MODELLING PUBLIC TRANSPORT MODE CHOICE FOR LOW-INCOME RESIDENTIAL SUBURBS IN HARARE, ZIMBABWE.

Masimba Tutsirai Mapfurira

February 2022

A thesis submitted as fulfilment of the requirements for the degree of

Master of Engineering in Transport Studies

in the

Faculty of Engineering and the Built Environment

of the University of Cape town

Declaration

I, Masimba Tutsirai Mapfurira declare that the findings, interpretations, and conclusions expressed in this study are essentially my work or otherwise referenced.

Date: 14 February 2022

Signature: Signed by candidate

Dedications

I dedicate this work to my mom, Mrs Ndaiziveyi Mapfurira, who has supported my dreams and every step towards achieving them.

Acknowledgements

Firstly, I would like to highlight the limitation of language in expressing my gratitude towards the Lord Jesus Christ for His mighty hand and remarkable providence towards the commencement and completion of my master's programme.

To my supervisors, Professor Mark Zuidgeest and Dr Hazvinei Tsitsi Tamuka Moyo, I would like to express my sincere and humble admiration for your diligent effort and guidance towards the success of this thesis. Thank you for helping me structure my ideas well and providing constructive feedback on my model estimations. The knowledge gained during this study, not found in written records but now under my custody courtesy of the remarkable guidance from my supervisors. A special appreciation goes to the apollo forum team, Professor Stephane Hess and Dr David Palma, for their expert advice and patient responses to my seemingly trivial questions. I would like to acknowledge the survey team from Harare, specifically Tino and Tadiwanashe, who have been incredibly generous with their time and energy in helping me to coordinate all the resources that we used for data collection. My gratitude similarly extends to my young brother Flavian and colleagues from One Pangaea Financial for their support in making this journey a success. Also not forgetting, my family and friends for their assistance and moral support throughout the research, may God abundantly bless you. I am forever indebted for the unprecedented support from my mother towards the studies. Thank you, mom, for insisting on paying for my flight to Cape Town.

Abstract

Modern-day life in developing urban cities is structured around the need to access goods and services outside the vicinity of residential areas, and transportation facilitates access to such services. Like many other African cities, most of the residents in Harare sorely rely on public transport, and while the government of Zimbabwe acknowledges the need for efficient public transportation systems in the country's urban environments, insufficient commitment and political will have been directed towards developing strategic plans with clear and well-defined objectives.

The development of public transport plans and policies requires a good understanding of the passengers' service quality needs and willingness to pay for service quality improvements. In this study, we use stated choice preference data collected from five high-density suburbs in Harare (n = 361) to investigate the influence of service quality indicators to travel mode choice decisions. Multinomial, mixed multinomial and latent class logit models are developed under the random utility maximisation framework and compared to identify the best model. The model is used to evaluate the willingness to pay indicators for public transport service improvements and outline the contributions of the findings to possible policy directives.

The results suggest that latent class models perform substantially better in explaining observed choices than both mixed and multinomial logit counterparts. With regards to public transport mode choice behaviour, the study classifies the population into two distinct groups on the basis of gender, income, employment status, and location. The willingness to pay indicators shows a substantial difference in the value of all the public transport attributes between the groups, except for waiting time. The willingness to pay for improvements in waiting time, which relates to service frequency, is standard at Z$65 per hour. Noteworthy is the classification of the suburbs between the groups; the posterior analysis indicates that Chitungwiza residents have the highest willingness to pay and Budiriro, the least.

This research is of value to ZUPCO and other potential private players in identifying service quality deficiencies and understanding the requirements of public transport service provision at the suburban level. The strong inertia towards kombis emphasises the general dissatisfaction with ZUPCO service quality levels while providing insights

into lagging areas that future policy deliberations could address. The research presents a potential performance framework to the Harare city council against which the public transport service provision can be assessed. Most importantly, the findings might be useful in further understanding the public transport landscape in other cities in Zimbabwe, similar to the high-density suburbs used as study areas in this research.

Table of contents

Lists of Tables

Lists of Figures

List of Appendices

Acronyms and abbreviations

PT	Public transport
MNL	Multinomial logit
IIA	Independence of Irrelevant Alternatives
GEV	Generalised Extreme Value
MMNL	Mixed multinomial logit
LC	Latent Class
LC-MNL	Latent Class with multinomial logit
LC-MMNL	Latent Class with a mixed multinomial logit
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
SP	Stated Preference
ESAP	Economic Structural Adjustment Programme
ZUPCO	Zimbabwe United Passenger Company
GTFS	General Transit Feed Specification
COJ	City of Johannesburg
CBD	Central Business District
CPI	Consumer Price Index

NB: 1USD = $Z160 (Zimbabwean dollar at the time of survey)

1.0 INTRODUCTION

Apart from population growth, urbanisation has been the leading demographic trend since the late 20th century, as the growing share of the global population now lives in cities (Brown & Jacobson, 1987). Most developing countries have experienced annual urban population growths of more than six per cent for the past decades (World Bank. & Gwilliam, 2002), while forecasts indicate an increase in the world urban population to 68% by 2050 (Farvacque-Vitkovic & Kopanyi, 2019). As the urban population grows, cities expand horizontally, creating a demand for infrastructure and transport services, among other things. However, most cities in developing countries face financial and institutional challenges in providing these services to cope with the demand (Mitric, 2008; Amiegbehor et al., 2010).

Job opportunities and economic activities remain concentrated in the urban core while new settlements develop at the urban periphery. Such urban forms require efficient transportation systems to overcome the geographical separation between work opportunities and residential locations. The combination of poor public transport and low household private vehicle ownership in African countries highlights the extent of economic exclusion, which is highly dependent on the quality and network coverage of the public transport system (Amiegbehor et al., 2010; Porter, Abane & Lucas, 2020).

The public transport landscape in developing countries is dominated by bus rapid transit, metro rail, conventional bus service, and paratransit. Bus rapid transit and metro rail are operated formally on fixed routes and schedules. While conventional bus services operate formally on selected routes, their schedules are semi-fixed and contingent upon demand and network flow. The paratransit or informal public transport service, which commands the largest market share, is characterised by different types of vehicles: midi-buses, minibus taxis, microbus, motorcycles, and horse carts. The service is privately owned and operates illicitly on semi-fixed routes with flexible schedules, providing the much-needed level of service, especially for the marginalised population living in informal settlements (Cevero, 2000; Behrens, McCormick & Mfinanga, 2015).

On the other hand, formal public transport services receive disproportionately high subsidies from central governments, enabling them to charge lower fares than paratransit on the pretext of redressing social inequalities against the low-income population. However, the peak period unidirectional nature of travel demand in these countries undermines the financial viability of providing competitive services across the metropolitans. Paratransit competes for passengers with the standard service on parallel routes, offering a higher quality of connectivity at premium fares, limiting low-income households from accessing essential social activities. The competing public transport service providers offer different quality of service to fare price ratio packages, and their market share depends on their relative ratio packages (Vuchic, 2005) to competing counterparts.

Kittelson & Associates et al (2013), frame the role of public transport as to provide essential mobility to public transport dependent users and attract choice users who might have more travel options available to them; this means different public transport operators render services to meet the expectations of their intended "target market". Passengers choose between available alternative public transport modes, conditioned by attributes of different alternatives, and their choices reflect on their needs and preferences (Litman, 2013; Muro-Rodríguez, Perez-Jiménez & Gutiérrez-Broncano, 2017). When developing a public transport development plan, it is therefore essential to assess and understand the passenger's service quality needs and preferences. Quality of service is a subjective metric that reflects a passenger's perception of public transport performance, and the indicator is intuitively measured according to the level of service rendered, meeting the needs and expectations of the consumer. When making choices, the literature suggests that passengers place a value on different service attributes and subconsciously evaluate the generalised cost of the trip (Kittelson & Associates. et al., 2013; Litman, 2013). The choice decisions are thus subject to mode attributes, and discrete choice models have been developed to explain and predict choice behaviour. Moreover, the models have proven effective in traffic network assignment and assessing travel demand management strategies. This study seeks to determine the factors that influence public transport mode choice decisions in the context of low-income residential suburbs in Harare and examine their contribution to mode choice decisions.

1.1 Harare Overview

Zimbabwe has an urban population of 5.3 million projected to grow by 2.0% annually to over 6.2 million by 2030 (ZimStat, 2015). Like in many developing countries, the population growth in Harare has led to the expansion of the city spatial structure as characterised by infills, extensions, leapfrog developments, and uncontrolled urban sprawl (Kamusoko, Gamba & Murakami, 2013). The capital city inherited its urban layout from the colonial planning policies, which focused on the economic segregation of neighbourhoods along racial lines. The urban landscape bears classist fingerprints of the inherited urban design, where the opulent occupies well-developed low-density areas located close to economic hubs, and the urban poor in high-density minority neighbourhoods are situated in the urban fringe. The urban sprawl is partly driven by politically connected land barons who allocate land for political expediency disregarding the city guidelines (Matamanda et al., 2020). The continual proliferation of informal settlements at the urban fringe further undermines the provision of critical services, among which is transport, thus exacerbating spatial dislocation of the majority from the available opportunities in the economic nodes. The long distances travelled, as a result, highlight flaws in the inherited urban policies and planning practices.

For the past decade, Zimbabwe experienced a mass importation of relatively cheap second-hand cars from Japan, most destined for Harare (Mbara, 2015); this has increased pressure on the already struggling urban road network. Consequently, the level of service for transportation facilities and traffic networks has deteriorated to gridlock congestion levels during peak periods, thereby imposing social, environmental, and economic costs to businesses and residents. Most sub-Saharan cities face the same mobility deficiencies due to an unstable macroeconomic environment and a lack of transportation planning strategies to satisfy their travel needs (African Centre for Cities, 2015; Ndibatya & Booysen, 2020).

1.2 State of public transport services

Jordan (1983) and Maunder & Mbara (1995) well document the administrative history and development of public transport in Harare. Before the attainment of independence in 1980, a stage carriage service was offered exclusively by Salisbury United Omnibus Council (SUOC) under a franchise agreement with the local authority in the form of devolution. In Harare, the bus operator was responsible for bus shelters and depots, while the city council provided the locations for bus termini. This arrangement was however different from other cities like Bulawayo, Gweru and Mutare, where the city council provided supporting infrastructure to the private operators. As part of the franchise agreement, the operator would make a reasonable profit measured in return on investment calculated by an agreed formula, below which a subsidy was payable by the local authority. The low-density suburbs, nonetheless, did not enjoy these subsidies, for a "first-class" service was provided instead, outside the franchise agreement with slightly higher fixed fares.

Following independence in April 1980, the newly elected government embarked on a policy redressal program which aimed at addressing the socio-economic developmental inequalities which existed in the colonial era, and emphasis was placed on controlling major critical sectors of the economy in which urban transport was one of them (Maunder & Mbara, 1995). A decision was reached to take over the fare determination mandate from the Harare city council and revoke all operation subsidies, but with an agreement to establish and review new economic fares annually. According to Mbara, Dumba & Mukwashi (2014), the fare establishment arrangement did not work out well, and subsidies were agreed upon instead. In addition, the government temporarily legalised an informal shared taxi service, characterised by seven-seater station wagon Peugeot vehicles, to operate on pre-set routes (Maunder & Mbara, 1995). The central government later acquired a 51% shareholding in the conventional bus service, which became Zimbabwe United Passenger Company (ZUPCO) after independence.

After ten years of independence, the government introduced an economic structural adjustment programme (ESAP) that aimed to liberalise the economy to facilitate competition and enhance productivity (Mlambo, 1997). The deregulation of public transport in the early 90s, as a result, initiated an exponential growth of privately owned and informally operated minibus vehicles, locally known as commuter omnibus or kombis (Mbara, 2006). The development increased the supply of passenger transport services and expanded the public transport network substantially, connecting economically marginalised areas to economic hubs. Accessibility in the form of walking distance and waiting time improved, which resonated with passengers' service quality needs, especially for people who previously had to walk long distances and endure long queues to access the government-owned conventional service (Maunder & Mbara, 1995; Chikozho Mazarire & Swart, 2014). Due to operational challenges, the conventional bus service eventually collapsed, leaving the informal private players in complete control of the entire urban public transport sector (Bryceson et al., 2003). In the city centre, a fleet of small ex-Japanese cars infamously known as "mushika-shika"(*which translates to go faster*), offer an illegal shuttle service to get around town. The vehicles park haphazardly to pick up passengers at undesignated spots, causing congestion, and they drive recklessly with less regard for passengers, other motorists, and pedestrians (Charamba, 2016).

In the year 2019, the federal government reintroduced ZUPCO to provide a relatively cheap public transport service to civil servants and dependent users who could not afford kombi fares due to hyperinflation. With the subsidy put in place, kombi operators faced a price competition challenge as ZUPCO charged almost half their trip fares. Due to the resultant shift towards the cheaper mode, kombis lost more than half of their ridership regardless of relatively better service quality (Ndlovu, 2020). The global outbreak of the COVID-19 pandemic came along with so many changes in socio-economic activities. To curtail the spread of the pandemic, many countries-imposed travel restrictions and lockdown measures that financially crippled the public transport industry as people were required and encouraged to stay home.

The pandemic, nonetheless, came as a perfect opportunity for the government of Zimbabwe to get rid of the informal public transport service in a quest to formalise the urban transport sector. A Statutory Instrument (SI) 99 of 2020: Covid-19 Prevention, Containment and Treatment was enacted under the National Lockdown Amendment Order No.5, 2020 giving ZUPCO exclusive rights to provide public transport services in urban spaces (Zimcodd, 2020). In Harare alone, more than 12000 minibus taxis were barred from providing transport services, and they could only operate under the ZUPCO franchise agreement. The financing structure remains blurry as there is no coherent policy with clear objectives to support urban mobility services. To date, ZUPCO has managed to acquire 770 conventional buses and signed 1000 minibus taxis nationwide to operate under its franchise scheme (Mutongwiza, 2021; Zimbabwe United People's Company, 2020). Despite all this effort, the service provider remains incapacitated to meet daily travel demand in Harare alone, as the service is characterised by unbearable long queues, waiting, and commuting times. Consequently, a growing number of private car owners following the subsequent ban of kombis have seen the capacity problem as an opportunity to ferry the helpless passengers, who sometimes walk long distances, to and from work at premium fares.

1.3 Research problem

The primary role of public transportation is to compensate for disproportional imbalances in the availability of urban amenities and economic opportunities across the metropolis by facilitating sustainable mobility for transit captive passengers and choice riders. The National Transport Policy (NTP) of Zimbabwe rightfully acknowledges the impact of growing automobile use on existing capacity and environment and further outlines the need for an efficient public transport system in urban environments. However, little commitment has been directed towards developing strategic plans with clear and well-defined objectives (Ministry of Transport Communications and Infrastructural Development, 2011). The Ministry of Transport and Infrastructural Development focuses on national and regional connectivity instead, channelling all the infrastructure budget towards building national trunk, provincial, rural, and urban roads. Though it is important for landlocked countries like Zimbabwe to build resilient national road networks, they are not sufficient to address all the transportation needs for the growing urban population.

Policies and projects that affect the public transport users are often crafted and implemented without consultations with all stakeholders; for instance, the unexpected restriction of kombis from dropping off and picking up passengers in the central business district in February 2018, which aimed at decongesting the city. The kombis had to drop off passengers at designated spots, more than two kilometres walking distance outside the city centre; then, the local authority provide a shuttle service to complete the first-last mile service to the city centre at an additional cost. The ministry of local governance reversed the decision a day later as no shuttle service was availed to provide the last mile service (Mbara, 2015). The arrangement was extremely painful for connecting passengers, who ended up walking more than four kilometres across the central business district to access their second kombi. The general victimisation of the informal public transport sector, which is responsible for connecting the low-income and marginalised population to socio-economic zones, highlights the lack of the much-needed skillset within the local authority structures to address urban transport challenges (Mbara & Pisa, 2019).

The car-sharing service that emerged following the ban of kombis during the lockdown period cannot fill the supply vacuum that the statutory instrument created. In areas where kombis still operate illegally or under the ZUPCO franchise scheme, passengers have to choose from different service providers. Notwithstanding the ever-increasing demand for public transport services due to covid-19 travel restriction relaxations, the central government insists on discouraging the use of any form of public transport besides ZUPCO. The use of kombis and car-sharing services in town remains prevalent, especially during peak periods, despite being a punishable offence attracting a spot fine of Z$2000 (*Equivalent to US*$12.50 *at the time of study*).

The urban transport landscape in Harare needs policy reforms and integrated public transport plans, and the development thereof, requires a solid understanding of passenger service quality needs, the influential factors towards travel mode choices, and the passenger willingness to pay for service quality improvements. Surprisingly, little effort has been directed towards identifying the service quality indicators and forecasting passenger mode choice behaviour which is essential when drafting policies and travel demand management strategies.

1.4 Main research objective

The main objective is to develop a public transport mode choice model for high-density/low-income residential suburbs in Harare, Zimbabwe, and then examine the contribution of mode attributes to mode choice using a discrete choice experiment. The outcome of this research will be valuable to ZUPCO as the mandated public transport service provider in identifying service quality gaps in their provision. This study may add value by providing policymakers and private investors with guidance to facilitate a modal shift towards more sustainable mobility. In addition, the knowledge may, in turn, be used in traffic assignment when developing four-step trip generation models.

1.4.1 Specific research objectives and research questions

1. **To identify the key service quality indicators that influence mode choice behaviour in the context of the low-income population.**
    a. Which service quality indicators influence passenger satisfaction, and can the relevant indicators be identified from literature review and interviews?
    b. How can the service quality indicators be expressed as physical and metric attributes?
2. **To design and conduct a stated preference survey to elicit public transport mode choice preferences.**
    a. How can choice behaviour be explained by observing responses to hypothetical choice scenarios?
    b. How can a questionnaire be constructed to mimic the real-world mode choice structure?
    c. Which sampling strategy best represents the population of public transport dependent users in Harare?
3. **To estimate different choice models from the data to understand the mode choice behaviour.**
    a. Can generic parameters for attributes be estimated from observed mode choice data?
    b. How do the parameters vary across people across the population?
4. **To determine the contribution of public transport service quality indicators to observed choice behaviour.**
    a. How can the parameter estimates be interpreted in terms of sensitivity to change in variables?
    b. How substantial is the contribution of the variables towards the observed choice behaviour?
5. **To evaluate the willingness to pay for improvements in public transport service level.**

   a. How can the interpretation of the results be applied to transport policies?
   b. How much are the respondents willing to pay for improving the service quality?
   c. How does willingness to pay vary across the population?

6. **To evaluate transport policies and propose recommendations to improve the level of service and ridership.**
   a. What conclusions can be drawn from the study?

1.5 Thesis outline

The rest of the research is divided into four chapters, and the flow is organised as follows:

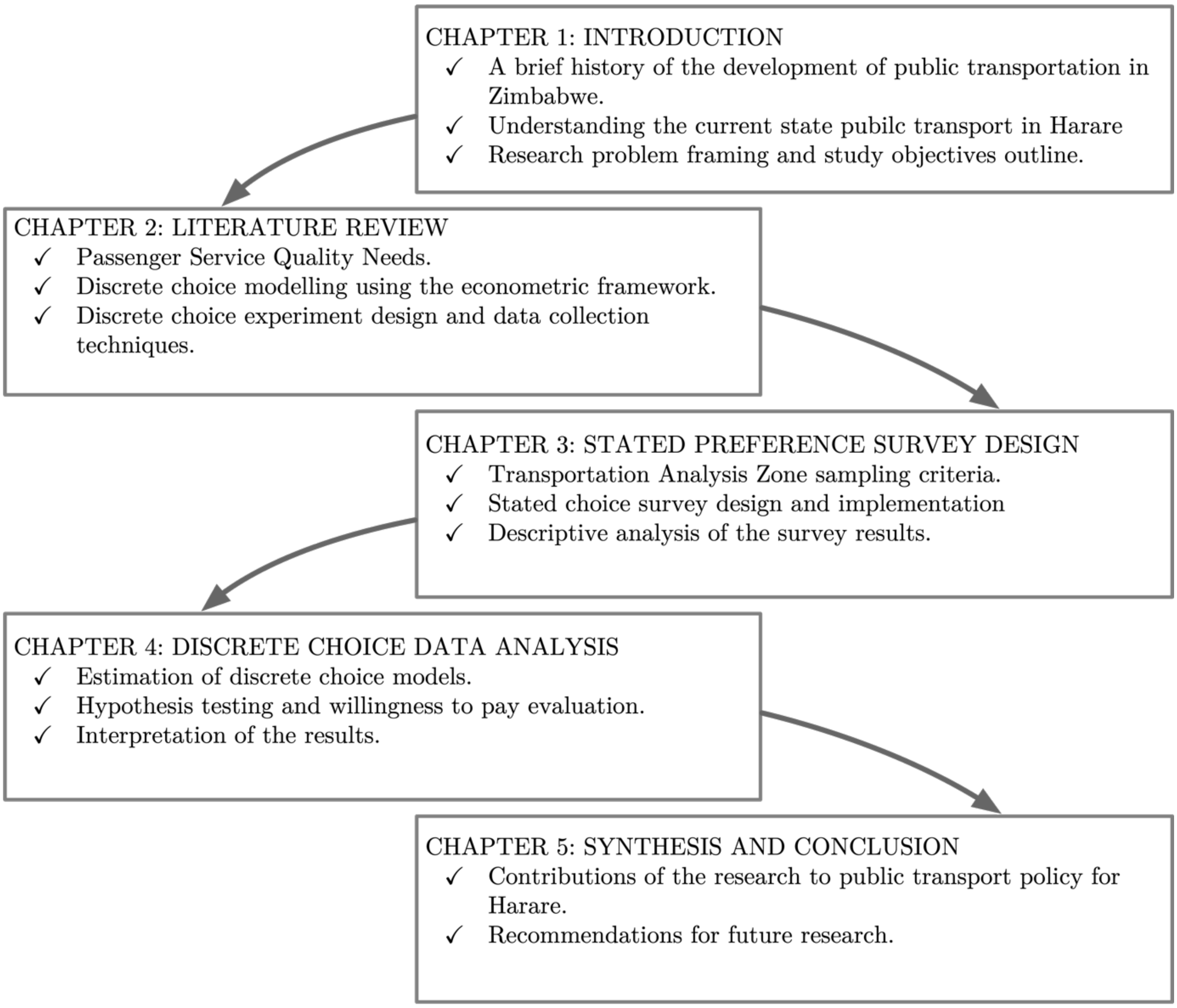


Figure 1: Research flow chart

Chapter 2.0 reviews both empirical and theoretical literature on passenger service quality indicators, the econometric choice modelling framework, and discrete choice experiment design. The primary goal is to identify the key factors that influence mode choice decisions and establish how their contribution to mode choice behaviour can be assessed. Chapter 3.0 describes how this study was conducted in terms of the discrete choice experiment design and survey administration, sampling procedures, and data collection. The chapter also covers how the attributes and levels for each mode were determined for each sampled suburb. The section concludes by giving a detailed statistical descriptive analysis of revealed preference data collected from the survey. Chapter 4.0 focuses primarily on estimating the discrete mode choice models and testing the statistical significance of the influence of both sociodemographic and mode-specific variables on the observed choices. The models are interpreted in terms of behavioural meaning and the willingness to pay for service improvement. Finally, chapter 5.0 concludes the thesis with a summary of the research methodology, data analysis, and results. The final chapter also indicates the implementation of the results and recommendations for future research directions.

