight due to Covid-19 https://www.businessinsider.co.za/gautrain-expansion-during-covid-19-2020-7

Bhat, C. (1997). Work travel mode choice and number of non-work commute stops. *Transportation Research Part B: Methodological, 31*(1), 41-54.

Cervero, R. (2001). Walk-and-ride: factors influencing pedestrian access to transit. *Journal of Public Transportation, 3*(4), 1.

Cervero, R. & Gorham, R. (1995). Commuting in Transit Versus Automobile Neighborhoods. *Journal of the American Planning Association, 61*(2), 210-225.

Chakour, V. & Eluru, N. (2014). Analyzing commuter train user behavior: a decision framework for access mode and station choice. *Transportation, 41*(1), 211-228.

Chia, J. & Lee, J. (2015). Variation in the walking time to bus stop by the degree of transit captivity. In: *Australasian Transport and Research forum.*[online] Sydney. Available at: http://www.atrf.info/papers/index.aspx

Chia-Wen, Y., Jin-Long, L. & Chun-Yen, H. (2014) Modeling joint airport and route choice behavior for international and metropolitan airports *Journal of Air Transport Management 39* (2014) 89e95.

Cronkleton, E., (Healthline), What is the Average Walking Speed of an Adult, Available at https://www.healthline.com/health/exercise-fitness/average-walking-speed, date accessed (01-06-2019)

Cushing, B. & Cushing, C. (2007). Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration. Paper presented at the 46th meeting of the *Southern Regional Science Association Charleston*, South Carolina, Charleston, March 29-31

Debrizion, G., Pels, E & Rietveld., P (2007). Modelling the Joint Access Mode and Railway Station Choice. *Tinbergen institute discussion paper*

Del Mistro. R. & Aucamp, C. (2000) Development of a Public Transport cost model. Paper presented at *South African Transport Conference,* Pretoria, South Africa, 17-20 July

Fieldstadt, E., (CBS News), The most dangerous cities in America ranked (online), Available at https://www.cbsnews.com/pictures/the-most-dangerous-cities-in-america/, date accessed (17-04-2019)

Foddy, W.H. (1994). Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research, Melbourne: Cambridge University Press.

Frazer, L. & Lawley, M. (2000). *Questionnaire Design and Administration: A Practical Guide*. Brisbane: Wiley.

Plus 94 research (2017). Market Segmentation Audit Quantitative Report, Gautrain Management Agency, South Africa.

Givoni, M. & P. Rietveld. (2007). The Access Journey to the Railway Station and Its Role in Passengers' Satisfaction with Rail Travel. *Transport Policy, 14*(5): 357-365.

Global Fuel Economy initiative, South Africa: GFEI releases new assessment of fuel economy potential (2018), Available from https://www.globalfueleconomy.org/blog/2018/january/south-africa-gfei-releases-new-assessment-of-fuel-economy-potential, Date accessed (09-05-2019)

Hayes, G. & Venter, C. (2017). Meta-Analysis of Gauteng Stated Preference Surveys. Paper Presented at *5th International Choice modelling conference*. Cape Town, South Africa, 3 – 5 April

Hensher, D., Rose, J. & Greene, W. (2005). *Applied Choice Analysis: a primer*. Cambridge Press. pp. 74–193, 308-366, 479-601. Cambridge, UK

Hensher, D.A. & Greene, W.G. (2001). Specification and estimation of nested logit models. *Transportation Research*, B 36(1), 1–18

Hess, S., Rose, J. & Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transportation Research* Part D 15 (2010) 405-417.

Koh, P.P. & Wong, Y.D. (2016). Influence of Socio-Demography and Operating Streetscape on Last-Mile Mode Choice. *Journal of Public Transportation*, *19*(2)*, No. 2, 2016*

Loo, B.P.Y. (2008). Passengers' airport within multi-airport regions (MARs): some insights from a stated preference survey at Hong Kong international airport. *Journal of Transport Geography. 16*, 117e125.