2.0 LITERATURE REVIEW

People are constantly faced with choices in health, commerce, education, accommodation, transport, and many other sectors. Some options like grocery brands and commuting modes are repetitive, while choices about buying a house or attending a university are made occasionally. Repetitive choices usually develop into habits, which often change with various life-changing events or disappointment from poor service provision. In the context of urban mobility, a traveller chooses a transport service from a set of different modes, which may include public transport, non-motorised transport, and driving a personal car. As discussed in chapter 1, the principal objective of this research is to understand the travel choice behaviour of the commuters who sorely rely on public transport to get around. Different discrete choice models will be estimated using stated preference data to explore the relationship between observed choice decisions and public transport attributes. Literature on travel choice behaviour, choice model estimation, and stated preference surveys design are therefore reviewed in this chapter to provide in-depth knowledge on the subject and fulfil the objectives of this study.

2.1 Choice behaviour in transport

Understanding travel behaviour remains one of the most crucial yet complex transport planning and management tasks. Travel behaviour informs transport planning authorities about travellers' decision-making processes and travel patterns in relation to trip purpose, mode of travel, timing, destination location, and route choices (Garcia-Sierra, van den Bergh & Miralles-Guasch, 2015), which forms the back-borne of travel demand analysis. Often transport planners use the economic approach based on Becker's classical theory to understand travel demand dynamics (Richard J. Carrick, 1979).

The classical theory framework posits decision-makers as rational agents who consistently make choices that bring immediate or long-term happiness. (Becker, 1976) assumes that before deciding, accurate information describing the alternatives is availed to a perfectly rational decision-maker, who, based on the provided information, subconsciously evaluates the benefits associated with all the other options to select that which maximises the subjective utility (Simon, 1959; Ben-Akiva et al., 1985). The theory, often called "utility maximisation theory", has been widely used as the framework for analysing choice behaviour in transportation planning. However, behavioural researchers like (Richard J. Carrick, 1979; Tversky & Kahneman, 1986; Garcia-Sierra, van den Bergh & Miralles-Guasch, 2015), have challenged the rationale assumptions. They argue against the unaccounted contribution of intuitive reasoning, psychological and social factors to human behaviour. Another weakness as cited in the literature is how the theory assigns the deviation from actual behaviour due to real-world complexities and an individual's limited computational capabilities as a systematic error.

2.2 Passenger quality needs in public transport

Modern-day life in developing urban spaces is structured around the need to access goods and services that lie outside the vicinity of residential precincts, and transportation facilitates access to such services. The process of travelling, being a "derived demand" resulting from the need to overcome the spatial separation of services (McFadden, 1974), individuals aim to minimise their perceived generalised costs, in the form of financial cost, utilised time, and physical/mental effort, that are associated thereof. Given a set of transport mode alternatives, the random utility maximisation theory suggests that trip makers subconsciously choose transport modes that minimise the generalised costs to match their budget constraints (McFadden, 2001b; Onderwater & Kishoon, 2017), while maximising the service quality (Agyemang, 2017a).

On the other hand, the utility from the service provider's perspective relates to cost efficiency and utilisation. Therefore, in a competitive environment, it is in the best interest of the transport planners and service providers to understand how commuters perceive the level and quality of service in their determination of passenger satisfaction indicators. The first step is to acknowledge that passengers evaluate public transport service quality as a package of attributes, then assess each underlying feature's contribution to the overall level of service. The quality of service is a subjective measure based on individual perceptions and expectations, and several indices are used to determine passenger satisfaction. These methods include the Customer Satisfaction

Index (CSI), SERVQUAL, and Service Quality Index (SQI), which is founded on the discrete choice modelling framework (Hensher & Prioni, 2002; Hensher, Stopher & Bullock, 2003; Eboli & Mazzulla, 2009a, b; Luke & Heyns, 2020), and they all rely on the passenger service quality needs.

In the American public transport capacity and service quality manual, (Kittelson & Associates et al., 2013), identify the availability of service, its convenience and comfort, as the most contributing factors towards passenger satisfaction and public transport mode choice decision. The availability of service relates to accessibility, affordability, service span, and frequency, while comfort and convenience are classified as a function of passenger load, reliability, safety, security, customer service, and travel time (European Commission, 1998). Service accessibility refers to the proximity of public transport service to the passenger's origin and destination, and the universal access to transport facilities regardless of social status or physical disabilities (Balcombe et al., 2004). This indicator can be measured in both time and space dimensions, including the ability to accommodate mobility-impaired individuals. Frequency measures the time between successive service vehicles, and high frequency implies shorter waiting times. Service span relates to the availability of the service in time-space. Affordability refers to one's financial abilities and willingness to pay for the service. Passenger load determines the ability to get a seat or the average standing passenger space; this can also be measured in terms of crowding density. Service reliability describes the level of confidence in the service's ability to deliver with regards to on-time performance, headway adherence, and excess waiting time. Travel time is componentised into access, waiting, transfer, and in-vehicle travel times. Safety relates to the possibility of being involved in an accident while using the services and security, the chance of being a victim of crime. Customer service is related to the availability of information on how to use the transport service, scheduling, customer service response rate, and the general onboard interaction with staff (Balcombe et al., 2004; Kittelson & Associates. et al., 2013).

The importance of the above service quality needs varies across people, and a body of work on passenger satisfaction has directly linked some of the variations to sociodemographic differences. An empirical study by Mayo and Taboada (2020) discusses some of the influencing factors for commuters when making mode choice decisions. The authors rank the relative importance of mode-specific attributes, availability, comfort, cost, safety, and environmental concerns across people with unique sociodemographic features for different trip purposes. For instance, both male and female commuters are found to be more sensitive to safety but least to comfort and environmental concerns, respectively. In addition, female commuters prefer “for-hire motorcycles” more than Jeepneys, unlike their male counterparts. Safety is ranked first across all the trip purposes, while comfort, environmental concerns, cost, and availability are ranked the least for work and leisure-related, family visit and school-related, business-related, and other trips, respectively.

Allen, Muñoz & Ortúzar (2019), on the other hand, uses Maslow’s hierarchy of public transport needs to rank the attributes’ contribution to public transport satisfaction. They find reliability to be the most determining attribute, followed by safety and security, customer service, and comfort. Furthermore, the results of the study suggest that the higher-income group exhibits an additional need for reliability than the low-income group. The same trend for reliability is observed between long and short trips. In contrast to reliability, customer service gains relevance for the low-income group. From the studies above, we recognise a link between customer judgments ratings and passenger service quality indicators, trip purpose, and sociodemographic attributes. The discrete modelling approach uses physical and measurable transport attributes to explain choices.

2.3 Discrete choice modelling

Discrete choice models are econometric tools that use choice observation data to estimate the relative importance of different attributes when making choices between two or more discrete alternatives (Train, 2009). The choice framework involves a decision-maker with unique characteristics, a choice set, and a decision rule (Ben-Akiva et al., 1985). The choice set is exhaustive with determinate mutually exclusive options (Train, 2009) that are defined by different attribute values. The decision-maker, who can be an individual or organisation, identifies a set of alternatives to evaluate the overall associated benefits or losses by analysing the attributes, then chooses one option with their decision rule’s guidance. The decision rule describes the internal mechanism that maps a score to the decision made.

It is important to note that the actual choice behaviour is very complicated to model, and the decision rule is a mathematical representation of how an analyst presumes the decisions to have been made by the decision-maker. The decision rules are often categorised as non-trading decision-making, such as elimination by aspects, lexicographic, dominance (Hess, Rose & Polak, 2010), and compensatory decision-making. The choice behaviour in transportation is analysed chiefly based on the utility decision rule, in which the desirability of an alternative is contingent on the alternative attributes and their relative importance to the decision-maker. The utility theory assumes a compensatory decision-making framework, that is, gaining attractiveness due to an improvement in one attribute, at the expense of poor performance in another characteristic, for example, saving travel time at the expense of an additional trip cost.

2.3.1 Random Utility Maximisation Theory

The utility of an alternative, n to an individual, a denoted by $U_{a,n}$ indicates the value or benefit the individual enjoys by choosing the alternative, and the numerical value is derived from the attributes. The development of discrete choice analysis is founded on the concept of random utility maximisation, which originates from the classical economic theory (McFadden, 2001b), and the models derived from this concept are called random utility models (RUM). The utility is subject to the individual inherent taste preferences, which may partly be explained by the observable sociodemographic attributes, given by:

$$U_{a,n} = f(X_{a,n}, \beta_a) \qquad 1$$

Where$X_{a,n}$ is a vector of alternative and individual attributes as observed by the decision-maker, n and β_a the vector parameter measuring the marginal utility of the attributes to be estimated or the contribution of the attributes to the overall utility. Given a choice set S_a of discrete alternatives, the theory suggests that the individual will choose an option that yields the highest utility out of all the utilities of other options in the choice set. The utility value is a unitless expression for preference ranking of alternatives, hence alternative, n is chosen if,

$$U_{a,n} > U_{a,j} \ \forall n \neq j, j \in S_a \qquad 2$$

For simplicity, an additive utility function is used to model individual preference. The linear utility specification makes the human behaviour deterministic, which is not the case, and to account for behavioural complexities and inconsistency, a random error term is introduced. As a result, the choice process changes to probabilistic, leading to the random choice theory. The sources of behavioural inconsistency are unobserved taste variations, unobserved attributes that the decision-maker considers, and measurement errors because of misinterpretation of data.

$$U_{a,n} = V_{a,n} + \varepsilon_{a,n} \quad 3$$

where;

$V_{a,n}$ represents the deterministic utility component, while $\varepsilon_{a,n}$, is the stochastic part that captures uncertainty.

The choice probability for individual a and alternative n is:

$$P_a(n) = P\ (\varepsilon_{a,n} - \varepsilon_{a,j} \geq V_{a,j} - V_{a,n}\ \forall n \neq j) \quad 4$$

This expression means that the probability for the individual to choose n equates to the probability of the unobserved factors for n being sufficiently better than for alternative j. This can be expressed as a multi-dimensional integral over the density of the random error term, $f(\varepsilon_n)$ (Train, 2009).

$$P_a(n) = \int I(\varepsilon_{a,n} - \varepsilon_{a,j} \geq V_{a,j} - V_{a,n}\ \forall n \neq j) f(\varepsilon_n) d\varepsilon_n \quad 5$$

The indicator function, I reduces to 1 when the ratio of the differences of unobserved to observed terms is greater or equal to 1 and 0 otherwise. The above equation represents the generalised form of choice probabilities and the ability of the indicator function to take a closed form for specific ε_n distributions, makes it possible to derive different choice model frameworks. Logit, GEV, probit, and mixed logit model forms are all derived under different assumptions regarding the specification of the unobserved term density. The analysts, therefore, choose a model with an applicable error-term distribution for their study objectives.

2.3.2 Multinomial Logit model specification

The multinomial logit model specification is commonly used in transport discrete choice analysis because of its simplicity in statistical inference and computation. The model form was extended from the binomial logit model, following (Luce, 1959) independence from irrelevant alternatives axion work. McFadden (1977) classifies the model specification as the empirical specialisation of the generalised extreme value theorem. The derivation is directed by three underlying assumptions (Ben-Akiva et al., 1985; Train, 2009) with respect to the error term. The first assumption is that the error terms are distributed independently across the alternatives and respondents, while the second assumes the distribution to be identical. This, therefore, means that the random error is the same, and there is no correlation across all the alternatives and respondents. Finally, the error term is assumed to follow the Gumbel or type 1 extreme value distribution.

The density function of the unobserved term becomes

$$f(\varepsilon_{a,n}) = \mu e^{-\mu(\varepsilon_{a,n}-\eta)} e^{-e^{-\mu(\varepsilon_{a,n}-\eta)}} \qquad 6$$

and the cumulative distribution,

$$F(\varepsilon_{a,n}) = e^{-e^{-\varepsilon_{a,n}}} \qquad 7$$

with a mean $\eta + \frac{\gamma}{\mu}$ and variance $\frac{\pi^2}{6\mu^2}$, where γ is the Euler constant.

Normalising the MNL by assigning 1 and 0 to μ *and* η respectively, reduces the choice probability to a deterministic probability function (Small, 1987):

$$P_a(n) = \frac{e^{V_{a,n}}}{\sum_{j=1}^{J} e^{V_{a,j}}} \qquad 8$$

Utility function specification.

The attribute specification in a utility function, which is generally guided by intuition, represents the priori assumption concerning the effects of the change in attribute level on choice decisions. The linear function is commonly used for utility specification in discrete choice analysis, and while this simplifies the marginal substitution of variables, the parameter estimates can be misleading for attributes that exhibit gain-loss asymmetries or the diminishing marginal utility property (Lanz et al., 2010). Empirical investigations have noted significant overestimations in willingness to pay for continuous attributes when non-linearity is not accounted for and a relatively small bias when a non-linear specification is estimated for linear functions (Torres, Hanley & Riera, 2011; van der Pol et al., 2014). No assumptions are made regarding the contribution of categorical variables to utility, effects coding, or dummy variables are used instead to avoid over-specification.

To account for non-linearity in continuous variables, the log-transform, power series, and piecewise linear specifications can be used. Unlike in linear specification, the marginal effect in these functions is dependent on the variable. As illustrated by the equations below, the linear utility specification assumes a constant marginal disutility, while in the stepwise specification, the marginal utility of a variable x is allowed to vary across different x intervals (van der Pol et al., 2014). The quadratic and logarithmic transform specifications allow the marginal (dis)utility to vary continuously with the variable x depending on the value and sign of the parameter. However, the fundamental difference between the two is that while the marginal utility in the quadratic function follows a linear relationship with the variable, it diminishes/appreciates for the logarithmic function. This can be illustrated by taking partial derivatives of the equations below with respect to the variable x_1. The marginal utilities would be $2\beta_{x1}x_1$ and $\frac{\beta_{x1}}{x_1 \ln 10}$ for the quadratic and logarithmic transform specifications respectively.

Linear: $U_{a,n} = \beta_{x1}x_1 + \cdots + \varepsilon_{a,n}$

Quadratic: $U_{a,n} = \beta_{x1}x_1^2 + \cdots + \varepsilon_{a,n}$

Logarithmic: $U_{a,n} = \beta_{x1}\log(x_1) + \cdots + \varepsilon_{a,n}$

Stepwise: $U_{a,n} = \beta_{x1c1}(x_1 \leq c_1) + \cdots + \beta_{x1ck}(x_1 > c_{k-1}) + \cdots + \varepsilon_{a,n}$ where $c_1 \ldots c_k$ represents the critical values for the variable x_1, in the utility function.

Model Estimation

Model estimation is an iterative process employed to estimate coefficients that best explain the observed choices in collected data. In other words, the aim is to find beta values that minimise the absolute difference between the choices predicted by the probabilistic function and the observed choices, thus maximising the likelihood of the parameters.

The likelihood of a choosing alternative n is expressed as $L(\beta_{a,n}) = P(n|X_{a,n};\beta_{a,n})$and the likelihood of parameters for the entire population N_a is given by the product of choice probabilities for everyone in the population.

$$L(\beta) = \prod_{a=1}^{N_a} P_{a,n}(\beta) \qquad 9$$

The product of choice probabilities is very small for modest population size, N_a and to facilitate computation, the likelihood is transformed to log scale by taking the logarithms of both sides.

The log-likelihood transforms to:

$$LL(\beta) = \sum_{a=1}^{N_a} \ln P_{a,n}(\beta) \qquad 10$$

The search process for the coefficients that maximise the log-likelihood is iterative, and the batch gradient ascent is normally used; however, more optimisation algorithms exist like the Newton Raphson method. With the batch gradient ascent, an initial vector of parameters is continuously updated by the partial derivative of the log-likelihood until a maximum log-likelihood is reached.

$$\beta_j := \beta_j + \alpha \frac{\partial LL(\beta)}{\partial \beta_j} \qquad 11$$

The ascending step is given by the learning rate α. The learning rate for the Newton Raphson method is the inverse of a negative Hessian matrix defined as the matrix of second order partial derivatives of the log-likelihood with respect to β_j as the ascending step (Zaki, Meira & Meira, 2014). Convergence is reached as the ascending step approaches zero.

Informal Tests

Informal tests are basic checks on the estimated model results to assess the behavioural consistency of the estimates with the econometric theory, judgment, and intuition. The sign of parameters relates to the direction of marginal utility, which indicates the behavioural response to changes in attributes. The sign is expected to be consistent with the rational behaviour; for instance, in a mode choice study, we expect the cost coefficient estimate to be negative because individuals always choose a cheaper mode alternative, all variables being equal. The alternative-specific constants measure preference bias towards options when all variables are kept constant across the alternatives. To avoid over-specifying parameters, the alternative specific constants are normalised by specifying one constant as a dummy or reference. The highest estimate value signifies the most preferred alternative, whilst the least indicates aversion. Besides the above tests, the analyst can also check the consistency of systematic taste variation across different sociodemographic groups. For example, the marginal disutility of costs is expected to increase with a decreasing income.

Hypothesis testing

With subject to a chosen model and utility specification, for every estimated parameter value, there exists a unique true value around which the distribution of parameter estimates approaches a normal distribution with increasing sample size N_a (Hess et al., 2020). In this context, the deviation of the estimated parameters from the true value informs the analyst about the uncertainty and precision of the estimated values. The hypothesis test checks the statistical significance of the attribute coefficients estimated against the null hypothesis that the estimate equals zero. In other words, the test is used to test for the substantial contribution of the attribute whose parameter is under investigation to the observed choices. The standard t-statistic value, which is essentially the ratio of the estimated value to the standard error, tests the hypothesis that the estimated parameter value is not h_0 , where $t_{statistic} = \frac{\hat{\beta}-h_0}{\sigma_\beta}$. Using the t-statistic, a $p - value$, defined by (Benjamin et al., 2017) as the probability, calculated under the null hypothesis, of observing the outcome or an extreme value, is then extrapolated from the standard normal distribution to indicate the level of significance. The H_0 is rejected, and the estimated parameter is confirmed to be statistically substantial if the $p - value$ falls below a type 1 error, α. When the estimate has an expected or known priori sign, the parameter significance is tested against the null hypothesis of the coefficient taking the opposite sign to avoid behavioural irrationality. In this case, the hypothesis test is naturally one-sided; as such corresponding threshold values should be used. Apart from significance testing, Hess et al. (2020) emphasise the importance of reporting estimate results in the context of precision represented by the confidence interval. This is important for real-world applications, like policy formulation where a narrow range is implementable.

Willingness to pay indicators

Discrete choice experiments are structured to encourage trade-offs between attributes, making it easier to measure the relative importance between attributes. When one of the attributes is in the monetary index, the willingness to pay for a service improvement can be evaluated. The most widely used willingness to pay measure in the transportation sector is the value of time savings (Bliemer & Rose, 2013). The willingness to pay is expressed mathematically as the ratio of the parameter under investigation to the marginal utility of cost:

$$WTP_x = \frac{\partial U}{\partial x} \Big/ \frac{\partial U}{\partial cost} = \frac{\beta_x}{\beta_{cost}}$$

Goodness fit testing

The goodness fit test indicates how well the model, which is essentially the combination of estimated parameters and variables, successfully predicts the observed choices in the data. A likelihood ratio index test statistic or rho-square is typically used for model fit, and this approach compares an initial model, with equal probability for all options in the choice set, to the estimated model. The comparison is based on the log-likelihood at equal probabilities and the log-likelihood at convergence. The rho square value, ρ^2 measures gain in log-likelihood at convergence relative to the initial log-likelihood, $LL(0)$.

$$\rho^2 = \frac{LL(0) - LL(\beta)}{LL(0)} \qquad 12$$

The rho square takes values between 0 and 1, where 0 represents a purely random model, and the latter, a perfect model fit. The gain in log-likelihood increases with additional parameters, and the rho-square scale tends to optimistically estimate model fit, especially for parameters with no influence on observed choice. To correct this overestimation, McFadden (1973) adjusted rho square, which penalises the introduction of additional parameters, is used instead.

$$Adjusted\ \rho^2 = \frac{LL(0) - \ LL(\beta) - K}{LL(0)} \qquad 13$$

Where K represents the total number of estimated parameters.

Model fit comparison for model selection

The estimation process is an iterative empirical method, where more than one model is estimated, and the best model selected. In that regard, various model-fit tests have been developed over the years to derive a parsimonious model. One of the methods is the likelihood ratio test which assesses the impact of additional parameters on the log-likelihood gain to check if the improvement is sufficient. A test statistic is formulated to compare the two models, provided that one model can be easily collapsed into the other by imposing restrictions on parameters. The test statistic, which is twice the gain in log-likelihood, is supposedly distributed chi-squared with degrees of freedom equal to the sum of the additional parameters. The test statistic, $-2[LL_2(\beta) - LL_1(\beta)]$ is then compared against a critical value from the chi-squared distribution to indicate the level of confidence at which the null hypothesis that the generic or unrestricted model provides a better fit can be rejected. The comparison of non-nested models is based on the McFadden (1973) adjusted ρ^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion values (BIC). The BIC imposes stricter criteria for additional parameters with larger sample sizes to counter massive changes in log-likelihood associated with large sample sizes (Hess et al., 2020).

$$AIC = -2LL(\beta) + 2K$$

$$BIC = -2LL(\beta) + K \ln N_a$$

Limitations of the independent and identical alternative property.

The multinomial logit model estimation assumes an identical and independent distribution of the stochastic terms across alternatives and respondents. The independence assumption nullifies the existence of unobserved utility components across the alternatives. Although the assumption simplifies parameter estimation, it remains inappropriate in many choice situations where the respondents usually assign higher utility to alternatives with common unobserved variables. For example, in a mode choice study with a car, bus, and train as alternatives, the independence assumption violates the relatively strong unobserved correlation between the two public transport modes. The identical random error distribution assumption across alternatives is violated when an unobserved variable differs across alternatives. Using the same mode choice illustration above, different levels of flexibility and privacy between car and bus violate the equal error variances assumption (Bhat, 2003).

The GEV family of choice models like the nested logit (McFadden, 1977) relaxes the independent distribution of the unobserved utility component assumption while maintaining the identical distribution property by grouping similar alternatives in nests. The error terms are only allowed to correlate across options within the same nest but denied across nests. Each nest is associated with a log-sum parameter, λ_{nest} indicating the strength of covariance within the nest. The log-sum parameter varies inversely proportional to the covariance, and as a result, the nested logit collapses to a multinomial logit model when the log-sum is equal to one (Bhat, 2003; Hess, 2005). The nested model generally develops as a hierarchical structure that can be easily illustrated by an econometric tree diagram shown in the figure below.