Nielsen Consumer Insights (2019), *Gautrain User Behaviour and Preferences,* Pretoria, University of Pretoria [Powerpoint Presentation & Audio recording]

Numbeo, Crime index by city (2019), Available from www.numbeo.com/crime/rankings, Date accessed (03-04-2019)

Ortuzar, J. & Willumsen, L. (2014). *Modelling Transport.* 4th Edition. pp.227-241., John Wiley & Sons, United States

Rose, J.M. & Bliemer, M.C.J. (2004). The design of stated choice experiments: the state of practice, Working Paper, Institute of Transport Studies, University of Sydney.

Ryan, S. & Frank, L.F. 2009. Pedestrian Environments and Transit Ridership. *Journal of Public Transportation, 12*(1): 39-57.

Van Soest, D., Tight, M.R. & Rogers, C.D. (2020). Exploring the distances people walk to access public transport. *Transport Reviews, 40*(2), 160–182.

Venter C (2020). Measuring the Quality of the First/Last Mile Connection to Public Transport. *Research in Transportation Economics. Volume 83, November 2020. (https://doi.org/10.1016/j.retrec.2020.100949).*

Wen, C., Wang, W. & Fu, C. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review,* [online] 48(2), 545-554. Available at: https://www.sciencedirect.com/science/article/pii/S136655451100113X

Yang, C.W., Lu, J.L. & Hsu, C.Y. (2014). Modeling joint airport and route choice behavior for international and metropolitan airports. *Journal of Air Transport Management*, 39, 89-95. https://doi.org/10.1016/j.jairtraman.2014.05.01

Appendix A: Sample of 2019 Gautrain Passenger Survey

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Welcome to the Online Gautrain Survey!

Welcome to the online Gautrain Survey and thank you for participating. We estimate that this survey should take you about 15 minutes to complete. We wish to emphasize that your feedback is totally confidential, and will only be used for planning better Gautrain bus and rail services. If you are taking the survey on your phone please rotate your screen to landscape orientation.

* 1. Please confirm the **last six digits** on your Gautrain Card Number. We need this information to make you eligible for the **R50 takealot voucher**. To qualify you need to fully complete the questionnaire:

We would like you to think of **your last completed Gautrain trip that started at your home.** We would like to firstly ask you a series of questions about the details of that trip.

Please note that your **progress is saved** after completing each page.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

General Questions on Your Use of the Gautrain

* 2. Thinking about that last completed Gautrain trip, at what time did you **start** your trip (e.g. from your home) ?

Time

Time AM/PM

hh mm -

* 3. At what time did you **arrive at your final destination** (e.g. your work, college or university, or at your business appointment)?

Time

Time AM/PM

hh mm -

4. Which Suburb did you begin this trip in?

5. Which suburb did your trip end in?

* 6. What is your best estimate of the Gautrain fare you paid (in Rands) for that trip to work, college, university, etc. (i.e. for the train fare only)?

0 South African Rands (R) 200

* 7. How **often** do you make this trip?

* 8. At which Gautrain **station** did you **board the Gautrain?**

* 9. At which Gautrain **station** did you **get off the Gautrain?**

* 10. For this last trip you are describing, how did you travel to **the Gautrain station** where your train trip **began**? Please note Gautrain Bus refers to any size vehicle operated by Gautrain for getting people to their destination.

* 11. How **long** have you been **using** the Gautrain?

* 12. Do you ever use **metered taxis**? If so, **how often**?

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Not Considering OR TAMBO

Thank you for participating in the online Gautrain Survey. Unfortunately, we are not considering trips taken to or from OR Tambo International Airport.