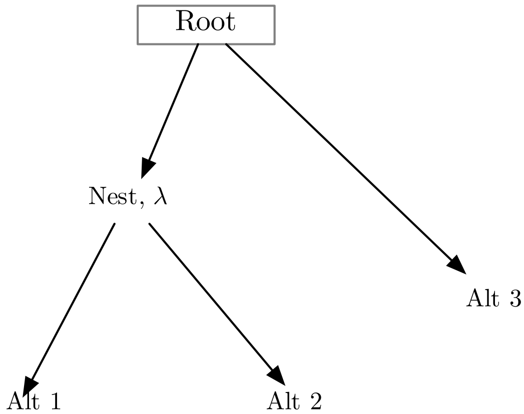


Figure 2: Nested logit econometric tree illustration.

The choice probability for individual a and alternative n belonging to nest S_m is evaluated by the product of the nest probability and the conditional probability within the nest.

$$P_a(n) = \frac{e^{\frac{\lambda_m}{I_m}}}{\sum_{m=1}^{M} e^{\frac{\lambda_m}{I_m}}} \left(\frac{e^{\frac{V_{a,n}}{\lambda_m}}}{\sum_{j \in S_m} e^{\frac{V_{a,j}}{\lambda_m}}} \right) \qquad 14$$

$$I_m = \sum_{j \in S_m} e^{\frac{V_{a,j}}{\lambda_m}} \qquad 15$$

2.3.3 Mixed multinomial logit specification.

The mixed logit presents a highly flexible class of models that relaxes the three assumptions with regards to the distribution of the unobserved utility term across alternatives and the population (Bhat, 2003; Train, 2009). McFadden & Train (2000) define a mixed multinomial logit function as a multinomial model with random parameters drawn from a cumulative distribution function. In other words, the model captures the additional random variation in taste preference across the population which cannot be explained deterministically using observed information. A priori statistical random distribution of the parameters across the population is set with a density function, $f(\beta|\Omega)$, where Ω represents the type of distribution.

The choice probability for alternative n and decision-maker a, conditional on β_n is given by the integral of the logit probability at a given set of parameters over the density of the parameters that exist in the population.

$$P_a(n) = L_{a,n}(\Omega) = \int P_a(n|\beta_n) f(\beta_n|\Omega) d\beta_n \tag{16}$$

Unlike with MNL, the continuous distribution of parameters results in an integral without a closed-form expression, making the choice probability computationally impossible.

The log-likelihood for choosing the alternative across the entire population is given by:

$$LL_n(\beta, \Omega) = \sum_{a=1}^{N_a} \ln P_a(n|\beta_n, \Omega) \tag{17}$$

The log-likelihood is simulated over a finite number of parameters draws picked systematically from a continuous domain of the parameters. Letting $\beta^{(r)}$ with $r = 1, \ldots, R$ be independent random draws from domain $f(\beta|\Omega)$, the simulated log-likelihood is reduced to:

$$SLL(\Omega) = \sum_{a=1}^{N_a} \ln \left[\frac{1}{R} \sum_{r=1}^{R} P_a\left(n|\beta_n^{(r)}\right) \right] \tag{18}$$

For panel data, where an individual makes repeated choices over periods of time, the mixed logit structure is sufficiently flexible to accommodate correlation over multiple choice scenarios for one person. Assuming the taste preference to vary across people and remain constant across repeated choice scenarios, $t = 1, .., T$, for the same individual, the simulated choice probability for n and a becomes:

$$\boldsymbol{\widehat{P_a(n)} = SL(\Omega) = \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} P_{a,t}\left(n \middle| \beta_n^{(r)}\right)} \tag{19}$$

and the simulated log-likelihood for the entire population,

$$SLL(\Omega) = \sum_{a=1}^{N_a} \ln \widehat{P_a(n)} \tag{20}$$

2.3.4 Latent Class models

Apart from GEV and MMNL, latent class models provide an alternative methodology to account for heterogeneity in discrete choice analysis. The model specification derives from the general theory of choice behaviour, which suggests that choice decisions accurately reflect both unobserved latent heterogeneity and observed information (W. H. Greene & Hensher, 2003). While the mixed multinomial logit assumes a parametric prespecified continuous distribution of unobserved taste preference across the population, the latent class logit structure takes a discrete and semiparametric approach in modelling latent heterogeneity (W. Greene, 2001; Martínez-Cruz, 2015). The sample population is assigned to a finite number of heterogeneous classes that are distinguished by different parametric vectors, with a possibility of some parameters being constant across classes, and homogeneous taste preference and sensitivity are assumed within latent classes. The individuals in the population are probabilistically grouped into different classes by a latent class model, and the utility of classification can be specified as a function of the observed sociodemographic features.

The class membership probability of an individual a being a member of latent class c specified as a combination of sociodemographic features Z_a is given by a logit function,

$$\pi_{a,c} = \frac{e^{\delta_c + f(\gamma_c, Z_a)}}{\sum_{i=1}^{C} e^{\delta_i + f(\gamma_i, Za)}} \qquad 21$$

where $\delta_c$represents a class-specific coefficient, and γ_c a parametric vector capturing the contribution of the observed individual class membership variables, Z_a to the class assignment probabilities.

The choice probability of a and n conditional on latent class c is given by

$$P_a(n|\beta_c) = \frac{e^{\beta_c x_{n,a}}}{\sum_{j=1}^{J} e^{\beta_c x_j}} \qquad 22$$

The unconditional probability of the decision-maker and the alternative can therefore be expressed as:

$$P_a(n|\beta_c) = \sum_{c=1}^{C} \pi_{a,c} \, P_a(n|\beta_c) \qquad 23$$

An empirical study by Jin Lee & Zhang (2003) combines the latent class and mixed logit specifications to relax the restrictive homogeneity assumption within a latent class. The mixed multinomial logit specification is exploited in both class allocation and within a class to allow for a continuous variation of preference across all members. The model estimation can be computationally expensive; however, it gives more insights into preference behaviour dynamics by allowing inter-and intra-latent class heterogeneity. The unconditional choice probability of an individual a choosing the alternative n is conditional on the random distribution of both class-membership and within-class parametric vectors, i.e., $f1(\pi|\Omega_\pi)$ and $f2_c(\beta_c|\Omega_{\beta_c})$.

$$P_a(n|\Omega_\pi, \beta_c) = \int \sum_{c=1}^{C} \pi_{a,s} \left[\int P_a(n|\beta_c) f2_c(\beta_c|\Omega_{\beta_c}) \right] f1(\pi|\Omega_\pi) d\pi \qquad 24$$

2.4 Discrete choice experiment

A discrete choice experiment is a process of gathering choice data for taste preference studies. In discrete choice structures, an individual, based on their taste preferences, chooses an alternative, which can be a product or service, from a set of finite mutually exclusive possible alternatives. Attributes characterise the alternatives, and the choices are based on attributes trade-off (Mcfadden & Train, 2000; Train, 2009). Discrete choice experiments are commonly used in applied economics, marketing, health, environment, transport, and many other sectors to understand taste preferences for welfare, and policy assessments and interventions to ensure efficient resource allocation. The analysis of discrete choices is an empirical method, and the choice data is collected from revealed (RP) and stated preference (SP) surveys.

2.4.1 Revealed Preference versus Stated Preference survey

The revealed preference analysis methods draw statistical inferences on choice decisions made in real life, while the latter involves choice observations from hypothetical scenarios (Ben-akiva et al., 1994). In a stated preference survey, respondents are presented with hypothetical alternatives, some of which may not be available, and asked to select one option by trading off the attributes. While the approach gives the analyst control over the selection of alternatives and their characteristics, the observed preferences may not reflect the actual choice behaviour as captured by revealed data. The reliability of stated choice experiments has been a concern to many researchers who cite the lack of realism in the hypothetical structure, as there are no consequences to the decision-maker (Cherchi & Hensher, 2015a). Revealed preference data on the contrast is much more valid and reliable; however, there is limited information with regards to the possible alternatives that might have been considered in the decision-making process. Moreover, the attributes are highly correlated with little variability, which makes it challenging to explore the trade-offs between alternatives. Stated preference and revealed preference data can be combined to exploit the advantages of both approaches while addressing the disadvantages, which has proven to improve the accuracy of parameter estimates (Ben-akiva et al., 1994). The hypothetical bias cannot be eliminated in its entirety; however, building complex survey tasks can significantly reduce the potential bias. The survey task complexity relates to the number of attributes, levels, and choice scenarios. As the complexity increases, the hypothetical choice structure in the stated choice experiment asymptotically converges to real world choice setting, and so does respondent engagement. However, when presented with complex tasks, respondents use simplifying strategies to reduce the mental effort required to evaluate options (Elisabeth Kløjgaard et al., 2012). The level of complexity should therefore avoid cognitive burden to respondents while minimising the bias.

2.4.2 Design of a discrete choice experiment

Discrete choice experiments are conducted to determine how different factors influence choice decisions; as such, the design of these experiments is important because it impacts the amount and quality of information that can be extracted from the survey (Choice Metrics, 2018; Kjaer, 2005). The two main important aspects of discrete choice experiments are statistical design and the selection of the alternatives, attributes, and levels. The combination of attributes and levels plays a vital role in encouraging respondents to make trade-offs, while the range and distribution of levels impact the statistical significance of the model output. Poor distribution of levels can, however, be compensated by using large sample sets at high costs (Choice Metrics, 2018).

2.4.3 Design of Alternatives, Attributes, and Levels

When designing discrete choice experiments, the first and vital step is defining the alternative attributes of interest. The potential attributes can be identified from literature review, direct questioning of individuals, focus group discussions, and the selection should be consistent with the study objectives, at the same time remaining meaningful and relevant to the target population (Elisabeth Kløjgaard et al., 2012; Mazur & Bennett, 2008). An optimum number of attributes should be selected, and it is crucial that all the relevant attributes are included. In line with Kjaer (2005), the formulation of attributes should ensure that the combination covers all the important aspects of the study question while remaining mutually independent. The issue of causality between attributes has been pointed out in literature as a common problem related to attribute establishment. According to Blamey et al (2002), the effect can be eliminated by combining the attributes into a single composite attribute. Alternatively, framing statements designed to reduce causal strategies can be included in the questionnaire.

Once the attributes have been established, appropriate levels for each attribute are determined and then assigned. The researcher needs to decide how best the levels can be expressed and presented to the target population. Like with attributes, the levels need to be reasonable, policy-relevant, and relatable to respondents' experiences. Most importantly, as Sanko (2001) suggests, the level of attributes should present competitive trade-offs that cover a range of valuations. Attribute levels can be expressed quantitatively or qualitatively in the form of words, percentages, numbers, pictures, charts, graphics, etc. When a combination of textual and graphical attributes is used, the researcher should ensure that the pictorial does not dominate the texts to reduce the potential survey response bias, in which respondents focus on the dominating attributes in making choice decisions. The other problem with the graphical presentation of attribute levels is potential bias emanating from the subjective decryption of the embedded numerical information, which risks the validity of the discrete choice experiment (Veldwijk et al., 2015). The range of attribute levels should be sufficiently wide to accommodate respondents from different sociodemographic groups and prevent them from ignoring some attributes. Particular attention needs to be paid to the relationship between the number of attribute levels and the possible effect of attributes on estimation; for example, two levelled attributes can only allow linear effect estimation, while more than two can estimate non-linear effects (Lancsar & Louviere, 2008b).

2.4.4 Generation of experimental design

The identified levels are assigned to the attributes to generate a set of hypothetic choice scenarios to be presented to respondents in the stated choice experiment. There exists a number of experimental designs, but the aim is to determine a befitting one. The choice of an experimental design depends on the analyst's decisions regarding the experiment being labelled or unlabelled and the need to satisfy the attribute level balanced property. Most of the experimental designs are however created to satisfy the attribute level balance property. With attribute level balance, each attribute level is supposed to appear an equal number of times, which minimises the variance in the coefficients (Choice Metrics, 2018). The two commonly used statistical experimental designs are full factorial design and fractional factorial design. The former consists of all possible attribute level combinations, some of which can be dominated, while with fractional factorial, a sample of the full factorial design is selected (Choice Metrics, 2018). The number of possible scenarios is given by the product of the levels across all the attributes of all alternatives, and in mathematical language, this is expressed as:

$$Scenarios = \prod_{\forall n} \prod_{\forall x} Level_{n,x} \qquad 25$$

Where:

n is the alternative and x the attributes

To put that into context, the number of possible choice scenarios for a choice experiment with three alternatives, three attributes, and three levels would be eighty-one. Although the full factorial design allows estimation of the effects of each attribute interaction, it is practically impossible for one respondent to face all the hypothetical choice scenarios in a stated choice experiment (Lancsar & Louviere, 2008; Choice Metrics, 2018). The art of experimental design is therefore to generate a sufficiently small number of behaviourally meaningful combinations of levels from the full factorial design that satisfy attribute level balance (Rose et al., 2013). Choice Metrics (2018) identifies orthogonal and efficient designs as the two main approaches used for locating a good design. The orthogonality property is satisfied when the attributes are balanced, and the parameter estimates are statistically independent. This is only achievable at limited attributes and levels combinations, and for other combinations, a near-orthogonal design is used (Choice Metrics, 2018; Mangham et al., 2009). The efficient design approach, on the other hand, requires prior information about parameters to create an asymptotic variance-covariance matrix that predicts minimum standard errors under which efficiency is reached. While the orthogonal design minimises the correlations between attributes levels for choice scenarios, the efficient design aims to minimise the determinant of asymptotic variance covariances in parameter estimates, sometimes at the expense of level attributes orthogonality.

Furthermore, (Bridges et al., 2011; Choice Metrics, 2018) argue that the correlation structure is irrelevant when deriving non-linear statistical models, like MNL models. More importantly, orthogonal designs often generate identical and dominated choice scenarios, yielding high standards errors even with large sample size. This is not the case with efficient designs, which, when provided with priori parameter values, create scenarios with relatively low standard errors at smaller sample sizes. As such, efficient design is used to structure this discrete choice experiment. Statistical efficiency is often expressed as $D-error$ which is computed as the determinant of the AVC matrix, so the efficient design only arrives at the lowest $D-error$ (Mariel et al., 2021). The Ngene manual, (Choice Metrics, 2018), lists a number of efficiency measures that have emerged in literature over the years. Apart from $D-error$, there is $S-\ error$ that aims to minimise the sample size, A-error for variance, $C-error$ for WTP, and $B-error$ for utility balance.

2.5 Conclusion

The literature review provides a solid theoretical foundation for the development of discrete choice models in public transport economics; from experimental design to model estimation and results interpretation. The econometric framework based on the random utility maximisation theory is often used to understand choice decisions. The choice framework involves (i) a decision maker with unique sociodemographic features; (ii) a choice set of finite mutually exclusive alternatives defined by different characteristics; and (iii) a decision rule which maps a score derived from the evaluation of the alternatives based on their features to the decision made. It was important to explore the concept of *service quality* in public transport to understand passenger satisfaction measurement criteria, and the contributing public transport characteristics. An additional empirical studies review provided substantial evidence of the systematic variance in the relative importance of the attributes across populations with different sociodemographic features, and it is critical to understand these behavioural nuances.

The random utility maximisation framework suggests that the mode with the highest utility score is chosen. Four discrete choice model types were identified; multinomial logit, mixed multinomial logit, nested multinomial logit and latent class, and their utility equation specifications were reviewed to understand their derivation assumptions and limitations. Furthermore, the literature looked at the model estimation framework, hypothesis testing and the behavioural interpretation of the estimates. The aim was to understand the applicability of the choice models to public transport and evaluation of the willingness to pay for service upgrades. Lastly, the literature reviewed the design, generation, and implementation of discrete choice experiments, with the aim to identify the different choice data collection methods and explore their fundamental differences and limitations. Two survey techniques were identified; revealed and stated preference, and the latter was found to be applicable to the study. Discrete choice modelling is an empirical exercise which relies on the quality of data, so the generation of experimental design and delivery of the survey are fundamental in minimising the possible data collection bias.

3.0 SP SURVEY DESIGN AND DATA COLLECTION

3.1 Questionnaire Development and Survey

The literature review on discrete choice experiments provided a solid background on different data collection methods and given that kombis are not allowed to offer services in Harare, a stated preference survey was found to be appropriate to fulfil the objectives of this study. A series of hypothetical choice scenarios described in terms of public transport attributes were constructed, and respondents' choice behaviour was analysed based on the random utility maximisation theory framework.

3.2 Study Area and Sampling strategy

The study targets high-density suburbs in the City of Harare, where most public transport captive users live. The Greater Harare province comprises three administration-level districts, the urban district, Harare rural, and the peri-urban, as shown below.

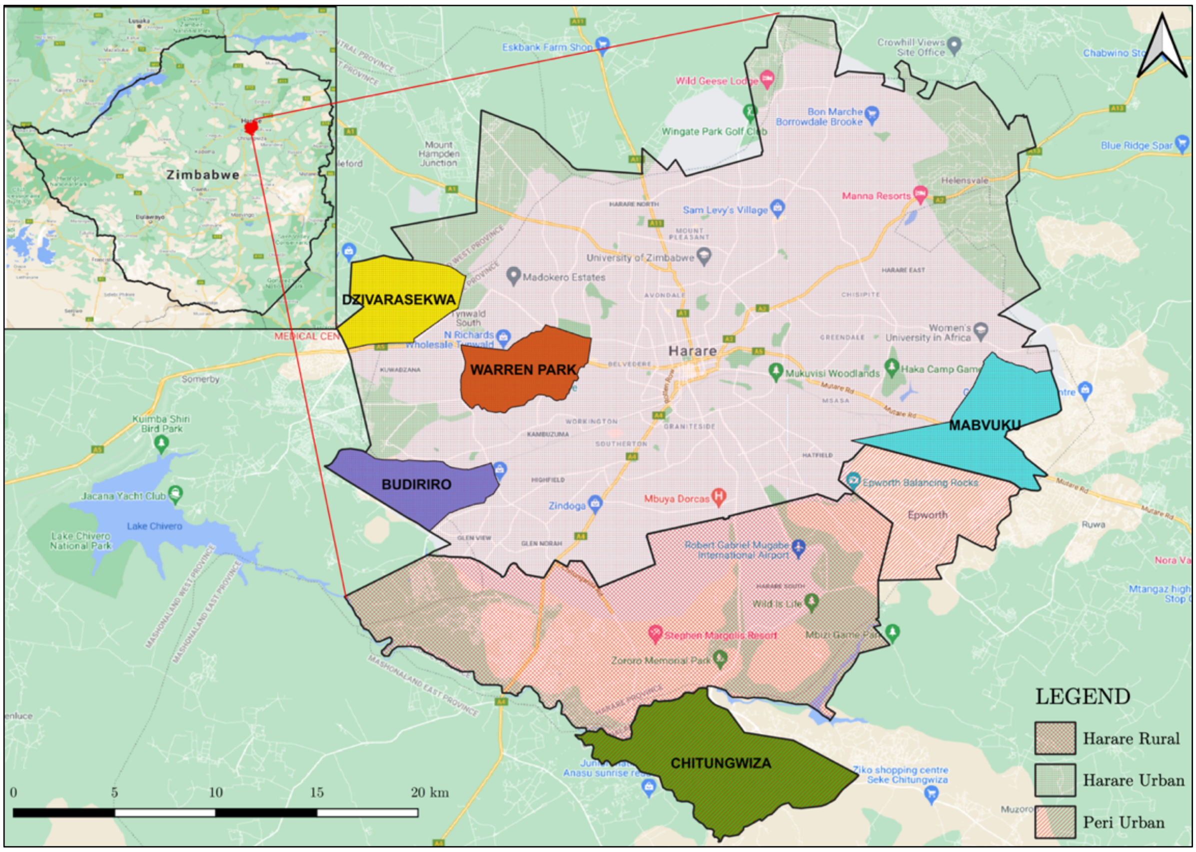


Figure 3: The location of greater Harare and sampled residential suburbs.

The high-density residential suburbs are mainly located to the south and western sides of the city centre, while the wealthy reside on the northern side. The peri-urban suburbs are predominantly low-income residential areas relatively far from the employment centres. A multistage sampling strategy was employed at five high-density residential suburbs across Harare to assess heterogeneity across neighbourhoods. The location of the sampled suburbs with reference to the central business district ranges from ten to twenty-five kilometres, which helps to capture mode choice decisions for both short and long trips.

3.2.1 Defining Attributes and Levels

The art of constructing alternative attributes and levels is derived from the need to capture as much relevant information regarding choice behaviour as possible. A combination of literature review, General Transit Feed Specification (GTFS) data analysis, and focal groups were used to define attributes and levels. (Redman et al., 2013) classifies transport attributes as physical and perceived variables. Physical attributes can only be measured without user involvement, and their impacts on the decision-maker are based on assumptions, e.g., frequency, speed, cost, etc. On the contrary, perceived attributes are measured directly from user experience or perception, for example, comfort, safety, convenience, etc. The table below lists sociodemographic, trip, mode-specific and perceived attributes identified from the literature.

Table 1: Discrete mode choice variables for previous studies

Attributes	Alternatives	Country
Sociodemographic: gender, age, income, education, occupation status, vehicle ownership, driving licence *Travel mode*: travel time, cost	bus, auto, auto shared,	Chennai, India
Sociodemographic: gender, age, income, education, employment status, car ownership *Travel mode*: time (travel, access, waiting, transfer), distance, cost *Subjective*: environmental awareness, attitude	walk, bus, car, taxi, metro, bicycle	China
Sociodemographic: route *Travel mode*: time (Access, transfer, in-vehicle, waiting), fare, crowding, car park fee, security, Method of Payment	bus, train, car, taxi	South Africa

Sociodemographic & trip: gender, age, education status, trip destination, trip distance, occupation, car ownership, household *Travel mode*: availability, fare, frequency, speed *Pecieved Variables*: availability, pollution, comfortability, convenience, safety, evade congestion	trotro, car, walking, motorcycle, bicycle, bus, taxi	Ghana
Sociodemographic & trip: Household Income, trip purpose, level of mode captivity *Travel mode*: travel cost, in-vehicle travel time, walking time, seat availability, number of transfers	bus, BRT, car, taxi, gautrain, train	South Africa

(van Zyl & Hugo, 2002; Venter, 2016; Agyemang, 2017b; Hayes & Venter, 2017; Kunhikrishnan & Srinivasan, 2017; Liu et al., 2020)

The focus group discussions were instrumental in refining the attributes to mimic the real-world choice environment. Many experiences with the current public transport service (ZUPCO) were shared in these discussions, and most of the responses were in reference to the kombi level of service. People expressed their dissatisfaction with the current system on the grounds of safety, crowding, long walking distances, and waiting times to access the buses; however, some were grateful for the cheap bus fares. Due to the economic environment in Zimbabwe, the average monthly income in high-density residential areas is approximately equivalent to USD$100, and the focus group members raised concerns over informal public transport prices that cost close to USD$1 per seat per direction. The bus termini in the CBD are always overcrowded and chaotic during the evening peak periods; young men and women struggle to board the buses. In support of this, women shared horrific experiences of sexual assault and theft at these facilities.

Attribute levels can be expressed as categorical, continuous, or percentage and should cover the full range of possibilities for choice options under investigation. Kombis GTFS data collected by "*WhereIsMyTransport*" in 2019 was extracted from the World Bank website[1] and then analysed to evaluate kombi trip attributes[2]. ZUPCO fare's structure was captured from the company website. Kombi trip routes originating

[1] City of Harare GTFS data link: https://datacatalog.worldbank.org/dataset/harare-zimbabwe-general-transit-feed-specification-gtfs

[2] GTFS data analysis methodology, (Prommaharaj et al., 2020)

from the suburbs to the city centre are shown in Figure 4 below, and trip attributes in the table that follows.

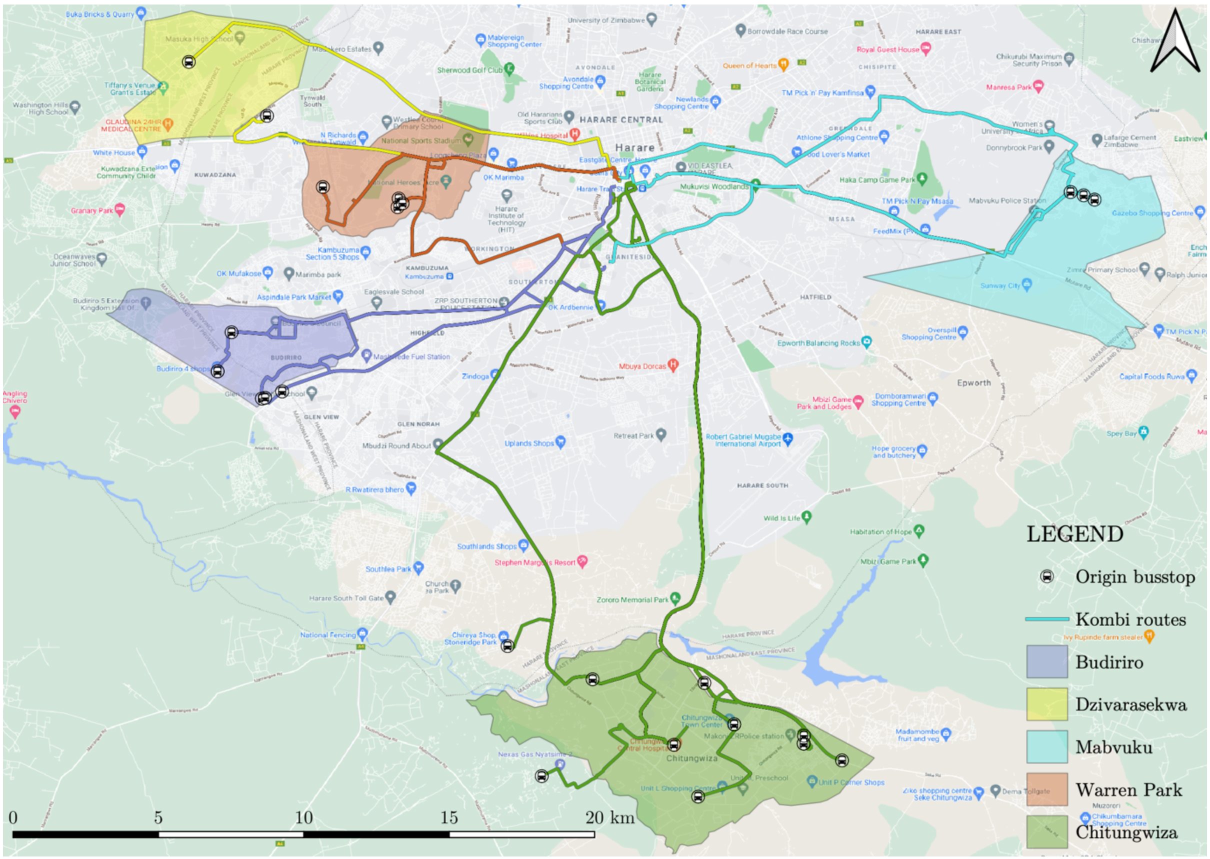


Figure 4: Kombi trip routes from the study area.

Table 2: Kombis and ZUPCO trip attributes

Trip Origin	Average distance	GTFS frequency	GTFS travel time	ZUPCO Fare[3]
Dzivarasekwa	15.7km	5 minutes	45 minutes	Z$60
Warren Park	10.0km	9 minutes	28 minutes	Z$60
Budiriro	15.4km	7 minutes	44 minutes	Z$60
Chitungwiza	25.3km	5 minutes	57 minutes	Z$90
Mabvuku	18.1km	5 minutes	41 minutes	Z$60

Travel time components levels were referenced/pivoted on kombi attribute levels extracted from GTFS data, while fare levels are referenced on ZUPCO fare.

[3] 1USD = $Z160 at the time of survey

Referencing attribute levels to current or known trip characteristics makes it easier to develop a plausible and relatable hypothetical mode alternative to subjects. In-vehicle travel time is different across the study areas, so the levels are referenced on respective rounded GTFS travel times; that is, **T** = 30 minutes for Warren Park, 45 minutes for Dzivarasekwa, Budiriro, and Mabvuku, and 60 minutes for trips originating from Chitungwiza. In the same way, fare levels were defined as a function of **F**, the current ZUPCO fares shown in Table 2 above. The level of seat arrangement is a categorical qualitative measure expressed on four levels for the ZUPCO bus (seat available, standing), car passengers (seat available, pick-up trailer), and kombi (seat available, "pakadoma"[4]), as shown in the Table 3 below.

Table 3:Proposed attributes and levels

Attribute	*Description*, [*Symbol*]	*Levels*
Access time	Time from origin to a bus stop, [AT]	3 (5, 10, 15) minutes
Waiting time	Time at the bus stop to vehicle departure., [WT]	3 (5, 10, 15) minutes
Seat Arrangement	available, standing, "pakadoma", pick-up trailer [SArr]	3 (0, 1, 2,3)
Travel time	In-vehicle travel time, T: *kombi travel time*, [IVT]	3 (0.75**T**, 1**T**, 1.5**T**)
Fare	The fare paid, **F**: *ZUPCO trip fare in Z$*, [F]	3 (1**F**, 1.5**F**, 2**F**)
Alternatives	Car hitchhike, Bus, Kombi	3 choices

3.3 Generation of Experimental design in Ngene.

The attributes and levels defined in the table above give a total of 14,348,907 possible combinations, which are practically impossible for one respondent to handle; as such, the efficient design approach was used to generate a manageable and behaviourally meaningful choice set using *Ngene*, a software used to create experimental designs (Choice Metrics, 2018). Most importantly, the efficient design approach yields reliable results with relatively low standard errors with small samples size when provided with

[4] "Pakadoma" is a colloquial name for an engine seat located just behind the driver seat row

prior information on the utility function and parameter values. Assuming rational behaviour, the design hypothesis was that the population is sensitive to all the attributes with a strong aversion to ZUPCO, as indicated in the focus group discussions; hence all the attributes have negative marginal utilities. The priori parameter signs and values were obtained from a combination of past studies and intuition. A mode choice study conducted by Venter (2016) for the City of Johannesburg was used to guide the definition of priori parameter values for the design. The author followed the random utility maximisation framework to understand the difference in mode preference between choosers and public transport captives.

$asc_{kombi} = 0.45,$ $\quad asc_{bus} = -0.53$

$\beta_{accessTime} = -0.045,$ $\quad \beta_{waitingTime} = -0.039,$

$\beta_{available} = 0,$ $\quad \beta_{standing} = -0.091$

$\beta_{pickup-truck} = -0.095,$ $\quad \beta_{kadoma} = -0.085$

$\beta_{InvehicleTime} = -0.033,$ $\quad \beta_{Fare} = -0.291$

$where\ asc_n$ represents the alternative specific coefficient for transport mode n.

Utility specification

The utility functions for the alternatives specified below are coded in the Ngene software package to generate an $S - efficient$ design.

$$U_n = asc_n + \beta_{accessTime} \times AT_n + \beta_{InvehicleTime} \times IVT_n + \beta_{waitingTime} \times WT_n + \beta_{SArr_n} \times SArr_n + \beta_{Fare} \times Fare_n \qquad 25$$

The minimum number of choice scenarios S is determined by $\frac{K}{J-1}$, where K is the maximum number of parameters and J is the number of alternatives. In our study, since $K = 9$, the minimum number of scenarios, $S = 5$. A respondent can easily handle five hypothetical choice scenarios in a survey; three blocks of eight scenarios are used to increase efficiency.

Ngene Output and Statistical efficiency

The table below gives an efficiency index summary for the study. The $D-error$ is finite and less than 1, indicating that the data collection is valid, and a model can be estimated. The minimum sample size at 95% level of significance is 361

Table 4: MNL efficiency measures.

D optimality	61.39%						
D error	0.2343						
A error	0.3723						
B estimate	78.6273						
S estimate	**361.267**						
Priori	$\beta_{accessT}$	$\beta_{waitingT}$	$\beta_{seat(d1)}$	$\beta_{seat(d2)}$	$\beta_{seat(d3)}$	β_{InvT}	β_{Fare}
Fixed priori	-0.045	-0.039	-0.091	-0.095	-0.085	-0.033	-0.291
Sp estimates	208.9597	251.8970	304.979	355.352	361.267	308.485	6.2050
$Sp\ t-ratio$	0.1356	0.1235	0.1122	0.1040	0.1031	0.1116	0.7868

The levels were coded as generic values -1,0,1, and the efficient design output is shown in Appendix 1. The outcome was then used to generate four[5] survey designs for all the suburbs by simply replacing the corresponding values of in-vehicle time and fare levels.

3.4 Construction of the Questionnaire

A discrete choice analysis is an empirical exercise that relies on data quality; as such, the survey structure should be eligible and easy to follow (Cherchi & Hensher, 2015b; Veldwijk et al., 2015). A paper-based draft version of the survey questionnaire was developed for the blocked efficient design output. Although time-consuming, the method was chosen because it allows the interviewer to answer questions, should respondents seek clarity, leading to accurate data collection. The first section provides an introductory statement about the purpose of the study and the hypothetically structured context of the survey. The respondents were assured of anonymity and requested consent to proceed with the survey.

[5] Budiriro and Dzivarasekwa suburbs have the same IVT and Fare levels.

In the second section, revealed preference data on the recent previous frequent trip was collected and later analysed to understand the travel patterns and estimate travel demand. In addition to the trip attributes in the stated preference, information pertaining to how often the respondents use public transport was collected, including the trip purpose and destination. The third section introduced the respondents to eight hypothetical stated choice preference questions, in which they had to choose one preferred public transport mode. The choice scenarios were presented to the respondents in the format shown in Figure 5 below.

The focus group members lamented over the safety, security, and reliability of ZUPCO services, so in the fourth section, attitudinal information towards the service quality indicators was collected to understand the general perception of the safety and reliability of the current public transport systems. Two attitudinal statements about safety and security, and the reliability for each mode are presented to the respondents to rate on a five-point Likert scale, with ratings from "strongly agree" to "strongly disagree".

The questionnaire ends with four final questions aimed at sociodemographic attributes. Information about gender, age, employment status, and monthly income will give more insights into possible systematic differences in taste preference across the population.

attributes	 **kombi**	 **zupco**	 **hitchhike**
Access Time	10 minutes	5 minutes	10 minutes
Waiting Time	5 minutes	5 minutes	15 minutes
Seating arrangement	seating	standing	pick-up trailer
Travel Time	33 minutes	45 minutes	45 minutes
FARE	$Z 120	$Z 90	$Z 60
CHOICE (X)			

Figure 5: Sample choice scenario design and layout.

3.4.1 Pilot Survey

We conducted a pilot survey in Tafara, a high-density suburb that shares boundaries with Mabvuku, to determine its time to complete and test for wording clarity, ease of the tasks, and respondent engagement. Twenty-five questionnaires were randomly distributed, and the results were analysed. We noticed that respondents understood access time better in terms of walking distance to access public transport services, so in the main survey, the levels were expressed in both walking time and distance at an average walking speed of 1.5m/s from ITDP (2018).

3.4.2 Main Survey

A sample of the questionnaire is attached in Appendix 2. Ten university students were recruited and trained in October 2021 ahead of the survey to assist with implementation. The training aimed to explain the context of the study and how to conduct a stated preference survey. The interviewers were divided into pairs per suburb, and we decided that surveys be done at bus stops during the morning peak period as commuters will be waiting for transport. A total of 375 completed questionnaires were obtained from the exercise. The data was compiled and captured into an excel spreadsheet for cleaning and editing.

3.5 Description of data and findings

3.5.1 Sociodemographic statistics

After cleaning the data, 361 individual datasets were found to be useable for analysis. A descriptive analysis of the data was done to understand the distribution of the population across different sociodemographic groups and trip characteristics. 43% of the respondents were female and 57% male; this matches the 41%: 59% gender distribution for the economically active group recorded in the 2012 census findings (Zimstat, 2013). The economically active referred to the population available for services and goods production as realised in the national income statistics, and this included unpaid family workers, employers, unemployed, paid employees, and the own-account workers.

Employment status

The questionnaire classified employment status into four categories; formally employed, informally employed, students, and unemployed. The respondents who were by the law, not subject to national labour legislation and income tax were classified as informally employed, and those who worked for a legal entity as formally employed. We found that 42% of the respondents were formally employed, 27% stated that they were informally employed, and 16% were students.

Income

The figure below shows a representative chart for the monthly income and the respondents' employment status. Due to the economic challenges that the country is currently facing, the remuneration packages are generally low. Only 17 respondents out of 361 reported having a monthly income of more than Z$30,000, with 24% of the top earners being formally employed and the remaining 76% informally employed. Forty-four respondents asserted that they had no income, and this group comprises of the unemployed and students. A closer look at the employed respondents shows that 93% of the group have an income of less than Z$30,000 (*an equivalent of* $187.5*USD*); however, the monthly income for most of the respondents ranges between Z$20,000 and Z$30,000.

Figure 6: Sample distribution of income.

Age

The age of the population was grouped into five subgroups, as shown in the figure below. Most of the respondents were between the age of 25 – 34 years, and only 5% reported being over 54 years. More than 64% of the respondents were less than 35 years old, indicating the most active population.

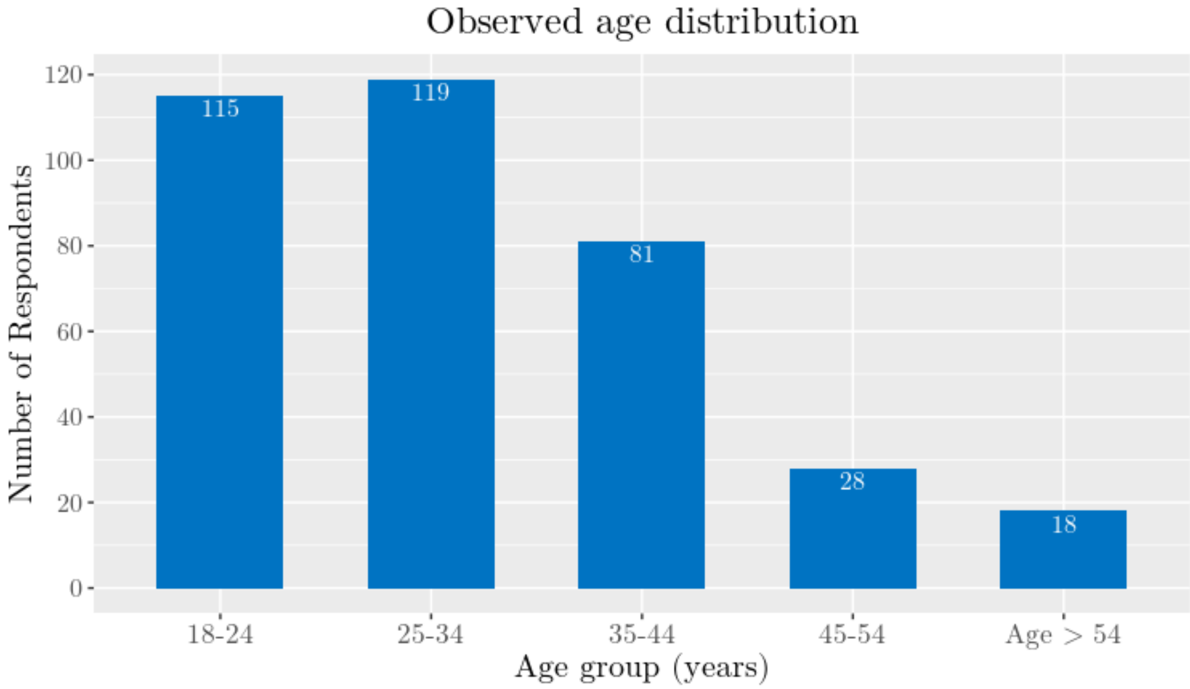


Figure 7: The distribution of age in the sample.

3.5.2 Trip characteristics

Trip purpose and destination

Harare is generally monocentric in nature and the most vibrant economic zone is around the central business district (CBD). Analysis of results with regards to destination revealed that 64% of the respondents passed through the city centre, and the destinations of the remaining 36% were before the CBD. After classifying trip purposes into three groups as work trips, education trips, and leisure/shopping trips, we found work-based trips to account for 71% of the trips, while only 12% and 16% were leisure/shopping and education trips. We found that out of the 257 work trips, 56% ended in the city centre whilst the remainder alighted along the road. At least 80% of both education and leisure/shopping trip purposes passed through the city centre. The figure below shows the distribution of trips purpose across the destinations.

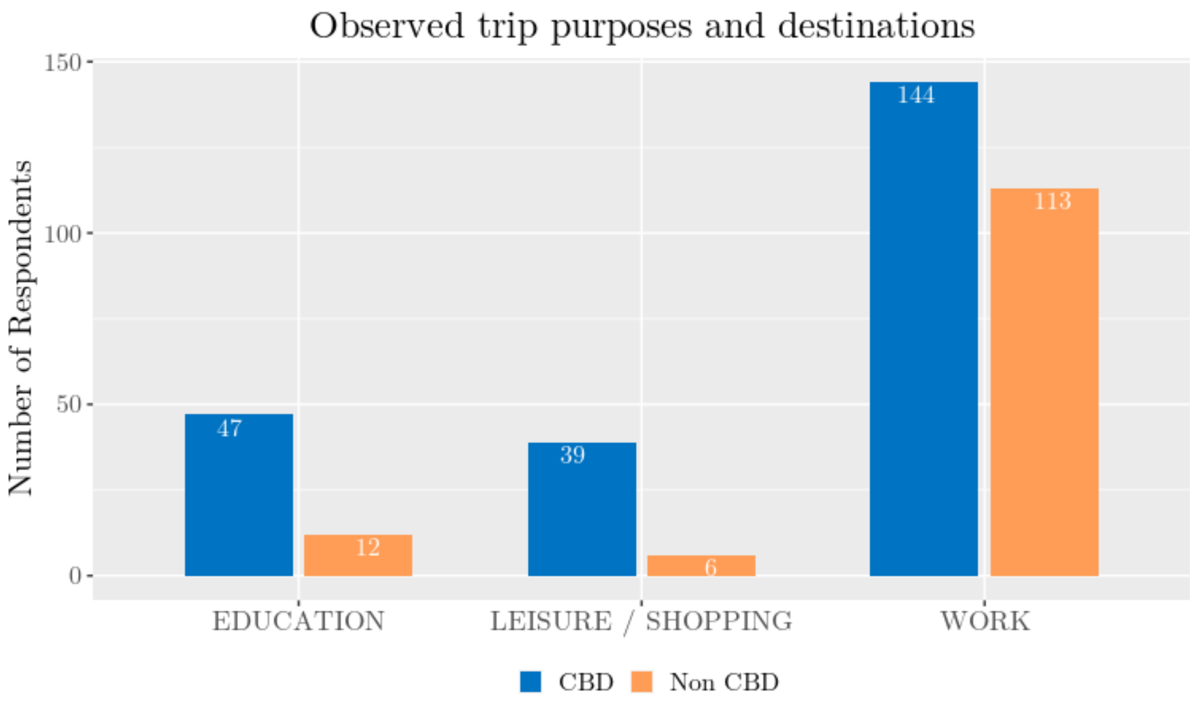


Figure 8: The distribution of trip purpose and destination.

Access and waiting time

Figure 9 below shows the distribution of access time across the sample suburbs.

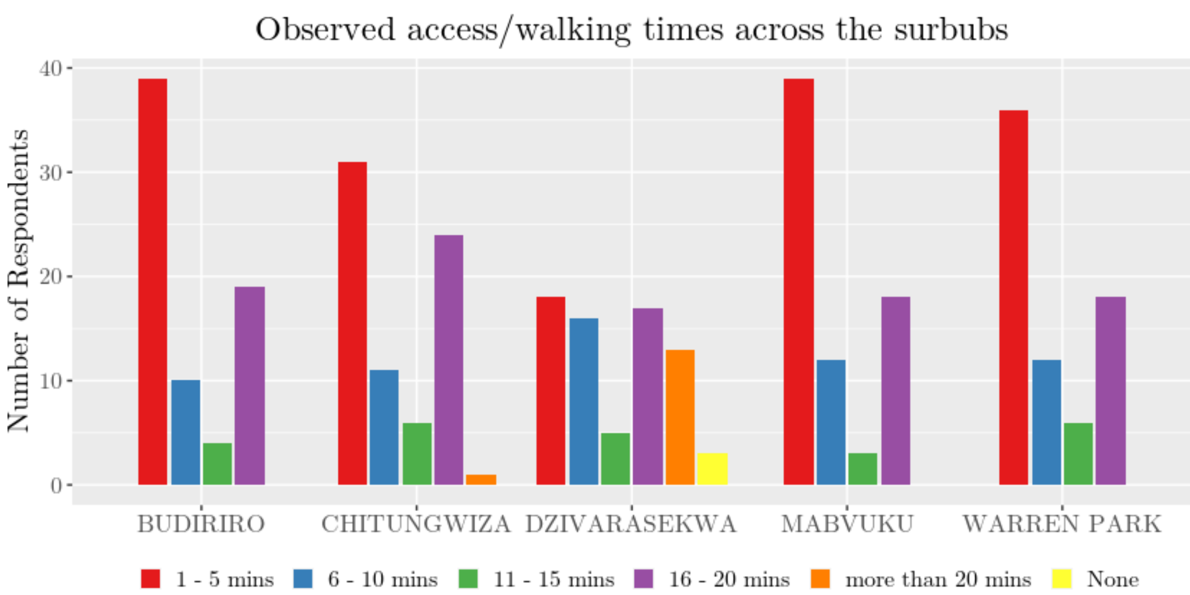


Figure 9: The distribution of accessibility across the suburbs.

Public transport is generally accessible, 62% of the respondents managed to access public transport within 10 minutes or 900 metres, and only 4% had to walk for more than 20 minutes or 1.8 kilometres. Comparing accessibility across the suburbs, the results reveal that Mabvuku has the highest accessibility index, with more than 71% of the respondents having access within 10 minutes. The least accessibility index was recorded in Dzivarasekwa, not more than 47% of respondents could access public transport within 10 minutes, and 18% had to walk for more than 20 minutes

The respondents experienced long waiting times in their previous trip, only 46% could get transport within 20 mins of arriving at the bus stop, and more than 17% had to wait for more than 40 minutes to get transportation. Chitungwiza had the most unbearable waiting times; 15% of the respondents waited for more than an hour at the bus stop, while only 34% could get transport within 20 minutes. The distribution of waiting time across the suburbs is shown in the figure below.