Please continue on to fill in the demographics portion of this survey.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Access Mode: Drop Off

13. You said you were **dropped off** at the station. Which other modes were available to you when making this trip?

- [ ] Non E-Hail Taxi
- [ ] Private Car
- [ ] E-Hail (Uber/Bolt)
- [ ] Walk
- [ ] None of the above
- [ ] Gautrain Bus

* 14. Assuming the following modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your starting Gautrain station? (Select one of the following).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Access Mode: Walking

15. You said you **walked** to the station. Which other modes were available to you when making this trip?

- [ ] Non E-Hail Taxi
- [ ] Private Car
- [ ] E-Hail (Uber/Bolt)
- [ ] Being Dropped off
- [ ] None of the above
- [ ] Gautrain Bus

* 16. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your starting Gautrain station? (Select one of the following).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Access Mode: Private Car

17. You said you took your **car** to the station. Which other modes were available to you when making this trip?

- [ ] Non E-Hail Taxi
- [ ] Walking
- [ ] E-Hail (Uber/Bolt)
- [ ] Being Dropped off
- [ ] None of the above
- [ ] Gautrain Bus

* 18. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your starting Gautrain station? (Select one of the following).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Access Mode: Gautrain Bus

19. You said you took **the Gautrain bus to** the station. Which other modes were available to you when making this trip?

- ☐ Non E-Hail Taxi
- ☐ Walking
- ☐ E-Hail (Uber/Bolt)
- ☐ Being Dropped off
- ☐ None of the above
- ☐ Private Car

* 20. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your starting Gautrain station? (Select one of the following).

[ ]

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Access Mode: E-Hail (Uber / Bolt)

21. You said you took an **E-Hail ride** to the station. Which other modes were available to you when making this trip?

- [ ] Non E-Hail Taxi
- [ ] Walking
- [ ] Gautrain Bus
- [ ] Being Dropped off
- [ ] None of the above
- [ ] Private Car

* 22. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your starting Gautrain station? (Select one of the following).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Car/Gautrain Bus Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Trip Attribute	Description
Bus walk time to stop (min)	This is time you spend walking to the bus stop from your trip origin.
Safety level on walk to bus stop	This describes the level of personal safety you experience during your walk to the bus stop. **Somewhat Secure** means a walking environment with not many people around, no visible security presence and poor lighting at night. **Very Secure** means an environment with a lot of people around, very good lighting at night and a visible security presence.
Waiting time at Gautrain bus stop (min)	This is the time you spend waiting at the Gautrain bus stop for the bus to arrive.
Travel time car & bus (min)	This is the travel time in your car or the Gautrain bus to get to the Gautrain station.
Bus fare (Rand)	This is the fare you pay for your Gautrain bus trip to the Gautrain station.
Car petrol cost to station (Rand)	This is your car's petrol cost to get to the Gautrain station
Car parking fee (Rand)	This is the parking charge for leaving your car parked at the Gautrain station park and ride.

Given different combinations of the following attributes please select the option you would make in the following situation.

* 23.
The choice sets have been split into three equals blocks of 12 choice sets each- please **randomly** select a block to answer.

- ◯ Block One
- ◯ Block Two
- ◯ Block Three

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Car or Gautrain Bus travel to Gautrain station: B3 Scn 1/12

* 24. Given these conditions, would you prefer taking your own car or the Gautrain bus to your starting station?

Choice Set 1 Block 3	Car	Gautrain Bus
Bus walk time to stop (min)	-	5
Safety level on walk to bus stop	-	Somewhat Secure
Bus wait time at stop (min)	-	5
Travel time car & bus (min)	20	20
Bus fare (Rand)	-	6
Car petrol cost to station (Rand)	6	-
Gautrain park & ride fee (Rand)	18	-

- I would prefer taking my **car** to the starting station.
- I would prefer taking the **Gautrain Bus** to the starting station

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Drop off / Gautrain Bus Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walking time to bus stop (min)	This is the time spent walking to the bus stop
Waiting time at bus at stop (min)	This is the wait time for a bus
Travel time drop off & bus (min)	This is the time spent in the bus or in your lift to reach the station
Drop off & bus cost (Rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip. On the bus column this is the fee you are charged for the bus ride
Safety level on walk to bus stop	This describes the level of safety you experience during your walk to the bus stop . **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Drop-Off or Gautrain Bus travel to Gautrain Station: Scenario 1/8

* 60. Would you prefer being dropped off or taking the Gautrain bus to the station?