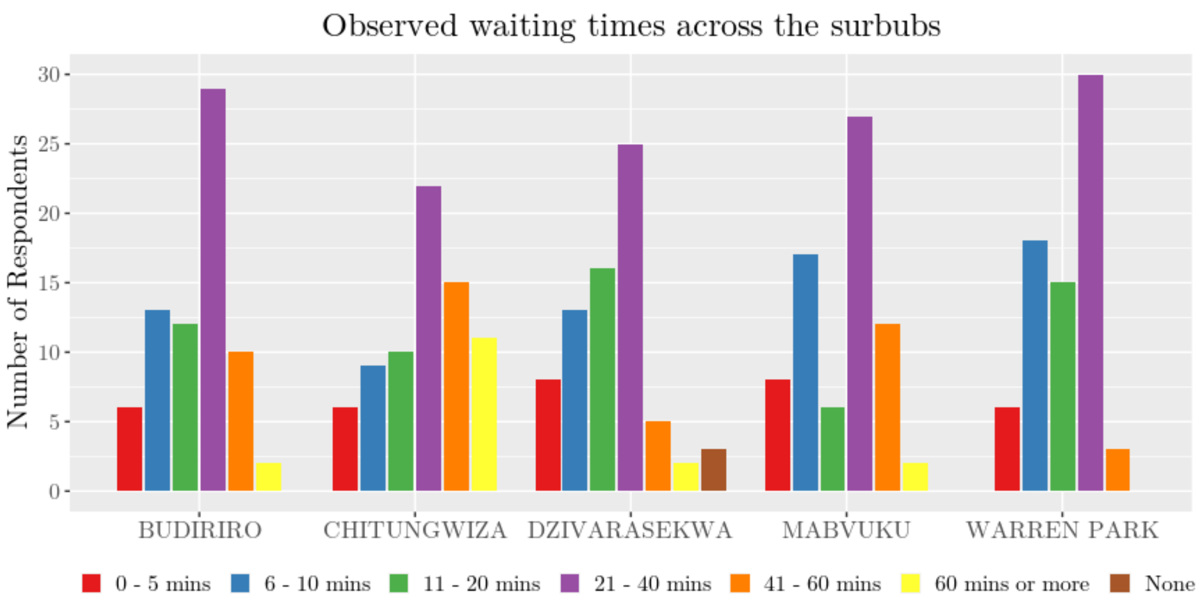


Figure 10: The distribution of waiting time across the suburbs.

Mode used and seat arrangement

In the previous trip, the modal split was 0.8%, 27.1%, 27.4%, and 44.6% for NMT, car hitchhiking, kombi, and ZUPCO, respectively. ZUPCO had the most significant market share because the service operates under a monopoly arrangement, whereas kombis are restricted from offering services. Two respondents reported to have cycled to work the previous day, and one person walked. Most of the respondents managed to get a seat, and 50% of hitchhikers sat in the pick-or lorry trailer while only 20% who used kombi sat “pakadoma”, as shown in the figure below. The difference in seat arrangement is wide for kombis because a kombi can only accommodate four “kadoma” and eighteen seating passengers. This is not the case for cars and ZUPCO buses. On the other hand, ZUPCO buses and open-truck vehicles can accommodate more passengers. The high “seat available” to alternative ratio reflects the crowding and discomfort associated with using these public transport modes.

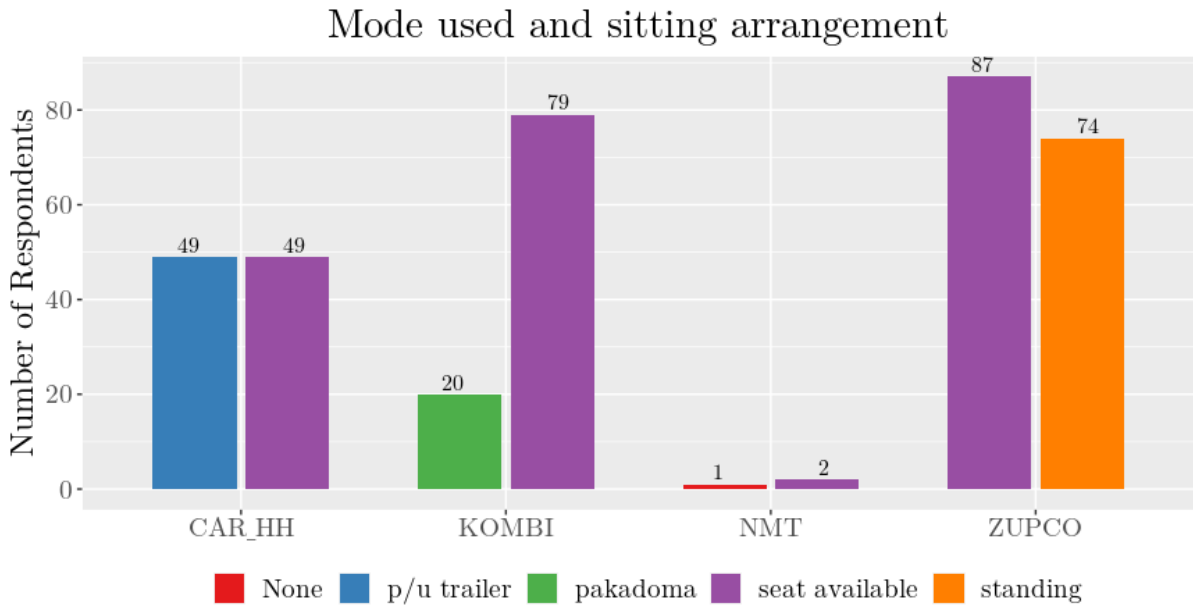


Figure 11: The modal split and seat arrangement in the previous trip.

Perceived variable responses

In the fourth section of the questionnaire, alternative specific perceived variables were presented to capture the general perception for each mode regarding safety & security and reliability. The respondents were asked how they felt about the safety and reliability on a 5-point Likert scale. The figure below shows how the respondents answered regarding the latent variables.

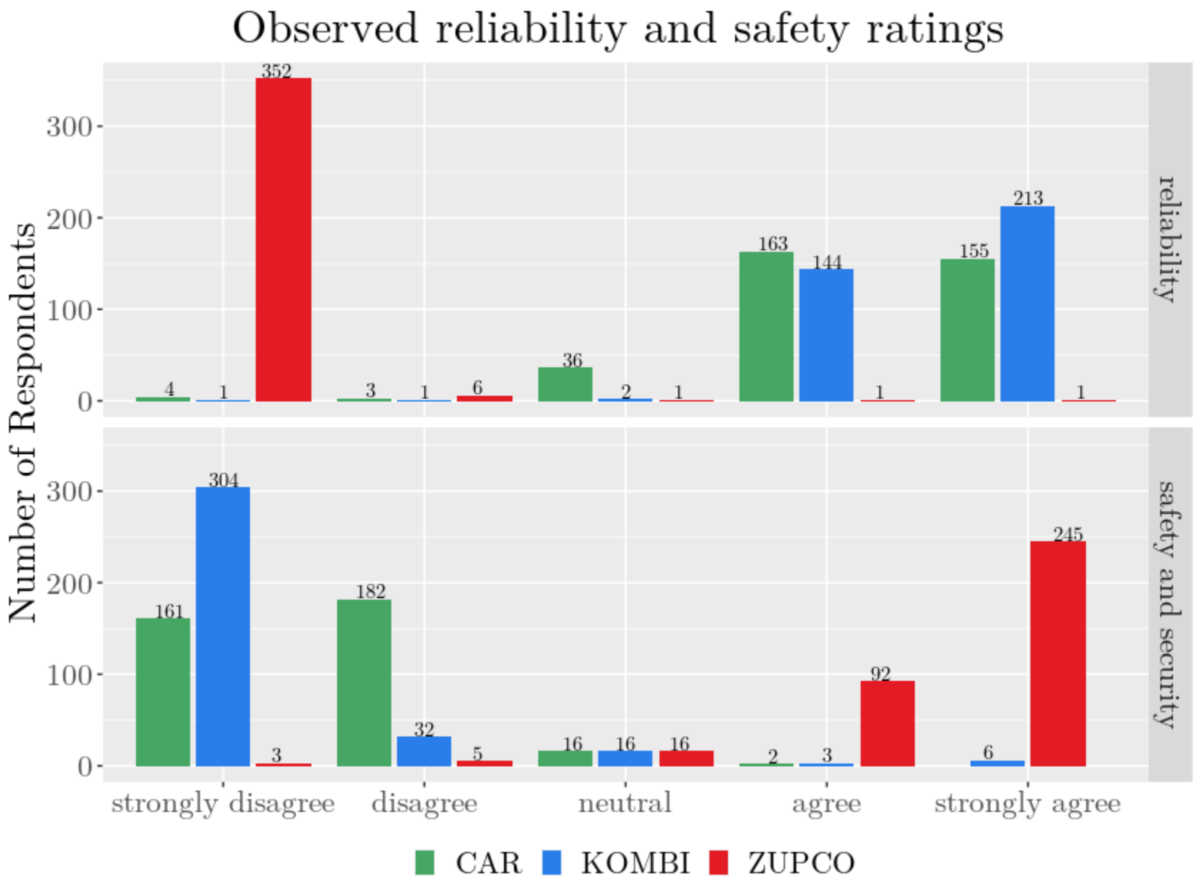


Figure 12: The distribution of individual responses to the psychometric questions.

It is apparent from the responses to the first three safety and security questions, "*I worry about losing my belongings each time I use* ***mode x***" that ZUPCO buses offer a poor service. 93% of the respondents about the security of buses, while only 2.2% disagree with the statement. Regarding car-hitchhike and kombis, at least 92% of the sampled population responded positively that they do not worry about the safety and security of the services; 4% of the respondents were however neutral across all the modes. The reliability and efficiency were found to be positive for kombis, and car hitchhike, while negative for ZUPCO buses. 99% of the respondents disagreed with the statement that "***Mode x*** *provides a reliable and efficient service*". On the contrary for car-hitchhike and kombis, 88% and 99% agreed with the statement, respectively. These results align with the focus group discussions on the state of public transport systems and supporting infrastructure in Harare.

4.0 DATA ANALYSIS: DISCRETE CHOICE MODELING

This section examines the observed choice behaviour as a function of the mode attributes, last trip attributes, and observed individual characteristics. The data is analysed using *apollo*, a customisable freeware package used to analyse discrete choices in the R computer software environment package (Hess & Palma, 2019, 2021). The survey dataset was coded into a *.csv* data file to be compatible with apollo. A summary of the dataset is shown in Appendix 3. Table 5 below shows that majority of the respondents chose kombi in most of the choice scenarios.

Table 5: Overview of choices for MNL model component.

Public transport mode	zupco	car hitchhike	kombi
Times available	2888	2888	2888
Times chosen	973	703	1212
Percentage chosen overall	33.69	24.34	41.97

The random utility theory framework is used to analyse the observed choices and provide the behavioural meaning to the estimated parameters. The model estimation is structured in three steps; at first, a generic multinomial logit model is estimated to investigate the influence of alternative attributes on choice behaviour, then interacted with covariates to explore deterministic taste heterogeneity. In the second and third, the mixed logit and latent class models are developed respectively to relax the IIA assumptions.

4.1 Multinomial Logit specification

As discussed in the literature review, multinomial logit choice models are easy to estimate and interpret due to the closed-form nature of the choice probabilities. The underlying assumptions imply that the observed choices are subject to the data that was collected in the questionnaire, and there is no correlation across respondents and public transport modes. The disaggregated choice probability of a respondent choosing kombi can be expressed mathematically as,

$$P(kombi) = \frac{e^{U_{kombi}}}{e^{U_{kombi}} + e^{U_{zupco}} + e^{U_{carHH}}} \tag{26}$$

The kombi is chosen when the choice probability in *equation* 26 above is greater than for zupco and car hitchhike. This is only satisfied when the utility for kombi as perceived by the respondent is greater than the other two modes. The utility for each alternative is specified as a function of seat arrangement, trip fare and the travel time components.

4.1.1 Attributes coefficients hypotheses

The expected signs of the marginal utility related to the respondent's sensitivity to the attribute are based on the economic theory.

Travel time components –time is one of the important mode choice predicting attributes, and the influence on the utility is expected to be negative. The perception of time is however dependent on the associated activities, and so is the level of impact on utility. Travel time was componentised into three parts to represent access, waiting, and in-vehicle travel. The first component measures the time taken to access public transport services, and most of the respondents walk to the bus stop to access the services. The waiting time component measures the time from arriving at the bus stop to the time of boarding a vehicle and is associated with the anxiety of waiting, especially since there is no bus timetable. However, some use that time to relax and catch up with friends. The in-vehicle travel time, as the name suggests, measures the duration of the trip from the origin to the destination bus stop in the vehicle, and the influence on utility may be contingent on the level of comfort in the vehicle (Kittelson & Associates et al., 2013).

Seat Arrangement – informs the respondent if an empty seat to occupy is available in the vehicle or not, which relates to comfort. The attribute is a qualitative measure, and four dummy variables were created in the questionnaire to represent all the possible sitting positions across the modes. Normalised at *seat available*, the sign is expected to be negative for "*kadoma*", "*trailer*" and "*standing*". This means assuming the expectation to find an empty seat available; an alternative seating arrangement would reduce the utility.

Fare – a negative sign is expected for cost as sensitivity is likely to increase with the trip fare.

4.1.2 Model 1: BASE MNL

The base model assumes homogeneity in taste preference across both the respondents and alternatives; that is to say, an hour in ZUPCO has the same influence on utility as kombi and hitchhiking. This model hypothesises that all the mode attributes influence choice decisions following the above expectations.

For each respondent, the utility for option n is given by the sum product of parameters and attributes as observed on the questionnaire, mathematically expressed as:

$$\begin{aligned} U_n = asc_n + \; & \beta_{AccessTime} * AccessTime_n + \beta_{WaitingTime} * \; WaitingTime_n \\ & + \sum_{x=1}^{4} \beta_{SittingArr_z} * \; dummy_z \begin{cases} dummy_z = 1, \forall \; SittingArr_n = z \\ dummy_z = 0, \qquad Otherwise \end{cases} \\ & + \beta_{InvehicleTime} * IVT_n + \beta_{Fare} * Fare_n \end{aligned} \tag{27}$$

Where;

asc_n represents the alternative specific coefficient for PT mode n

β_z represents the marginal utility or sensitivity to mode feature z to be estimated.

4.1.3 Model 2: MNL model with covariates

The base model assumes a constant sensitivity to the mode features across all respondents and alternatives, which is rarely the case in reality. Observed covariates can explain taste preference heterogeneity across the respondents, so interacting the sociodemographic and trip features with the generic model gives insights into how different covariate groups make preferential choices. For instance, two respondents from the low- and high-income groups are expected to have different sensitivities to trip fare; as such, the variables are specified as influencing attributes in the utility equation. In that regard, two hypotheses are established.

Hypothesis 1

Income group, gender, trip type, age, employment status, and suburb influence mode choices decisions. Therefore, the utility is specified by interacting the above variables with the fare, access time, and seating arrangement.

Hypothesis 2

The second hypothesis is that the model with covariates explains the observed choices much better than the generic model.

Taste Heterogeneity Specification

After estimating a number of interaction combinations, we observed that employment status and trip purpose were correlated, so we omitted the former variable; the age variable was found to have no sufficient influence on the sensitivity to attributes and was therefore excluded from the final utility equation. To capture fare heterogeneity, the fare component, seat arrangement and access time were modified as follows:

$\underline{Fare\ component\ modification}$

$$\beta_{Fare} = \beta_{Fare_{Base}} + \sum_{i=1}^{5} \beta_{Fare,IncomeGrp_i} * D_{IncomeGroup_i} + \sum_{i=1}^{5} \beta_{Fare,Taz_i} * D_{Taz_i} + \sum_{i=1}^{2} \beta_{Fare,Gender_i} * Gender_i + \sum_{i=1}^{3} \beta_{Fare,Trip_i} * D_{Trip_i} \quad 28$$

$\underline{Seat\ arrangement, Trailer\ modification}$

$$\beta_{Trailer} = \beta_{Trailer,Base} + \sum_{i=1}^{2} \beta_{Trailer,Gender_i} * Gender_i \quad 29$$

$\underline{Access\ Time\ modification}$

$$\beta_{AccessTime} = \beta_{AccessTime,Base} + \sum_{i=1}^{2} \beta_{AccessTime,Gender_i} * Gender_i \quad 30$$

Where:

$D_{y_i} = 1$ when the respondent falls under the sociodemographic group y_i and 0 otherwise

4.1.4 MNL estimation results, Model 1, and Model 2

The model's objective is to estimate β parameters, which gives the lowest log-likelihood. The results of the MNL models developed are summarised in Table 6 below:

Table 6: Model 1 and 2 parameter estimates.

Model name	MNL with no covariates	MNL with covariates
Parameters	9	20
Null LL	-3172.792	-3172.792
Final LL	-1714.785	-1541.902
ρ^2_{Adj}	0.4567	0.5077
AIC	3447.57	3123.8
BIC	3501.28	3243.17

	Est.	WTP (Z$)	t-ratio	Est.	t-ratio
asc_{zupco}	0.000		NA	0.000	NA
asc_{kombi}	1.354		15.721	1.441	14.145
$asc_{car,Hitch\ Hike}$	0.135		0.924	0.200	1.125
$\beta_{AccessTime}$	-0.173	Z$161.22	-16.022		
$\beta_{AccessTime\,Base}$				-0.212	-11.834
$\beta_{AccessTime\,Male}$				0.034	2.186
$\beta_{WaitingTime}$	-0.055	Z$51.61	-5.287	-0.059	-5.012
$\beta_{Available}$	0.000		NA		
$\beta_{Standing}$	-2.012	Z$31.24	-13.701	-2.154	-11.986
β_{Kadoma}	-1.585	Z$24.61	-15.037	-1.769	-12.823
$\beta_{Trailer}$	-2.158	Z$33.50	-12.271		
$\beta_{Trailer\,Base}$				-2.934	-9.400
$\beta_{Trailer\,Female}$				0.000	NA
$\beta_{Trailer\,Male}$				0.680	3.164
$\beta_{InvehicleTime}$	-0.027	Z$25.11	-9.053	-0.032	-7.436
β_{Fare}	-0.064		-25.561		
$\beta_{Fare\,Base}$				-0.072	-7.607
$\beta_{Fare\,NoIncome\&\$10k}$				0.000	NA
$\beta_{Fare\,\$10,000-\$20,000}$				0.037	3.620
$\beta_{Fare\,More\ than\ \$20,000}$				0.078	6.539
$\beta_{Fare\,WorkTrip}$				0.000	NA
$\beta_{Fare\,EducationTrip}$				0.031	3.491
$\beta_{Fare\,ShoppingTrip}$				0.036	2.880
$\beta_{Fare,Female}$				0.000	NA
$\beta_{Fare,Male}$				-0.018	-3.787
$\beta_{Fare\,Chitungwiza}$				0.000	NA
$\beta_{Fare\,Mabvuku}$				-0.023	-4.538

$\beta_{Fare_{Budiriro}}$		-0.029	-3.211
$\beta_{Fare_{Warren\ Park}}$		-0.018	-3.650
$\beta_{Fare,Dzivarasekwa}$		-0.015	-3.206

4.1.5 Hypothesis Testing

The utility function specifications for the first two MNL models above were based on three fundamental hypotheses. The first assumption was that individual behaviour aligns with the econometric theory. The MNL with covariates was specified on the hypothesis that the sociodemographic information could explain part of individual sensitivity to fare, access time, and "*trailer*". The third hypothesis was that the second model fits the observed choices better than the generic model.

We test the hypothesis that the parameters for the quantitative variables are less than zero against the null hypothesis that the parameters are not negative, and for the seat arrangement, we test the hypothesis that the parameters are not zero. The testing criterion for fare and time components is therefore a one-tailed test while two-tailed for seat arrangement, as shown in the figure below. Using the two-tailed rejection criterion for a variable with a known priori sign would reject the extreme expected behaviour (Hess et al., 2020).

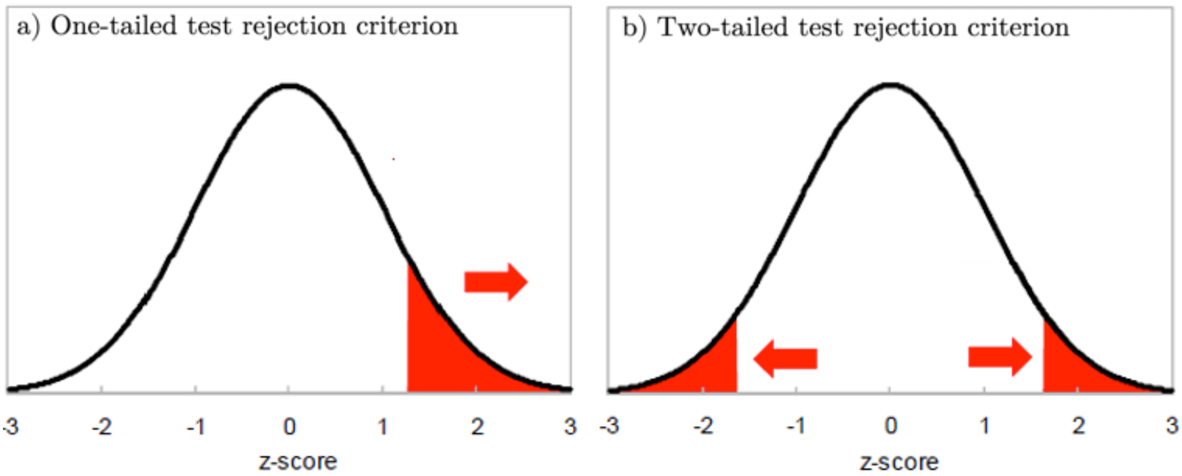


Figure 13: Hypothesis rejection criteria.

Generally, all parameter estimates are significantly consistent with the econometric theory of utility maximisation, implying that respondents are substantially sensitive to all the variables and prefer alternatives with lower costs, travel time, and comfortable seating arrangement, at more than 95% confidence level. The alternative specific constant estimates reveal strong cognitive inertia towards kombis and a relatively weak aversion to ZUPCO buses, ceteris paribus. However, there is not enough evidence to tear up the preference choice between ZUPCO buses and cars. The models suggest that the car is only preferred at not more than 62.6% and 73.9% confidence levels, respectively.

Seat arrangement is coded as a dummy variable fixed at "*seat available*", which is standard across all modes. The substantial difference between the "*seat available*" and the other alternatives provides sufficient evidence at a 100% level of confidence that the variable influences choice decisions, including "*seat available*" as a level. The marginal utility of seating in a car trailer ($\beta_{Trailer}$) is the highest, followed by being a standing passenger in ZUPCO ($\beta_{Standing}$), then sitting "*pakadoma*" in the kombi (β_{Kadoma}). Behaviourally, when all seats are occupied, especially during peak periods, people prefer using kombis to buses or pick-up trucks. Nonetheless, by using the delta method, we observe that there is not sufficient statistical evidence to suggest a difference in sensitivity between standing in ZUPCO and seating in a pick-up truck trailer (49.2% level of confidence).