	Drop off	Gautrain Bus
Walking time to bus stop (min)	-	10
Waiting time at bus stop (min)	-	10
Travel time drop off & bus (min)	15	15
Drop off & bus cost (Rand)	30	6
Safety level on walk to bus stop	-	Somewhat Secure

- ◯ I prefer to be **Dropped Off** at the station.
- ◯ I prefer to take the **Gautrain Bus to the station.**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Walk/Gautrain bus access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walk time to station (min)	This is the time required to walk to the station
Level of safety during walk	This describes the level of safety you experience during your walk to the station. **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Walk time to bus stop (min)	This is the time required to walk to the bus stop.
Wait time at stop (min)	This is time spent waiting for a bus at the stop.
Travel time in bus (mini)	This is the time spent in the bus
Bus Fare (Rand)	This is the fee for your trip on the bus

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Walk or Gautrain bus travel to Gautrain station: Scenario 1/8

* 68.

Would you prefer walking, or taking the Gautrain bus to the Gautrain Station?

	Walk	Bus
Walk time to station (min)	15	-
Level of safety during walk	Very Secure	Very Secure
Walk time to bus stop (min)	-	10
Wait time at stop (min)	-	5
Travel time in bus (mini)	-	5
Bus Fare (Rand)	-	6

- I prefer to **walk** to the Gautrain station.
- I prefer to take the **Gautrain Bus** to the Gautrain station**.**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- E-Hail (Uber/Bolt)/Gautrain Bus Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Trip Attribute	Description
Bus walk time (min)	This is the walk time to the Gautrain bus stop
Bus walk safety level	This describes the level of safety you experience during your walk to the bus stop . **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Bus wait time (min)	This is the waiting time for the Gautrain bus at the Gautrain bus stop.
Bus travel time (min)	This is travel time in the Gautrain bus to get to the Gautrain station.
Bus fare (rand)	This is the fare you pay for your Gautrain bus trip to the Gautrain station.
E-Hail (Uber/Bolt) travel time (min)	This is the travel time in the E-Hail car to reach the station.
E-Hail (Uber/Bolt) fare (rand)	This is the fare you pay for your E-Hail Car.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

E-Hail (Uber/Bolt) or Gautrain bus to travel to Gautrain Station: Scenario 1/8

* 76.
Would you prefer to E-Hail or take the Gautrain bus to get to the Gautrain Station?

	Bus	E-Hail (Uber/Bolt)
Bus walk time (min)	5	_
Bus walk saftey level	Very Secure	_
Bus wait time (min)	10	_
Bus Travel time (min)	25	_
Bus fare (rand)	6	_
E-Hail (Uber/Bolt) travel time (min)	_	15
E-Hail (Uber/Bolt) fare (rand)	_	20

- ◯ I'd prefer to take a **Gautrain Bus** to the Gautrain Station.
- ◯ I'd prefer to **E-Hail** to the Gautrain Station.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Private Car/ E-Hail (Uber/Bolt) Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Travel time (min)	This is the travel time in your car or the E-Hail vehicle to reach the Gautrain station.
Car travel cost (Rand)	This is the cost of the petrol for your car to reach the Gautrain station.
Car parking cost (Rand)	This is the cost of parking your car at the Gautrain station park and ride.
E-hail fare (Rand)	This is the fare for your E-Hail trip to the Gautrain station.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Private Car or E-Hail (Uber/Bolt) travel to Gautrain station: Scenario 1/8

* 84. Would you prefer to take your car or E-Hail for your trip to the station?

	Car	E-Hail (Uber/Bolt)
Travel time (min)	25	15
Car travel cost (Rand)	15	-
Car parking cost (Rand)	18	-
E-hail fare (Rand)	-	20

- ◯ I prefer to take **my own private car** to the Gautrain station.
- ◯ I prefer to take an **uber** to the station**.**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Drop off/E-Hail (Uber/Bolt) Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Travel time (min)	This is the time spent in your lift or E-Hail vehicle to reach the station.
Drop Off cost (rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip.
E-Hail (Uber/Bolt) fare (rand)	This is the fee you are charged for your E-Hail ride

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Drop Off or E-Hail (Uber/Bolt) travel to Gautrain station: Scenario 1/8

* 92.
Would you prefer to E-hail, or be dropped off to get to your starting station?

	Drop Off	E-Hail (Uber/Bolt)
Travel time (min)	15	15
Drop Off cost (rand)	15	-
E-Hail (Uber/Bolt) fare (rand)	-	20

- ◯ I prefer **being dropped off at** the Gautrain station.
- ◯ I prefer **to E-Hail** to the Gautrian station.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Walk/Drop off Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walking time (min)	This is the time spent walking to the station
Level of safety during walk	This describes the level of safety you experience during your walk to the station . **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Car travel time (min)	This is the time your lift takes to bring you to the station.
Car travel cost (Rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Walk or Drop Off travel to Gautrain station: Scenario 1/8

* 100.
Would you prefer walking or being dropped off to get to the Gautrain Station?

	Walk	Drop Off
Walking time (min)	15	-
Level of safety during walk	Somewhat Secure	-
Car travel time (min)	-	5
Car travel cost (Rand)	-	15

- ◯ I'd prefer **walking** to the Gautrain station.
- ◯ I'd prefer to be **dropped off** at the Gautrain Station.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Private Car/Drop Off Access

We now want you to make several imaginary choices on how to make your journey to the station. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Travel time (min)	This is the time spent in your car or your lift to reach the station.
Trip Cost (Rand)	This is the cost for the fuel for your car. In the Drop off column it is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip.
Parking fee (Rand)	This is the cost of parking your car at the station.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Private Car or Drop Off travel to Gautrain Station: Scenario 1/8

* 108.
Would you prefer taking your car or being dropped off to get to your Gautrain station?

	Drop-off	Car
Travel time (min)	20	15
Trip Cost (Rand)	30	6
Parking fee (Rand)		18

- ◯ I'd prefer **taking my own car** to the Gautrain Station.
- ◯ I'd prefer to be **dropped off** at the Gautrain Station.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Egress Mode

Egress mode refers to how you traveled from your final station to get to your destination (work, school etc.).

* 116. How did you **reach your destination** for this trip **from your final station?**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Egress Mode- E-Hail(Uber/Bolt)

117. You said you **E-Hailed** to your destination. Which of the following modes were available to you for your trip to your destination?

- [ ] Gautrain Bus
- [ ] Being picked up by a colleague
- [ ] Walking
- [ ] Normal (non-Ehail) Taxi
- [ ] None of the above

* 118. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your final destination? (Select one of the following).

- [ ] Gautrain Bus
- [ ] Be picked up by a friend/colleague

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction: E-Hail(Uber/Bolt)/Gautrain Bus Egress

We now want you to make several imaginary choices on how to make the journey to your destination. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

	Description
Bus walk time (min)	This is the walk time to your destination
Bus walk saftey level	This describes the level of safety you experience during your walk to your destination. **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Bus wait time (min)	This is the wait time for a bus at the station.
Bus Travel time (min)	This is time spent in the bus to get to reach your destination.
Bus fare (rand)	This is the fee you pay for your bus trip to your destination.
E-Hail (Uber/Bolt) travel time (min)	This is the time spent in the E-Hail car to reach your destination
E-Hail (Uber/Bolt) fare (rand)	This is the fare you pay for your E-Hail Car.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

E-Hail (Uber/Bolt) or Gautrain bus to reach final destination: Scenario 1/8

* 119.
Would you prefer to E-Hail or take the Gautrain bus to get to your final destination?