Looking at the travel time components, the first observation that can be made is the difference in sensitivities across the components, which is significant at not less than 99.3% confidence level. Walking time to a bus stop is more than six times as painful as in-vehicle travel time, while the waiting time is just twice as much. The difference reflects the inherent discomfort of utilising physical effort to walk and the anxiety that comes with waiting at a bus stop, further reinforcing the notion of travel being a derived demand and time being a subjective measure.

4.1.6 Covariate's interaction

In the modified MNL model, we explore the systematic taste variations to fare, access time, and seat arrangement across different sociodemographic groups to give more insights into the respondents' travel behaviour to trip fare. Just like with the generic MNL model, the econometric theory holds at not less than 95% level of confidence. The delta method and the two-tailed rejection criterion are used to test the second hypothesis.

Gender

Gender interacted with access time, fare, and seating arrangement: "*trailer*" as a dummy variable fixed at the female, so we test the hypothesis that men and women behave differently when choosing public transport modes against the hypothesis that they have the same sensitivity to the variables. The results provide sufficient evidence that there is a deterministic variation in taste preference that can be explained by gender. The MNL with covariates model shows that women are significantly more sensitive to walking ($\beta_{AccessTime_{Male}}$) and seating in a pick-up "trailer" than their male counterparts, ($\beta_{Trailer_{Male}}$). However, with regards to fare, men seem to be more sensitive than women, ($\beta_{Fare_{Male}}$).

Trip purpose

Trip purpose data was collected as revealed data on respondents' most frequent trips with four levels: home-based work-trip, education-trip, leisure and shopping, and others. The data was later cleaned to the first three variables. The variable was interacted with fare to investigate the sensitivity variation across trip types, and the utility specification fixed the trip purpose variable at work-trip. Trip fare appears to be most painful for work trips and lowest for shopping trips which behaviourally mean that respondents who frequent work trips are more sensitive to fare than the other respondents. Fare heterogeneity is quite substantial between work trips and the other two, and less significant between shopping and educational trips at not more than 41.4% confidence level.

Income group

No income and less than Z\$10,000 were grouped together because their sensitivity to fare was found to be statistically not different, the same for Z\$20,000 and more than \$30,000. By fixing the covariate at $\beta_{Fare_{NoIncome\&\$10k}}$, the sensitivity to trip fare appears to be lowest for people with a monthly income of not less than Z\$20,000 (*which is equivalent to US*\$125) and highest for people with a monthly income of less than Z\$10,000 and the "*no income*".

Transportation Analysis Zone

Regarding the place of origin, respondents from Chitungwiza are substantially less sensitive to fare than the other four suburbs, which may partly be attributed to its administration level and the distance from town. Chitungwiza is a dormitory town of Harare, and as shown in Table 2, the town is located about 25.3km from the city centre, while the other suburbs are less than 19km from the city centre. The most fare sensitive respondents originate from Budiriro, followed by Mabvuku, Warren Park, and then lastly, Dzivarasekwa. The variation of fare sensitivity across the four suburbs is relatively small compared to Chitungwiza and statistically significant at very low levels of confidence: for instance, Dzivarasekwa and Warren Park at 88.5%, Warren Park and Mabvuku, at 62.7% and, Budiriro and Mabvuku, at 48.2%. The confidence levels are too low to lend significant weight to any comparisons across the suburbs.

4.1.7 Goodness Fit test

The art of utility specification is to find a functional form such that the parameter estimates would minimise the absolute value of the log-likelihood. To compare the goodness of fit for models above, the log-likelihood ratio test, AIC, BIC, and the McFadden adjusted rho squared values can be compared. The two MNL models are nested; as such, the log-likelihood ratio test method can compare goodness fit. Using the log-likelihood ratio test, we can conclude with more than a 95% level of confidence that the second model outperforms the base model with generic coefficients. The log-likelihood improved by 173 units from the generic model at the expense of 11 more parameters, giving a log-likelihood ratio test value is 345.3 against the critical log-likelihood value of 95%, $\chi^2_{(11)} = 19.6$. This is highly significant since the test value is greater than the critical value. In support of the log-likelihood ratio test findings, the decrease in AIC and BIC by 324 and 258 units respectively and increase in the adjusted Rho square by 0.051 units respectively show a massive improvement in the explanatory power. We can therefore reject the hypothesis that the MNL base model is sufficient.

4.1.8 Willingness to Pay

The willingness to pay for the generic MNL can be easily calculated by finding the ratio of the marginal utility estimates to the fare estimate. For the MNL model with covariates, the mean willingness to pay is calculated from everyone's desire to spend. The table below shows the willingness to pay indicators for the two MNL models. The MNL model with covariates suggests that respondents are willing to pay Z$179, Z$54, and Z$29 for an hour of service improvement in access time, waiting time, and in-vehicle travel time, respectively. In addition, the respondents are willing to pay an additional of between Z$27 and Z$39 to get a seat. While the generic model simplifies the willingness to pay evaluation, the values are underestimated by up to 16%.

Table 7: The willingness to pay indicators for MNL models

Attribute	WTP Estimates Model 1	WTP Estimates Model 2
Access time	Z$161.22	Z$178.92
Waiting time	Z$51.61	Z$54.37
In-vehicle time	Z$25.11	Z$29.80
Seat available	(reference/ standard)	(reference/ standard)
Kombi: "*kadoma*"	Z$24.61	Z$27.11
ZUPCO: "*standing*"	Z$31.24	Z$33.01
Pick-up: "*trailer*"	Z$33.50	Z$39.70

WTP across suburbs and gender

The table below presents the willingness to pay indicators across gender and suburbs.

Table 8: The mean willingness to pay indicators across gender and suburbs

Attribute	Access	waiting	in vehicle	"kadoma"	standing	trailer
Chitungwiza	Z$234.82	Z$71.79	Z$39.34	Z$35.80	Z$43.59	Z$51.95
Budiriro	Z$144.42	Z$44.51	Z$24.40	Z$22.20	Z$27.03	Z$31.82
Dzivarasekwa	Z$187.90	Z$55.81	Z$30.59	Z$27.83	Z$33.89	Z$42.18
Warren Park	Z$174.75	Z$53.07	Z$29.08	Z$26.46	Z$32.22	Z$38.79
Mabvuku	Z$151.87	Z$46.40	Z$25.43	Z$23.14	Z$28.18	Z$33.61
Male	Z$144.83	Z$48.19	Z$26.41	Z$24.03	Z$29.26	Z$30.62
Female	Z$224.22	Z$62.58	Z$34.29	Z$31.20	Z$38.00	Z$51.77

The willingness to pay indicators are lower across all the indicators for men compared to women. Interesting to note is how women are willing to pay Z$13.77 (an additional 27%) for a seat upgrade from sitting in a pick-up *trailer* to *standing*, while men are only willing to pay Z$1.36 (extra of 4%) for the same upgrade. This reflects on female respondents' strong aversive behaviour towards pick-up *trailers*. Across the suburbs, the table shows vividly that Chitungwiza has the highest willingness to pay for all the attributes and Budiriro respondents have the lowest.

4.2 Mixed multinomial logit models

While using sociodemographic information to explain sensitivity differences across people gives more insights into their choice behaviour, the approach only captures the systematic taste variation across the observed variables, neglecting additional stochastic variation, which cannot be explained deterministically by the empirical data. The mixed multinomial logit assumes that there's not just a single fixed standard coefficient for the entire population or sociodemographic group but instead distributed continuously across the population (Train, 2009). The class of models captures random taste variation by making use of different prespecified parametric continuous mixing distributions. The choice of the continuous mixing distribution plays a crucial role in the estimation, and a body of literature has raised concerns over the behavioural meaning of some distributions (Hess, 2005). For instance, assuming a normal distribution of parameters often yields counterintuitive behaviour when used to specify parameters with an expected sign like trip fare and the travel time components. On the other hand, the lognormal distribution is restricted to take the expected sign, but the long tails and very slow convergence thereof exaggerate the sensitivity. Furthermore, when these continuous distributions are used to specify a cost coefficient, the values that are close to zero may result in an enormous willingness to pay. The non-existence of the willingness to pay moments has implications on policy appraisal. Similar challenges arise with uniform distribution specification of attributes, and even so, the distributions are widely in practice across several sectors. (Scarpa, Thiene & Train, 2008; Daly, Hess & Train, 2012) suggests that parametrising the model in the willingness to pay space can ensure finite WTP moments.

$$U_n = \beta_{fare} * Fare_n + \sum_{\forall attr} WTP_{attr} * x_{attr,n} \quad 31$$

Where;

β_{fare} represents a density function, $f(\beta|\Omega)$ of a statistical random distribution of fare parametres across the population, and Ω represents the type of distribution.

$$WTP_{attr} = \frac{\beta_{attr}}{\beta_{fare}}$$

4.2.1 The specification of random coefficients

The generic MNL utility specification in *equation* 27 is modified to accommodate random heterogeneity with regards to the parameters across the population to develop four mixed logit models in Table 9.

Table 9: MMNL models and parameter distribution across the population.

MMNL Model	*Lognormal*	*Loguniform*	*Normal*	*Uniform*
$\beta_{access\ Time}$	$\beta = -e^{\mu+\sigma*r_n}$	$\beta = -e^{a+b*r_u}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
$\beta_{waiting\ Time}$	$\beta = -e^{\mu+\sigma*r_n}$	$\beta = -e^{a+b*r_u}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
$\beta_{Standing}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
β_{Kadoma}	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
$\beta_{Trailer}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
$\beta_{InvehicleTime}$	$\beta = -e^{\mu+\sigma*r_n}$	$\beta = -e^{a+b*r_u}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$
β_{Fare}	$\beta = -e^{\mu+\sigma*r_n}$	$\beta = -e^{a+b*r_u}$	$\beta = a + b * r_u$	$\beta = a + b * r_u$

Where:

$r_n \sim N(0,1)$

$r_u \sim U(0,1)$

The estimated parameters for the MMNL models are summarised in the table below.

Table 10: The estimation results for MMNL models

Model name	MMNL neg. Lognormal	MMNL neg. Log uniform	MMNL Uniform	MMNL Normal
Parameters	16	16	16	16
Null LL	-3172.792	-3172.792	-3172.792	-3172.792
Final LL	-1541.772	-1544.171	-1538.331	-1540.26
Adj.Rho-square	0.5090	0.5083	0.5123	0.5095
AIC	3115.54	3120.34	3108.66	3112.52
BIC	3211.04	3215.84	3204.15	3208.01

	$\mu(\sigma)$	$bound(spread)$	$bound(spread)$	$\mu(\sigma)$
asc_{kombi} $est.$	1.77 (0)	1.76 (0)	1.86 (0)	1.87 (0)
$t-ratio$	12.14 (0)	12.42 (0)	11.41 (0)	11.64 (0)
asc_{car} est	0.2 (0)	0.18 (0)	0.24 (0)	0.4 (0)
$t-ratio$	0.79 (0)	0.69 (0)	0.94 (0)	1.65 (0)
$\beta_{accessTime}$ est	-1.42 (-0.21)	-1.08 (-0.7)	-0.17 (-0.17)	-0.27 (-0.03)
$t-ratio$	-14.86 (-2.82)	-5.55 (-1.79)	-2.91 (-1.34)	-11.4 (-0.66)
$\beta_{waitingTime}$ est	-2.91 (0.81)	-1.13 (-4.39)	0.13 (-0.41)	-0.08 (0.1)
$t-ratio$	-9.87 (4.18)	-5.09 (-3.53)	2.63 (-4.55)	-5.23 (2.86)
$\beta_{available}$ est	0 (0)	0 (0)	0 (0)	0 (0)
$\beta_{standing}$ est	-2.96 (1.29)	-4.89 (4.02)	-1.18 (-3.61)	-3.05 (1.37)
$t-ratio$	-11.03 (5.48)	-8.09 (4.52)	-2.01 (-2.71)	-10.88 (5.7)
β_{kadoma} est	-2.36 (-0.32)	-2.79 (0.93)	-1.53 (-1.89)	-2.48 (-0.77)
$t-ratio$	-11.81 (-0.81)	-3.5 (0.62)	-3.53 (-2.07)	-11.81 (-3.13)
$\beta_{trailer}$ est	-3.91 (-1.95)	-7.42 (7.05)	-0.5 (-6.97)	-4 (1.84)
$t-ratio$	-8.81 (-10.22)	-10.71 (9.31)	-1.12 (-9.97)	-9.19 (8.38)
$\beta_{invehicleTime}$ est	-3.22 (0.45)	-2.43 (-1.62)	-0.09 (0.09)	-0.05 (-0.02)
$t-ratio$	-18.4 (4.74)	-16.16 (-3.46)	-8.49 (5.58)	-7.94 (-4.95)
β_{fare} est	-2.34 (0.78)	-3.59 (2.4)	-0.01 (-0.22)	-0.11 (0.06)
$t-ratio$	-27.34 (11.25)	-26.09 (13.15)	-0.98 (-11.29)	-13.32 (12.32)

4.2.2 Model fit comparison

The table above shows estimates for the four MMNL models, with μ, σ, a, and b defining the continuous distribution parameters for the mode attributes and the alternative specific constants, the mode-specific coefficients referenced at ZUPCO buses. Firstly, the mixed logit shows a slight improvement in model fit from the multinomial logit with covariates. The MMNL log uniform has the least fit lost in log-likelihood from -1541.902 by 0.15% at the expense of seven fewer parameters yielding gain on the adjusted rho square, BAS, and AIC indices. This implies that the mixed logit models fit the choice observations better than the MNL models, reinforcing the random distribution hypothesis. Now comparing the mixed logit models, the unrestricted random distributions clearly outperform the restricted log distributions, with the uniform distribution surpassing all the distributions in model fit. To that effect, the Ben-Akiva & Swait test provides sufficient evidence at more than 97% level of confidence (p-value for Ben-Akiva & Swait test is 0.02551).

4.2.3 Estimation results

The first observation that can be made from Table 10 regarding the estimates is the general statistical efficacy of the parameters and their contribution to the observed choices. The models indicate a strong preference towards kombis with an insignificant difference between hitchhiking cars and ZUPCO buses. All the variables contribute to the observed choice decisions; however, only the negative lognormal model suggests a random sensitivity variation towards access time at not less than 95% level of confidence (92.6% for negative log uniform, 81.8% for Uniform, and 49% for normal distribution). In addition, the log-transformed models indicate that "*kadoma*" variation estimates are only significant at not more than 58.3% and the uniform distribution provides not more than 86.9% evidence that the bound parameters for trip *fare* and *trailer* are not zero. Behaviourally this means that, despite the uniform variation in sensitivity with regards to β_{fare} *and* $\beta_{trailer}$ as shown by the significant spread components estimates, there exist individuals in the population who are not sensitive to *fare* and some who rate sitting in a pick-up trailer and on the seat the same way.

Notwithstanding the improvements in model fit, the free distributions exhibit counterintuitive behaviour for some attributes. For instance, the normal distribution suggests the presence of individuals with counterintuitive behaviour with regards to fare, in-vehicle travel time, and waiting time, while the uniform distribution only displays such for access time. In the context of trip fare, this represents a situation where respondents would choose a more expensive mode, with all the variables being equal. The normal distribution indicates a 3.34% non-zero probability of a positive fare coefficient. In the same way, a positive waiting time coefficient value represents a situation where a respondent chooses a mode with longer headways, ceteris paribus and while this might reflect on their behaviour for leisure trips or shopping, the observed 31.5% (uniform distribution) and 20.9% (normal distribution) non-zero probability of positive coefficient or waiting time appears to be too high to be neglected. Regarding the in-vehicle travel time, the normal distribution suggests a 1.75% probability of an individual choosing a slower mode. All these behavioural inconsistencies could lead to misleading policy directives.

The log-transform random distributions, on the other hand, restrict the counterintuitive choice behaviour while exaggerating the sensitivity to variables; this is supported by the long tails for negative log attributes, as shown in Figure 14 below. The model fit choice criterion should therefore be based on both models' fit and the behavioural connotation of the estimates.

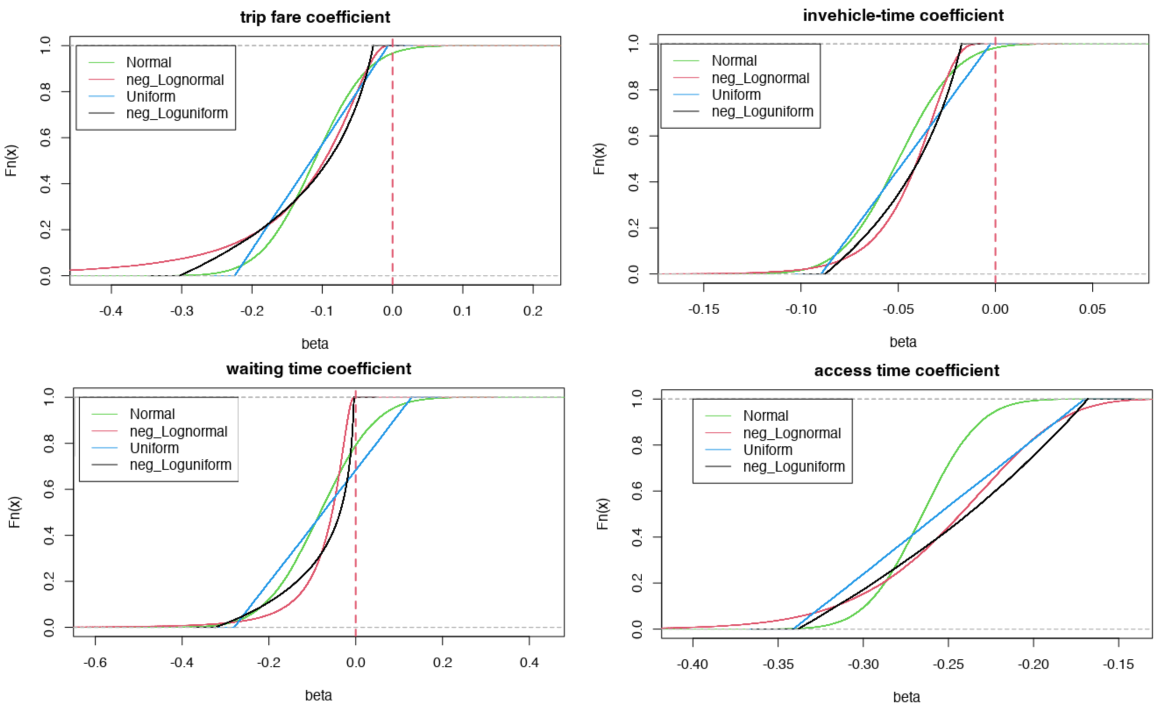


Figure 14: The implied distributions for the travel time components and trip fare.

4.3 Latent class models

The latent class modelling approach captures both systematic and random taste heterogeneity by dividing the population into a finite number of heterogeneous groups, distinguished by different parametric vectors, with some possibly constant across classes. The number of classes is determined based on a priori hypothesis, and the class allocation is probabilistic, evaluated for each class as a function of the socioeconomic features. The probability of falling in every class is considered to be non-zero to relax the specification while allowing random heterogeneity in taste preferences.

4.3.1 Model specification and estimation

Assuming S latent classes, defined as a function of income group, age, TAZ, gender, and employment status, to be sufficient to account for preference heterogeneity in this model, the class allocation utility function for individual n and class s can be expressed as:

$$U_{n,s} = \delta_s + \sum_{y=1}^{5} \gamma_{age_y} * D_{age} + \sum_{y=1}^{5} \gamma_{inc} * D_{inc} + \sum_{y=1}^{5} \gamma_{TAZ} * D_{TAZ} + \sum_{y=1}^{4} \gamma_{employment} * D_{employment} + \gamma_{Male} * gender \qquad 32$$

where δ_s represents the class specific constant for class, s.

The results of the estimation are summarised in Table 11 below.

Table 11:Estimatiton results for the latent class model

Model name	Latent Class -MNL Model		Latent Class – MMNL Model	
Parameters	27		42	
Null LL	-3172.792		-3172.792	
Final LL	-1501.254		-1435.684	
ρ^2_{Adj}	0.5183		0.5343	
AIC	3056.51		2955.37	
BIC	3217.65		3206.04	
Class	**class 1**	**class 2**	**class 1**	**class 2**

	Est.	t-ratio	Est	t-ratio	Est.	t-ratio	Est	t-ratio
asc_{zupco}	0.000	NA	0.000	NA	0.000	NA	0.000	NA
asc_{kombi}	1.898	8.684	1.351	10.350	2.624	2.921	1.516	6.572
$asc_{car,Hitch\ Hike}$	0.815	2.161	-0.107	-0.439	1.286	0.559	-0.084	-0.273
$\beta_{AccessTime}(\mu)$	-0.261	-13.03	-0.170	-9.025	-0.379	-5.498	-0.186	-8.285
$\beta_{WaitingTime}(\mu)$	-0.149	-3.840	-0.038	-0.439	-0.203	-1.737	-0.041	-2.409
$\beta_{Available}$	0.000	NA	0.000	NA	0.000	NA	0.000	NA
$\beta_{Standing}(\mu)$	-2.984	-9.438	-1.968	-9.025	-4.841	-4.291	-2.137	-5.129
$\beta_{Kadoma}(\mu)$	-1.902	-8.686	-2.116	-2.740	-2.956	-5.186	-2.409	-11.75
$\beta_{Trailer}(\mu)$	-4.017	-8.587	-2.009	-8.459	-6.186	-1.831	-2.621	-6.144
$\beta_{InvehicleTime}(\mu)$	-0.050	-6.507	-0.028	-11.79	-0.068	-2.609	-0.031	-5.426
$\beta_{Fare}(\mu)$	-0.139	-11.94	-0.035	-5.297	-2.956	-2.778	-0.038	-4.625
$delta_{\mu}$	0.000	NA	0.532	-5.236	0.000	NA	0.570	-4.625
$\beta_{AccessTime}(\sigma)$	0.000	NA	0.000	NA	0.054	0.428	0.014	0.217
$\beta_{WaitingTime}(\sigma)$	0.000	NA	0.000	NA	-0.176	-2.291	-0.035	-0.487
$\beta_{Standing}(\sigma)$	0.000	NA	0.000	NA	-1.029	-0.991	0.515	0.795
$\beta_{Kadoma}(\sigma)$	0.000	NA	0.000	NA	0.986	1.557	-0.057	-0.136
$\beta_{Trailer}(\sigma)$	0.000	NA	0.000	NA	-2.787	-2.269	-1.452	-5.691
$\beta_{InvehicleTime}(\sigma)$	0.000	NA	0.000	NA	-0.007	-0.644	-0.023	-1.969
$\beta_{Fare}(\sigma)$	0.000	NA	0.000	NA	-0.022	-0.688	-0.013	-2.477
$delta_{\sigma}$	0.000	NA	0.000	NA	0.000	0.000	0.046	0.183
γ_{Male}	1.404	3.583	0.000	NA	1.438	1.757	0.000	NA
$\gamma_{NoInc\$10K}$	0.000	NA	0.000	NA	0.000	NA	0.000	NA
$\gamma_{\$10,000-\$20,000}$	-4.484	-4.198	0.000	NA	-4.840	-3.025	0.000	NA
$\gamma_{More\ than\ \$20,000}$	-7.451	-5.185	0.000	NA	-8.129	-2.821	0.000	NA
$\gamma_{Not\ Employed}$	0.000	NA	0.000	NA	0.000	NA	0.000	NA
$\gamma_{Employed}$	4.492	3.900	0.000	NA	4.902	2.117	0.000	NA
$\gamma_{Chitungwiza}$	0.000	NA	0.000	NA	0.000	NA	0.000	NA
$\gamma_{Mabvuku}$	1.704	2.988	0.000	NA	1.783	2.217	0.000	NA
$\gamma_{Budiriro}$	3.761	4.280	0.000	NA	4.227	2.962	0.000	NA

$\gamma_{Warren\ Park}$	1.471	2.297	0.000	NA	1.819	1.419	0.000	NA
$\gamma_{Dzivarasekwa}$	0.868	1.513	0.000	NA	0.902	0.972	0.000	NA
Probability	56.8%		43.2%		57.3%		42.7%	
Null LL	-3172.8		-3172.8		-3172.8		-3172.8	
Final LL	-2245.9		-1989.9		-2405.1		-1913.0	

4.3.2 Model fit discussion.

The first observation to note is a general improvement in model fit from MMNL models to their latent class model counterparts, suggesting the presence of additional random taste heterogeneity. The LC-MNL gained log-likelihood from the lognormal distribution by 40.518 units at the expense of additional 11 more parameters, leading to an overall improvement by 59.03 and 0.0093 units on the AIC and adjusted rho square indices, respectively, despite losing by 6.61 BIC units[6]. In the same way, for the latent class with MMNL, gain in model fit by 0.02, 153.3, and 1.89 units on the adjusted rho square, AIC, and BIC scales, respectively, was acquired. Looking at the latent class specified models, we observe a benefit of 66 log-likelihood units at the expense of 15 more parameters, giving an infinitesimal log-likelihood ratio test $p-value$, and suggesting the presence of significant randomly distributed taste heterogeneity within classes.

4.3.3 Latent class estimation results[7].

LC Generic

All the sociodemographic attributes contribute to class allocation substantially except for age. The age groups' contribution to class membership allocation probability is only significant at not more than 45.1% confidence level and was therefore removed from the class allocation utility specification.

[6] The BIC tends to impose stricter criteria for including additional parameters with larger sample sizes, to counter bigger changes in log-likelihood associated with large sample sizes (Hess et al., 2020).

[7] Classes with same names are not necessarily equivalent, so we cannot compare the class one in LC generic to class one in LC MMNL directly.

The employment and income covariates were condensed to form two and three subgroups, respectively, and the estimates suggest a sufficient contribution to the class allocation probability. The *class one* alternative specific constants reveal a strong aversion towards ZUPCO at not less than 96.9% confidence level, while for *class two*, the dislike is towards the car at not more than 33.5%. Nonetheless, the two classes show a statistically significant preference towards kombis.

All the parameter estimates appear to be substantially significant with the expected negative sign, substantiating the hypothesis of similar choice behaviour within classes. Important differences arise between *class one* and *class two* in terms of the sensitivity estimates. The *class one* members are generally more sensitive to all attributes as compared to *class two* except for the seating arrangement. In relation to travel time components, all the respondents find walking time to be five times more painful than in-vehicle travel time, but the waiting time is three times more painful for *class one* members while comparable to in-vehicle time in *class two*.

The trade-off between access and waiting times is higher in *class two* by 250%, indicating a strong aversion to long walking times in *class two* as compared to *class one*. As many behavioural researchers have asserted, evaluating willingness to pay indicators is more informative and important for policymaking (Breidert, Hahsler & Reutterer, 2006; Eboli & Mazzulla, 2008; Miller et al., 2011; Schmidt, Tammo & Bijmolt, 2019). The table below summarises the willingness to pay indicator estimates indicating the value of time savings in $Z/hour and willingness to pay for a “seat”.

Table 12: Generic latent class model implied willingness to pay indicators.

Attribute	WTP estimate for class one	WTP estimate for class two
Access time	Z$112.94	Z$291.88
Waiting time	Z$64.50	Z$65.10
In-vehicle time	Z$21.70	Z$47.24
Seat available	(reference)	(reference)
ZUPCO: Standing	Z$21.50	Z$56.19
Kombi: Kadoma	Z$13.71	Z$60.41
Car: Trailer	Z$28.95	Z$57.37

The implied willingness to pay scale shows a clear distinction between the two classes. The willingness to pay is generally higher by more than 100% for all the variables in *class two*, except for waiting time, which is equal to the nearest Z$1.00 across the two classes. The implied willingness to pay for seat upgrades is approximately Z$60.00 across all the modes in *class two*, while it varies for *class one* members. Behaviourally, this means *class two* members are reluctant to compromise on getting a seat, whereas *class one* would consider seating on "*kadoma*" at a relatively lower price.

LC-MMNL.

Like in the LC-MNL model, the class allocation is based on gender, income, employment status, and suburb. The mean estimates of the coefficients are all statistically significant at not less than 95.9% (waiting time, *class* 1) level of confidence. However, there is not enough evidence to substantiate the statistical efficacy of the standard deviations for some variables. In *class* 1, the standard deviation is significant at more than 95% confidence level for *trailer* and *waiting time*, while the remaining variables are only substantial at not more than 88%. Similarity for *class* 2, the variation parameters for "*trailer*", $in-vehicle\ travel\ time$, and *fare* variation parameters are significantly different from zero at not less than 95.1% confidence interval and the rest only significant at not more than 57.4%. While the substantial variation indicates statistical merit in the exploitation of the residual heterogeneity within classes, the non-zero probabilities of 12.4% and 1.3% for waiting time and "*trailer*" respectively in *class* 1 and 3.6%, 9.3% and 0.2% for "trailer", in-vehicle travel time and trip fare in *class* 2 cannot be neglected.

4.4 Summary of the results

In the above section, the mode choice behaviour for the low-income population in Harare was assessed to investigate causal relationships between the observed variables and choices. To that effect, the hypothetical mode choice dataset collected from five suburbs in Harare was analysed using multinomial, mixed multinomial logit, and a latent class of models derived from the random utility maximisation framework. All the specified mode variables were consistently significant across the models, and the interaction with sociodemographic covariates provided more insights into the mode choice behaviour. The estimated coefficients are broadly consistent with (Venter, 2016)'s findings for the city of Johannesburg (COJ), showing significant negative impacts of increases in walking time, waiting time, in-vehicle travel time and travel cost. Model fit was examined by comparing the adjusted likelihood ratio index, log-likelihood, AIC, and BIC values, and particular attention was paid to possible counterintuitive behaviour. The goodness fit test reveals a gradual gain in model fit from the MNL class of models to the latent class, indicating an improvement in explanatory power as deterministic and random heterogeneity across classes is accounted for. The LC-MNL outclasses the LC-MMNL based on non-compliance of the results with the microeconomic theory; as such, the LC-MNL is considered the most appropriate model for the data.

The latent class multinomial logit model suggests that the population can be divided into two latent classes with similar mode choice behaviour; the allocation is estimated probabilistically as a function of gender, income, employment status, and suburb. We see that all the sociodemographic variables increase the probability of falling in *class one* except for respondents with a monthly income greater than Z$10,000 who are more likely to fall in *class two*, the opposite is true for *class two*. The results are used to compare the willingness to pay indicators to those calculated from the base MNL and the results for PT captives obtained from the City of Johannesburg study. Noteworthy is the substantial differences in willingness to pay indicators between the base MNL model and the weighted averages for the latent class model, highlighting the importance of accounting for random heterogeneity. The COJ study was conducted in 2016, and to account for inflation, values of timesaving were raised using the Consumer Price Index (CPI[8]) for public transport from September 2016 to September 2021. The value of walking time to start a trip from the study is 15% lower than the latent class weighted average, and for waiting and in-vehicles time savings, 230% and 71% higher, respectively. While the value of walking time for COJ data lies between the *class one – class two* range, the value of waiting time seems to be higher than walking time, which is behaviourally unusual.

---

[8]Consumer price index for PT: 2016/17=2.3%, 2017/18=7.1%, 2018/19=3.0%, 2019/20=1.6%, 2020/21=6.7%

Table 13: A comparison with previous studies' willingness to pay indicators.

Model	MNL	COJ[9]	Class 1	Class 2	Weighted Average
Access Time	Z$161.22	Z$161.68	Z$112.94	Z$291.88	Z$190.28
Waiting Time	Z$51.61	Z$218.97	Z$64.48	Z$65.11	Z$64.75
In-vehicle Time	Z$25.11	Z$56.11	Z$21.70	Z$47.24	Z$32.74
Standing	Z$31.24	-	Z$21.50	Z$56.19	Z$36.49
Kadoma	Z$24.61	-	Z$13.71	Z$60.41	Z$33.90
Trailer	Z$33.50	-	Z$28.95	Z$57.37	Z$41.23

The class allocation utilities are specified as a function of the sociodemographic, and while the above willingness to pay indicators provide a summary of mode choice behaviour, the interaction of the model results with sociodemographic covariates helps to review classification diagnostics. The class allocation parameters examine the contribution of the covariates to the non-zero membership probabilities for different classes. Posterior analysis can be used to analyse choice decisions and therefore evaluate preference at the individual level (Hess, 2014). We use data from the five suburbs and income to identify posterior class features. Conditional on the sociodemographic features and observed choices, the most probable class for each respondent can be identified. The average probabilities of the class allocation model accurately predicting classification can then be used to describe the composition of the latent classes as summarised in Table 13 below.

More than half of respondents from Budiriro and Mabvuku are likely to be allocated to *class one*, and the majority from Chitungwiza in *class two*. Warren Park and Chitungwiza residents are equally distributed between the two latent classes. 65% of the respondents classified in *class one* are male, and *class two* consist of 53% female respondents. Huge differences arise with the income group covariate; a minimum of 71% of respondents with No income and less than Z$20,000 monthly income is more likely to be classified in *class one*, and a minimum of 67% with more than Z$20,000 in *class two*.

[9] 1 rand was equivalent to Z$10.67 at the time of survey.

Table 14: Posterior class allocation by income, gender, and suburb

	Class	C1	C2
Suburb	Chitungwiza	7.50%	12.72%
	Mabvuku	12.46%	7.48%
	Warren Park	10.60%	9.34%
	Budiriro	16.90%	3.04%
	Dzivarasekwa	9.50%	10.45%
Gender	Male	36.92%	20.14%
	Female	20.04%	22.90%
Income	No income	9.18%	3.00%
	Z$0 – Z$10,000	14.63%	3.10%
	Z$10,000 – Z$20,000	18.94%	7.65%
	Z$20,000 – Z$30,000	12.76%	26.02%
	More than Z$30,000	1.45%	3.26%

Using the weights in Table 14, we calculate the willingness to pay for mode attributes for each suburb as a function of the specific posterior latent class allocation probabilities thereof. As indicated in the table below, the results highlight some variation in taste preference on the willingness to pay scale across all the suburbs, with a constant waiting time value of Z$65.00. The population group from Budiriro has the least value of time savings perception across all the attributes, and they are willing to be compensated the least for a downgrade from "*seat available*" to alternative seating arrangements. Chitungwiza, on the other hand, scores the highest willingness to pay values for all the attributes. Respondents from Warren Park and Dzivarasekwa exhibit similar sensitivities across all the attributes.

Table 15: Posterior willingness to pay indicators across the suburbs.

Attribute	Access	Waiting	In-vehicle	Kadoma	Trailer	Standing
Warren Park	Z$196.96	Z$64.74	Z$33.69	Z$35.64	Z$42.29	Z$37.79
Mabvuku	Z$180.23	Z$64.67	Z$31.29	Z$31.27	Z$39.63	Z$34.55
Dzivarasekwa	Z$206.86	Z$64.78	Z$35.10	Z$38.22	Z$43.86	Z$39.70
Budiriro	Z$140.37	Z$64.51	Z$25.60	Z$20.87	Z$33.30	Z$26.82
Chitungwiza	Z$225.74	Z$64.85	Z$37.80	Z$43.15	Z$46.86	Z$43.36

5.0 SYNTHESIS AND CONCLUSION

5.1 Synthesis

For the past three years, the federal government of Zimbabwe embarked on an urban transport restructuring journey that saw the prohibition of informal public transport and subsequent reintroduction of ZUPCO, a subsidised state-owned conventional bus service. While the traditional bus service arrangement seemed to work well during the early days of the COVID-19 pandemic, service reliability issues ensued as the stringent lockdown measures relaxed and the country returned to normalcy. All this was developed in the backdrop of a struggling economy as characterised by weak remuneration packages. As a result, some residents welcomed the bus service due to reduced fares, while most still rely on illegal informal transport services. To understand the choice behaviour of public transport captive users in Harare, we established six key objectives to guide a discrete mode choice study.

The first objective was to identify the drivers of mode choice and establish a methodology for assessing their contribution to mode choice behaviour. Chapter two addressed this objective. Chapter three fulfilled the study's second objective, which sought to design and conduct a discrete choice experiment suitable for mode choice analysis. We used a multistage sampling strategy to collect preference data from a specifically designed choice experiment in Greater Harare for empirical research. The experiment allowed us to uncover the contribution of preferences for comfort, travel time, and trip fare, free from the real-world complexities by implying a hypothetical free choice among different public transport modes, thereby responding to the second objective.

The fourth chapter estimated three classes of econometric models, including the MNL, MMNL, and latent class. The marginal utility results were interpreted in terms of parameter influence significance and behavioural meaning, thus fulfilling the research's third, fourth, and fifty objectives. Furthermore, the MNL and latent class models were interacted with covariates to explore preference across different sociodemographic groups. Chapter five outlines the policy interventions and service improvement recommendations, thus addressing the sixth objective. The discussion in this section is relevant to policymakers, transport planners, and public transport service providers which might be useful in further understanding the public transport landscape in other cities in Zimbabwe similar to the high-density suburbs used as study areas in this research.

5.2 Conclusion and recommendations

The study was motivated by the lack of a clear, integrated public transportation planning framework to address the urban transport inefficiencies in the country. The research implemented the utility maximisation framework to understand the mode choice dynamics in the Harare urban transport settings. The findings suggest rational behaviour while supporting the hypothesis of two heterogeneous latent classes in the population within which taste preference for public transport service quality is homogeneous. We evaluated the probability of class membership as a function of individuals' sociodemographic features. After several iterations, we found employment status, income, gender, and suburb to be the significant predictors of class allocation, with income being the most critical predicting covariate.

The heterogeneity in preference and willingness to pay across the population is key to public transport improvements that target specific transport interventions. The results indicate strong inertia towards kombis and aversion to ZUPCO buses within classes. The informal public transport sector remains the backbone of urban mobility in Africa, and the above finding emphasises the general dissatisfaction with the current public transport service provision in Zimbabwe. In addition, the respondents seem to be more sensitive to components of travel time like access time, which demands more physical effort. The implied willingness to pay estimates for service upgrades is generally higher in *class two* as compared to *class one*; however, the value of waiting time savings remains the same across the entire population. The perceived in-vehicle comfort appears to be more critical for *class two* members, as they are less willing to compromise on getting a seat. This finding contrasts with *class one* members, who appear to be more willing to accept a lesser amount to let go of the "*seat*".

It was essential to examine the class composition of the different sociodemographic populations; as such, we evaluated posterior latent class probabilities for gender, suburbs, and income levels. Interesting to note is the composition of suburbs; for every ten Budiriro respondents, more than eight were likely to be categorised in *class one*, which has most of the lowest income population. On the other hand, Chitungwiza had the highest probability of being classified in *class two*, while Dzivarasekwa and Warren had equal representation. Consequently, the average willingness to pay for service quality upgrades was high in Chitungwiza and lowest in Budiriro. The difference reflects the possible disparity in the service quality between the suburbs, being extremely poor in Chitungwiza and relatively better in Budiriro. That being the case, respondents from Chitungwiza will be more willing to pay for service improvement. In addition, the trip lengths to town may be a driving factor; however, further research can be done to test the hypotheses and investigate the influencing characteristics.

The findings of this study are relevant and applicable to the current public transport landscape in the capital city. The public transport restructuring programme in Harare aims to provide a reliable, affordable, and efficient public transport system; as such, the initiative's success should be assessed relative to a set of key performance indicators. Therefore, it is vital to understand the level of service that passengers need and how much the potential passengers are willing to pay for such services. The study identified two latent groups, and the transport authorities should develop an integrated public transport system with different services that accommodate the preferences of different population segments.

Policy provisions play a crucial role in shaping the urban mobility landscape. The intervention measures are more effective when tailored to account for heterogeneity in preference for the identified population segments. In that regard, the results from the posterior analysis can be helpful in the spatial allocation of different PT services. Since more than half of the respondents from Chitungwiza classify as *class two*, the local authorities should consider providing a relatively premium PT service. On the contrary, in Budiriro, where most of the population identify as *class one*, the planning agent can assign affordable services instead. Furthermore, the study findings can develop a level of service framework for each suburb or population segment against which PT service quality provision is assessed. The framework can be valuable to the public transport service providers, for example, ZUPCO, in understanding the service quality deficits and additional requirements of their service provisions.

The objectives of this research focused on understanding the mode choice dynamics for the public transport captive users, who, by the state of the economic landscape and public transport system in the country, are systematically deprived of access to essential socio-economic services. The research, however, acknowledges the need for transport demand management strategies that encourage choice users to shift towards more sustainable transport systems. To that effect, the study should therefore be extended in the future to explore the preferences of choice riders and different population segments.

The focus group discussions and the perceived variable descriptive analysis highlighted the general perception of the respondents on the efficiency and safety of all three PT services. ZUPCO services performed poorly on both scales, and policy interventions should advocate for the provision of safe environments and reliable services. The research did not investigate the contributions of safety and efficiency as latent variables. It will be interesting to explore the influence of the perceived variables on the observed mode choice decisions by employing integrated choice and latent variables models derived from (Ajzen, 1991) 's theory of Planned Behaviour. This advanced choice analysis will be critical, especially for choice riders whose latent variables will most likely influence their decisions.

6.0 REFERENCES

African Centre for Cities. 2015. *Urban infrastructure in Sub-Saharan Africa – Harnessing Land Values, Housing and Transport*.

Agyemang, E. 2017a. Mode choice for long distance trips: Evidence from the Greater Accra Metropolitan Area of Ghana. DOI: 10.1016/j.jtrangeo.2017.09.003.

Agyemang, E. 2017b. Mode choice for long distance trips: Evidence from the Greater Accra Metropolitan Area of Ghana. *Journal of Transport Geography*. 64:150–157. DOI: 10.1016/j.jtrangeo.2017.09.003.

Ajzen, I. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*. 50(2):179–211. DOI: 10.1016/0749-5978(91)90020-T.

Amiegbehor, D., Frederic, O., N'Gbesso, F., Gaye, C.O., Manful, E., Lami, M., Mokonyama, M., Analo, P., et al. 2010. Public Transport in Sub-Saharan Africa, Major trends and case studies. *UITP -International Association of Public Transport*. 10(8):2–24.

Balcombe, R., Mackett, R., Paulley, N., Preston, J. & Shires, J. 2004. *The demand for public transport: a practical guide*.

Behrens, R., McCormick, D. & Mfinanga, D. 2015. *Paratransit in African cities: Operations, regulation and reform*. DOI: 10.4324/9781315849515.

Ben-akiva, M., Bradley Morikawa J Benjamin T, M.T. & Oppewal V, N.H. 1994. Combining Revealed and Stated Preferences Data. *Marketing Letters*. 5(4):253–5324.

Ben-Akiva, M.E., Akiva, B., Lerman, S.L., Lerman, S.R., Press, M.I.T. & Lerman, S.R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. (MIT Press series in transportation studies). MIT Press.

Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E., Berk, R., Bollen, K.A., Brembs, B., et al. 2017. *Redefine statistical significance Strength of evidence from P values*. DOI: 10.1038/s41562-017-0189-z.

Bhat, C.R. 2003. Transportation Systems Planning Volume 20026459 (Methods and Applications) || Random Utility-Based Discrete Choice Models for Travel Demand Analysis.

Blamey, R.K., Bennett, J.W., Louviere, J.J., Morrison, M.D. & Rolfe, J.C. 2002. Attribute Causality in Environmental Choice Modelling. *Environmental and Resource Economics*. 23:167–186.

Bliemer, M.C.J. & Rose, J.M. 2013. Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological*. 58:199–214. DOI: 10.1016/J.TRB.2013.09.010.

Breidert, C., Hahsler, M. & Reutterer, T. 2006. *A REVIEW OF METHODS FOR MEASURING WILLINGNESS-TO-PAY*.

Bridges, J.F.P., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L.A., Regier, D.A., Johnson, F.R. & Mauskopf, J. 2011. Conjoint analysis applications in health - A checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health*. 14(4):403–413. DOI: 10.1016/j.jval.2010.11.013.

Brown, L.R. (Lester R. & Jacobson, J.L. 1987. *The future of urbanization: facing the ecological and economic constraints*. Washington DC: Worldwatch Inst.

Bryceson, D.F., Maunder, D., Mbara, T., Kibombo, R., Davis, A. & Howe, J. 2003. Sustainable Livelihoods, Mobility and Access Needs. *Department of International Development*. 57.

Cevero, R. 2000. *Informal Transport in the Developing World*. United Nations Centre for Human Settlements (Habitat).

Charamba, C., 2016. *More talk than action in dealing with mushikashika.* [Online] Available at: https://www.herald.co.zw/more-talk-than-action-in-dealing-with-mushikashika/[Accessed 13 April 2021].

Cherchi, E. & Hensher, D.A. 2015b. Workshop synthesis: Stated preference surveys and experimental design, an audit of the journey so far and future research perspectives. In *Transportation Research Procedia*. V. 11. 154–164. DOI: 10.1016/j.trpro.2015.12.013.

Chikozho Mazarire, G. & Swart, S. 2014. “The Pots on Our Roads”: The Diaspora Fleet and Harare’s Urban Commuter System. *African Diaspora*. 7:68–88. DOI: 10.1163/18725465-00701004.

Choice Metrics. 2018. *Ngene 1.2 USER MANUAL & REFERENCE GUIDE The Cutting Edge in Experimental Design*. Available: www.choice-metrics.com [2021, June 11].

Daly, A., Hess, S. & Train, K. 2012. Assuring finite moments for willingness to pay in random coefficient models. *Transportation*. 39(1):19–31. DOI: 10.1007/s11116-011-9331-3.

Eboli, L. & Mazzulla, G. 2008. *Willingness-to-pay of public transport users for improvement in service quality*.

Eboli, L. & Mazzulla, G. 2009b. A New Customer Satisfaction Index for Evaluating Transit Service Quality. *Journal of Public Transportation*. 12(3):21–37.

Elisabeth Kløjgaard, M., Bech, M. & Søgaard, R. 2012. Designing a Stated Choice Experiment: The Value of a Qualitative Process. *Journal of Choice Modelling*. 5(2):1–18. Available: https://www.econstor.eu/bitstream/10419/66822/1/730341690.pdf [2022, January 06].

European Commission. 1998. *Quattro Final Report: Synthesis and recommendations*.