	Bus	E-Hail (Uber/Bolt)
Bus walk time (min)	5	_
Bus walk saftey level	Very Secure	_
Bus wait time (min)	10	_
Bus Travel time (min)	25	_
Bus fare (rand)	6	_
E-Hail (Uber/Bolt) travel time (min)	_	15
E-Hail (Uber/Bolt) fare (rand)	_	20

- ○ I'd prefer to take a **Gautrain Bus** to my final destination.
- ○ I'd prefer to **E-Hail** to my final destination.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Egress Mode- Walk

127. You said you **walked** to get to your final destination from the Gautrain station for this trip.

Which of the following modes were available to you as an alternative for your trip to your destination?

- [ ] Gautrain Bus
- [ ] Being picked up by a colleague
- [ ] E-Hail (Uber/Bolt)
- [ ] Normal (non-Ehail) Taxi
- [ ] None of the above

* 128. Assuming the following modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your final destination? (Select one of the following).

- [ ] Gautrain Bus
- [ ] Be Picked Up

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Walk/ Pick Up Egress

We now want you to make several imaginary choices on how to make the journey to your destination. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walking time (min)	This is the time spent walking to your destination
Level of safety during walk	This describes the level of safety you experience during your walk to your destination. **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Car travel time (min)	This is the time your lift takes to bring you to your destination.
Car travel cost (Rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Walk or Picked up to travel to destination: Scenario 1/8

* 129.
Would you prefer to walk or be picked up to reach your final destination?

	Walk	Pick Up
Walking time (min)	15	-
Level of safety during walk	Somewhat Secure	-
Car travel time (min)	-	5
Car travel cost (Rand)	-	15

- I'd prefer **walking** to the my final destination.
- I'd prefer to be **picked up** at the Gautrain Station and dropped off at my final destination.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Walk/Gautrain Bus Egress

We now want you to make several imaginary choices on how to make the journey to your destination. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walk time to destination (min)	This is the time required to walk to your destination from the station.
Level of safety during walk	This describes the level of safety you experience during your walk to your destination. **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.
Walk time to destination from stop (min)	This is the walk time to the destination from the bus stop.
Wait time at station (min)	This is the time spent waiting for a bus at the station.
Travel time in bus (mini)	This is time spent in the bus to get to reach your destination.
Bus Fare (Rand)	This is the fee for your trip on the bus

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Walk or Gautrain Bus to travel to destination: Scenario 1/8

* 137.

Would you prefer to take the Gautrain Bus or walk to your final destination?

	Walk	Bus
Walk time to destination (min)	15	-
Level of safety during walk	Very Secure	Very Secure
Walk time from bus stop (min)	-	10
Wait time at station (min)	-	5
Travel time in bus (mini)	-	5
Bus Fare (Rand)	-	6

- I prefer to **walk** to my final destination.
- I prefer to take the **Gautrain Bus** to my final destination.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Egress Mode- Pick Up

145. You said you were **picked up** from the station to get to your final destination for this trip.

Which of the following modes were available to you as an alternative for your trip to your destination?

- [ ] Gautrain Bus
- [ ] E-Hail (Uber/Bolt)
- [ ] Walking
- [ ] Normal (non-Ehail) Taxi
- [ ] None of the above

* 146. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your final destination? (Select one of the following).

- ( ) Gautrain Bus
- ( ) E-Hail(Uber/Bolt)
- ( ) None of the above

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction: Pickup/E-Hail (Uber/Bolt) Egress

We now want you to make several imaginary choices on how to make the journey to your destination. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Travel time (min)	This is the time spent in your lift or E-Hail vehicle to reach your destination.
Pick Up cost (rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip.
E-Hail (Uber/Bolt) fare (rand)	This is the fee you are charged for your E-Hail ride

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Picked up or E-Hail (Uber/Bolt) to travel to destination: Scenario 1/8

* 147.

Would you prefer to E-Hail or be picked up to get to your final destination?

	Pick Up	E-Hail (Uber/Bolt)
Travel time (min)	15	15
Pick Up cost (rand)	15	-
E-Hail (Uber/Bolt) fare (rand)	-	20

- I prefer **being picked up** at the station by a friend/family/colleague and then dropped off at my final location.
- I prefer to **E-Hail** to my final destination.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Introduction- Gautrain Bus/Pick up Egress

We now want you to make several imaginary choices on how to make the journey to your destination. The choices are between the mode you currently use to get to the station, and the mode you said you would most likely switch to. Each mode of transport is described in terms of the following attributes.

Attribute	Description
Walking time to bus stop (min)	This is the time spent walking to your destination from the bus stop
Waiting time at bus at stop (min)	This is the wait time for a bus
Travel time pick up & bus (min)	This is the time spent in the bus or in your lift to reach your destination
Pick Up & bus cost (Rand)	This is the amount of money the person giving you a lift asks you to pay to help cover the cost of the trip. On the bus column this is the fee you are charged for the bus ride
Safety level on walk to destination	This describes the level of safety you experience during your walk to the bus stop . **Somewhat Secure** means an environment with not many people around, no visible security presence and poor lighting. **Very Secure** means an environment with a lot of people around, very good lighting and a visible security presence.

Given different combinations of the following attributes please select the option you would make in the following situation.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Gautrain Bus or Picked up to travel to destination: Scenario 1/8

* 155. Would you prefer to take a Gautrain Bus or be picked up to reach your final destination?

	Pick Up	Gautrain Bus
Walking time to bus stop (min)	-	10
Waiting time at bus stop (min)	-	10
Travel time Pick Up & bus (min)	15	15
Pick Up & bus cost (Rand)	30	6
Safety level on walk to bus stop	-	Somewhat Secure

- ◯ I prefer to be **picked up** by a friend/family/colleague at the station and dropped off at my final destination.
- ◯ I prefer to take the **Gautrain Bus** to a bus stop closest to my final destination.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Current Egress Mode- Gautrain Bus

163. You said you took the **Gautrain Bus** to get to your final destination for this trip.

Which of the following modes were available to you for your trip to your destination?

- [ ] E-Hail (Uber/Bolt)
- [ ] Being picked up by a colleague
- [ ] Walking
- [ ] Normal (non-Ehail) Taxi
- [ ] None of the above

* 164. Assuming these modes were available to you, which mode would you most likely switch to if you had to change mode, in order to reach your final destination? (Select one of the following).

- [ ] Be Picked Up
- [ ] Walk
- [ ] E-Hail(Uber/Bolt)
- [ ] None of the above

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

Social Demographics

Please complete the following social demographic questions before completing the online Gautrain Internet Survey.

* 165. Please indicate your **gender:**

- Female
- Male
- Prefer not to say
- Other (please specify)

* 166. Please indicate your **age** range:

- < 18
- 18 - 24
- 25-34
- 35-44
- 45 - 54
- >55
- Prefer not to say

* 167. What range does your **gross (before tax) monthly income** fall into? Please note this is only used to benchmark against other South African transport systems.

- < R10 000
- R10 000 - R17 000
- R17 000 - R30 000
- >R30 000
- Prefer not to say

* 168. What is your **employment** status?

- Home maker
- Full-time employment
- Part-time employment
- Unemployed
- Other (please specify)
- Scholar
- Retired
- Prefer not to say

* 169. How many **cars** are in your house hold?

- 0
- 1
- 2
- 3+

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Centre for Transport Development

IN ASSOCIATION WITH
GAUTRAIN MANAGEMENT AGENCY

End of Survey

170. Thank you for participating in this survey. In order to win a R50 takealot voucher, please provide us with an email address we can forward the coupon to. Please note you are free to not provide an email address. Also, in order to verify that there are no repeat takers of the questionnaire please allow up to 48 hours for delivery of the voucher.