Farvacque-Vitkovic, C. & Kopanyi, M. 2019. *Better Cities Better World: A Handbook on Local Governments Self-Assessments*. Washington DC: The World Bank.

Garcia-Sierra, M., van den Bergh, J.C.J.M. & Miralles-Guasch, C. 2015. Behavioural economics, travel behaviour and environmental-transport policy. *Transportation Research Part D*. 41:288–305. DOI: 10.1016/j.trd.2015.09.023.

Greene, W. 2001. Preliminary. Comments invited Fixed and Random Effects in Nonlinear Models. Available: www.stern.nyu.edu/~wgreene. [2022, January 03].

Greene, W.H. & Hensher, D.A. 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*. 37(8):681–698. DOI: 10.1016/S0191-2615(02)00046-2.

Hayes, G. & Venter, C. 2017. A META-ANALYSIS OF TWO GAUTENG STATED PREFERENCE DATA SETS. In *36th Southern African Transport Conference.* 335–349.

Hensher, D.A. & Prioni, P. 2002. A Service Quality Index for Area-wide Contract Perfomance Assessment. *Journal of Transport Economics and Policy*. 36(1):93–113.

Hensher, D.A., Stopher, P. & Bullock, P. 2003. Service quality - developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A: Policy and Practice*. 37(6):499–517. DOI: 10.1016/S0965-8564(02)00075-7.

Hess, S. 2005. ADVANCED DISCRETE CHOICE MODELS WITH APPLICATIONS TO TRANSPORT DEMAND.

Hess, S. 2014. *Latent class structures: taste heterogeneity and beyond*.

Hess, S. & Palma, D. 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*. 32:100170. DOI: 10.1016/J.JOCM.2019.100170.

Hess, S. & Palma, D. 2021. *Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application version 0.2.5 User manual*. V. 0.2.5. Available: www.ApolloChoiceModelling.com [2021, December 20].

Hess, S., Rose, J.M. & Polak, J. 2010. Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transport Research Part D*. 405–417. DOI: 10.1016/j.trd.2010.04.008.

Hess, S., Daly, A., Daziano, R. & Bliemer, M. 2020. *Statistical significance in choice modelling: a commentary on the use of confidence intervals, t-ratios, p-values and star measures*.

ITDP. 2018. *The BRT Planning Guide*. 4th ed. Available: https://github.com/ITDP/the-online-brt-planning-guide.

Jordan, J.D. 1983. Public Transport in Harare. *Zambezia*. 11(2):127–138.

Kamusoko, C., Gamba, J. & Murakami, H. 2013. Monitoring Urban Spatial Growth in Harare Metropolitan Province, Zimbabwe. *Advances in Remote Sensing*. 322–331. Available: http://dx.doi.org/10.4236/ars.2013.24035.

Kittelson & Associates., National Research Council (U.S.). Transportation Research Board., Transit Cooperative Research Program., United States. Federal Transit Administration. & Transit Development Corporation. 2013. *Transit capacity and quality of service manual*. Transportation Research Board.

Kittelson & Associates, Parsons Brinckerhoff, KFH Group, Texas A&M Transportation Institute & Arup. 2013. *TCRP Report 165: Transit Capacity and Quality of Service Manual, 3rd edition (TCQSM)*. Washington DC: Transportation Research Board.

Kjaer, T. 2005. A review of the discrete choice experiment-with emphasis on its application in health care.

Kunhikrishnan, P. & Srinivasan, K.K. 2017. Investigating behavioral differences in the choice of distinct Intermediate Public Transport (IPT) modes for work trips in Chennai city. DOI: 10.1016/j.tranpol.2017.10.006.

Lancsar, E. & Louviere, J. 2008. PRACTICAL APPLICATION Conducting Discrete Choice Experiments to Inform Healthcare Decision Making A User's Guide. *Pharmacoeconomics*. 26(8):661–677.

Lanz, B., Provins, A., Bateman, I.J., Scarpa, R., Willis, K. & Ozdemiroglu, E. 2010. Investigating willingness to pay–willingness to accept asymmetry in choice experiments. In *Choice Modelling: The State-of-the-art and The State-of-practice*. Emerald Group Publishing Limited.

Litman, T.A. 2013. Understanding transport demands and elasticities: how prices and other factors affect travel behavior. *Victoria Transport Policy Institute*. 1–76. Available: http://www.vtpi.org/elasticities.pdf.

Liu, X., Gao, L., Ni, A. & Ye, N. 2020. Understanding better the influential factors of commuters' multi-day travel behavior: Evidence from Shanghai, China. *Sustainability (Switzerland)*. 12(1):1–13. DOI: 10.3390/su12010376.

Luce, R.D. (Robert D. 1959. *Individual choice behavior : a theoretical analysis*. New York: Wiley.

Luke, R. & Heyns, G.J. 2020. An analysis of the quality of public transport in Johannesburg, South Africa using an adapted SERVQUAL model. *Transportation Research Procedia*. 48:3562–3576. DOI: 10.1016/J.TRPRO.2020.08.095.

Mangham, L.J., Hanson, K. & Mcpake, B. 2009. How to do (or not to do). .. Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning*. 24:151–158. DOI: 10.1093/heapol/czn047.

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Glenk, K., Dekker, T., Liebe, U., Olsen, S.B., et al. 2021. *Environmental Valuation with Discrete Choice Experiments Guidance on Design, Implementation and Data Analysis*.

Martínez-Cruz, A.L. 2015. Latent class logits and discrete choice experiments: Implications for welfare measures. *Revue d'Economie Politique*. 125(2):233–251. DOI: 10.3917/redp.252.0233.

Matamanda, A.R., Chirisa, I., Dzvimbo, M.A. & Chinozvina, Q.L. 2020. The political economy of Zimbabwean Urban informality since 2000 – A contemporary governance dilemma. *Development Southern Africa*. 37(4):694–707.

Maunder, D. & Mbara, T.C. 1995. The initial effects of introducing commuter omnibus services in Harare Zimbabwe. *Transport Research Laboratory*. (ISSN 0968-4107). Available: http://trl.demo.varistha.co.uk/uploads/trl/documents/trl123.pdf.

Mayo, F.L. & Taboada, E.B. 2020. Ranking factors affecting public transport mode choice of commuters in an urban city of a developing country using analytic hierarchy process: The case of Metro Cebu, Philippines. *Transportation Research Interdisciplinary Perspectives*. 4:100078. DOI: 10.1016/j.trip.2019.100078.

Mazur, K. & Bennett, J. 2008. Choice modelling in the development of natural resource management strategies in NSW. Available: http://www.crawford.anu.edu.au [2022, January 06].

Mbara, T.C. 2006. COPING WITH DEMAND FOR URBAN PASSENGER TRANSPORT IN ZIMBABWE : CHALLENGES AND OPTIONS. In *Southern African Transport Conference*. Pretoria. 121–129.

Mbara, T.C. 2015. Achieving sustainable urban transport in Harare, Zimbabwe: What are the requirements to reach the milestone? In *ENERGY, CLIAMTE AND AIR QUALITY CHALLENGES: THE ROLE OF URBAN TRANSPORT POLICIES IN DEVELOPING COUNTRIES*. Istanbul.

Mbara, T. & Pisa, N. 2019. An analysis of impediments to deliver sustainable transport in cities of developing countries: The case of Harare, Zimbabwe. *WIT Transactions on the Built Environment*. 182:241–252. DOI: 10.2495/UT180231.

Mbara, T., Dumba, S. & Mukwashi, T. 2014. Multi-stakeholder dialogue on formal and informal forms of public transport in Harare, Zimbabwe: Convergence or divergence perspective. *Journal of Transport and Supply Chain Management*. 8(1):1–9. DOI: 10.4102/jtscm.v8i1.140.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*. 105–142. Available: https://cir.nii.ac.jp/crid/1572824499423114240.bib?lang=en [2022, September 28].

Mcfadden, D. 1974. THE MEASUREMENT OF URBAN TRAVEL DEMAND. *Journal of Public Economics*. 3:303–328.

Mcfadden, D. 1977. Modelling the Choice of Residential Location. *Cowles Foundation Discussion Paper No 477*. Available: https://elischolar.library.yale.edu/cowles-discussion-paper-series/710 [2021, December 28].

Mcfadden, D. 2001b. Economic Choices. *The American Economic Review*. 91(3):351–378.

Mcfadden, D. & Train, K. 2000. MIXED MNL MODELS FOR DISCRETE RESPONSE. *Journal of Applied Econometrics*. 15:447–470.

Miller, K.M., Hofstetter, R., Krohmer, H. & Zhang, Z.J. 2011. How Should Consumers' Willingness to Pay be Measured? An Empirical Comparison of State-of-the-Art Approaches. *Journal of Marketing Research*. 48(1):172–184. DOI: 10.1509/jmkr.48.1.172.

Ministry of Transport Communications and Infrastructural Development. 2011. *National Transport Policy 2012-2016*.

Mitric, S. 2008. A Framework for Urban Transport Projects: Operational Guidance for World Bank Staff. Available: https://openknowledge.worldbank.org/handle/10986/17411.

Mlambo, A.S. 1997. *The economic structural adjustment programme: the case of Zimbabwe, 1990-1995*. Harare: University of Zimbabwe Publications.

Muro-Rodríguez, A.I., Perez-Jiménez, I.R. & Gutiérrez-Broncano, S. 2017. DOI: 10.3389/fpsyg.2017.01011.

Mutongwiza, L., 2021. *Harare City to ban kombis: Introduces new transport system.* [Online] Available at: https://www.263chat.com/harare-city-to-ban-kombis-introduces-new-transport-system/[Accessed April 2021].

Ndibatya, I. & Booysen, M.J. 2020. Minibus taxis in Kampala's paratransit system: Operations, economics and efficiency. *Journal of Transport Geography*. 88(August):102853. DOI: 10.1016/j.jtrangeo.2020.102853.

Ndlovu, H., 2020. *Problems with Public Transportation In Zimbabwe.* [Online] Available at: https://www.tuverl.com/blog/problems-with-public-transportation-in-zimbabwe/[Accessed 2021 April 2021].

Onderwater, P. & Kishoon, A. 2017. INFLUENCE FACTORS FOR PASSENGER TRAIN USE. Pretoria: 36th Southern African Transport Conference. 626–639. Available: https://repository.up.ac.za/bitstream/handle/2263/62742/Onderwater_Influence_2017.pdf?sequence=1&isAllowed=y [2021, May 19].

van der Pol, M., Currie, G., Kromm, S. & Ryan, M. 2014. Specification of the utility function in discrete choice experiments. *Value in Health*. 17(2):297–301. DOI: 10.1016/J.JVAL.2013.11.009.

Porter, G., Abane, A. & Lucas, K. 2020. User diversity and mobility practices in Sub-Saharan African cities: understanding the needs of vulnerable populations. The state of knowledge and research. Available: http://www.vref.se/macprogramme.

Prommaharaj, P., Phithakkitnukoon, S., Demissie, M.G., Kattan, L. & Ratti, C. 2020. Visualizing public transit system operation with GTFS data: A case study of Calgary, Canada. *Heliyon*. 6(4):e03729. DOI: 10.1016/J.HELIYON.2020.E03729.

Redman, L., Friman, M., Gärling, T. & Hartig, T. 2013. Quality attributes of public transport that attract car users: A research review. *Transport Policy*. 25:119–127. DOI: 10.1016/j.tranpol.2012.11.005.

Richard J. Carrick. 1979. Modelling Travel Behavior. *The Royal Geographical Society*. 11:90–91.

Rose, J.M., Michiel, • & Bliemer, C.J. 2013. Sample size requirements for stated choice experiments. *Transportation*. (February, 1):1021–1041. DOI: 10.1007/s11116-013-9451-z.

Sanko, N. 2001. Guidelines for Stated Preference Experiment Design (Professional Company Project in Association with RAND Europe) School of International Management Ecole Nationale des Ponts et Chaussées (Class 2000/01).

Scarpa, R., Thiene, M. & Train, K. 2008. UTILITY IN WILLINGNESS TO PAY SPACE: A TOOL TO ADDRESS CONFOUNDING RANDOM SCALE EFFECTS IN DESTINATION CHOICE TO THE ALPS. DOI: 10.1111/j.1467-8276.2008.01155.x.

Schmidt, J., Tammo, & & Bijmolt, H.A. 2019. Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*. 499–518. DOI: 10.1007/s11747-019-00666-6.

Simon, H.A. 1959. Theories of Decision-Making in Economics and Behavioral Science. 49(3):253–283.

Small, K.A. 1987. A Discrete Choice Model for Ordered Alternatives. 55(2):409–424.

Torres, C., Hanley, N. & Riera, A. 2011. How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments $. *Journal of Economics and Management*. 62:111–121. DOI: 10.1016/j.jeem.2010.11.007.

Train, K.E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. New York: Cambridge University Press.

Tversky, A. & Kahneman, D. 1986. Rational Choice and the Framing of Decisions. *Source: The Journal of Business*. 59(4):251–278.

Veldwijk, J., Lambooij, M.S., van Til, J.A., Groothuis-Oudshoorn, C.G.M., Smit, H.A. & de Wit, G.A. 2015. Words or graphics to present a Discrete Choice Experiment: Does it matter? *Patient Education and Counseling*. 98(11):1376–1384. DOI: 10.1016/j.pec.2015.06.002.

Venter, C. 2016. *ARE WE GIVING BRT PASSENGERS WHAT THEY WANT? USER PREFERENCE AND MARKET SEGMENTATION IN JOHANNESBURG.*

Vuchic, V.R. 2005.

World Bank. & Gwilliam, K.M. 2002. *Cities on the Move : a World Bank Urban Transport Strategy Review.* World Bank.

Zaki Wagner Meira Jr, M.J. 2003.

Zimcodd. 2020. Available: https://twitter.com/ZIMCODD1/status/1262713487777808384 [2022, February 11].

Zimstat. 2013. *Zimbabwe Population Census 2012 Report: Harare Province.*

ZimStat. 2015. *Population Projections Thematic Report.*

van Zyl, N.J.W. & Hugo, J. 2002. PUBLIC TRANSPORT IN CAPE TOWN: EVALUATING MODAL ATTRIBUTES WITH STATED PREFERENCE MODELS. 15–19.

Zimbabwe United People's Company, 2020. *ZUPCO Home.* [Online] Available at: http://www.zupco.co.zw[Accessed 12 April 2021].

7.0 APPENDICES

Appendix 1: Output from Ngene efficient design

Scenario	Kombi _AT	Kombi _WT	Kombi _Seat	kombi_ IVT	Kombi _Fare	Zupco_ AT	Zupco_ WT	Zupco_ Seat	Zupco_ IVT	Zupco_ Fare	car_hh .AT	car_hh _WT	car_hh _Seat	car_hh_ IVT	car_hh Fare	Block
6	0	-1	0	-1	1	-1	-1	3	0	0	0	1	2	0	-1	1
9	0	-1	1	1	0	-1	1	3	1	1	1	-1	2	0	-1	1
13	0	0	0	-1	0	-1	0	0	1	0	-1	0	0	1	-1	1
14	1	1	1	-1	1	1	-1	0	-1	-1	0	1	2	1	1	1
18	0	1	1	0	0	1	1	0	0	0	-1	1	2	-1	0	1
19	0	1	0	1	0	-1	0	0	1	0	0	-1	2	0	0	1
22	0	0	1	0	1	1	0	3	-1	-1	0	-1	0	1	1	1
24	0	0	0	-1	-1	0	1	0	1	1	-1	0	2	-1	-1	1
2	1	1	0	-1	0	0	1	0	1	-1	1	-1	2	1	1	2
3	-1	0	0	1	0	0	0	0	-1	-1	-1	1	2	-1	0	2
4	1	0	0	0	0	1	1	0	-1	0	0	-1	2	-1	0	2
7	-1	-1	1	-1	1	-1	-1	3	1	1	1	-1	2	-1	1	2
8	1	-1	1	1	1	-1	0	3	1	0	-1	1	2	-1	0	2
17	1	0	1	-1	-1	-1	-1	0	1	0	1	1	0	1	0	2
20	-1	1	1	1	1	0	0	3	1	-1	1	-1	0	1	1	2
23	1	0	0	1	1	1	0	3	1	0	-1	-1	0	-1	-1	2
1	0	1	1	1	1	-1	0	0	-1	0	0	-1	2	1	-1	3
5	-1	-1	0	0	1	-1	1	0	0	-1	-1	-1	2	1	1	3
10	1	0	1	-1	0	1	1	0	0	1	-1	-1	0	1	-1	3
11	-1	-1	0	0	-1	0	-1	3	1	0	0	-1	0	1	-1	3
12	0	0	0	-1	0	0	1	0	1	1	-1	-1	2	0	-1	3
15	1	1	1	0	0	-1	0	3	1	0	1	1	2	-1	0	3
16	-1	1	0	1	0	0	-1	3	1	0	-1	1	2	-1	0	3
21	1	-1	1	-1	0	-1	1	0	-1	1	1	0	0	-1	1	3

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Faculty of Engineering & the Built Environment
Centre for Transport Studies

Appendix 2: Dzivarasekwa sample survey questionnaire

ACADEMIC RESEARCH

My name is Masimba Tutsirai Mapfurira studying for a Master of Engineering in Transport Studies at the University of Cape Town. I am conducting research on issues that are related to personal preferences when making public transport mode choice decisions. The study seeks to understand the mode-specific factors that influence public transport mode choice decisions for the population from low income/high density suburbs and further assess the effect of their inherent attitude to the observed choices when making such trips to town.

Statement of Confidentiality: I can assure you that all the information that you provide during this exercise will solely be used for the intended purpose of this research project. Your name or any form of identity will neither be collected nor provided to anyone.

Voluntary Participation: Approximately 10-15 minutes of your time will be taken by this interview and a few questions related to your frequent trip and general mode choice behaviour will be asked. I am pleased that you have voluntarily agreed to participate in this research. Please note that you are not under any obligation to respond to all the questions that will be asked. Should you not feel comfortable, you are allowed to withdraw from the interview at any given time.

Benefits: While no direct compensation will be extended to you for your participation in this interview, the results will expand our knowledge base on this subject and the findings may be valuable in providing more efficient and reliable public transport systems which will benefit the public. Thank you very much for your time and support. Please start the survey by answering the questions below:

I, as the participant, volunteer to take part in this master's research questionnaire. I understand that the research aims to collect data on public transport user mode preferences and the data collected will be used in a master's thesis and give more insights into mode choice behaviour.

By proceeding to take this questionnaire, I freely consent to take part in this research project, and I agree with the statements above.

Signature of the participant
Date:

DISCRETE MODE CHOICE EXPERIMENT QUESTIONNAIRE

Section A: Information on previous main trip purpose

This section seeks to understand your most frequent trip. Kindly fill in the questions below in the spaces provided.

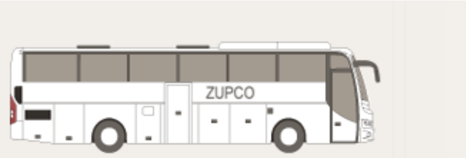
	What is your frequent trip purpose?					*Tick where applicable below*	
Work		**Leisure**		**Education**		**Other:** ***Comment***	

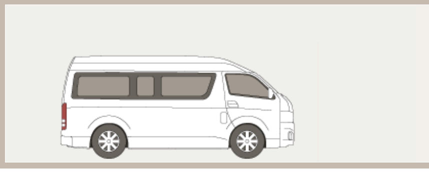
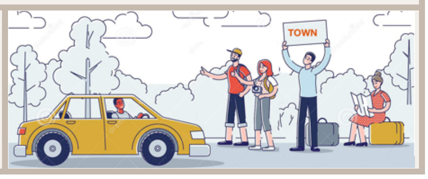
2	How often do you use public transport per week?			*Tick where applicable below*	
1 – 2 times		**3 – 4 times**		**5 times or more**	

3	What was the destination of your last frequent trip?	

4	How long did it take you to walk to the bus stop in minutes?							*Tick below*	
1 – 5mins		**6 – 10 mins**		**11 – 15 mins**		**16 – 20 mins**		**21 mins or more**	

5	Which mode did you use?	*Tick where applicable*



6. How long did you to wait at the bus stop?	
1 – 5 mins	
6 – 10 mins	
11 – 20 mins	
21 – 40 mins	
41 – 60 mins	
61 mins or more	

7. What was the seating arrangement?	
Seat available	
Standing	
pakadoma	
pick-up trailer	

8. How long did the vehicle take to your destination bus stop?	
0 – 15 mins	
16 – 30 mins	
31 – 45 mins	
46 – 60 mins	
1hr – 1hr 15 mins	
1hr 15 – 1hr 30 mins	
More than 1hr 30 mins	

9	How much did you pay for the trip?	**$Z**

Section B: Stated Choice Experiment

In section B, you are provided with eight hypothetical public transport mode choice scenarios, mark with **X** the mode you would use after considering the attributes.

Suppose you want to go to town for work, education or business and you can only use either zupco, kombi or hitch hike, given the attributes below which public transport mode would you use. **Mark with X**

CHOICE SET 1

attributes	 **kombi**	 **zupco**	 **hitch-hike**
Access Time	10 mins	5 mins	10 mins
Waiting Time	5 mins	5 mins	15 mins
Seating arrangement	seating	standing	pick-up trailer
Travel Time	33 mins	45 mins	45 mins
FARE	$Z 120	$Z 90	$Z 60
CHOICE (X)			

CHOICE SET 2

attributes	 **kombi**	 **zupco**	 **hitch-hike**
Access Time	10 mins	5 mins	15 mins
Waiting Time	5 mins	15 mins	5 mins
Seating arrangement	pakadoma	standing	pick-up trailer
Travel Time	1hr 7mins	1hr 7mins	45 mins
FARE	$Z 90	$Z 120	$Z 60
CHOICE (X)			

CHOICE SET 3

attributes	 kombi	 zupco	 hitch-hike
Access Time	10 mins	5 mins	5 mins
Waiting Time	10 mins	10 mins	10 mins
Seating arrangement	seating	seating	seating
Travel Time	33 mins	1hr 7mins	1hr 7mins
FARE	$Z 90	$Z 90	$Z 60
CHOICE (X)			

CHOICE SET 4

attributes	 kombi	 zupco	 hitch-hike
Access Time	15 mins	15 mins	10 mins
Waiting Time	15 mins	5 mins	15 mins
Seating arrangement	pakadoma	seating	pick-up trailer
Travel Time	33 mins	33 mins	1hr 7mins
FARE	$Z 120	$Z 60	$Z 120
CHOICE (X)			

CHOICE SET 5

attributes	 kombi	 zupco	 hitch-hike
Access Time	10 mins	15 mins	5 mins
Waiting Time	15 mins	15 mins	15 mins
Seating arrangement	pakadoma	seating	pick-up trailer
Travel Time	45 mins	45 mins	33 mins
FARE	$Z 90	$Z 90	$Z 90
CHOICE (X)			

CHOICE SET 6

attributes	 kombi	 zupco	 hitch-hike
Access Time	10 mins	5 mins	10 mins
Waiting Time	15 mins	10 mins	5 mins
Seating arrangement	seating	seating	pick-up trailer
Travel Time	1hr 7mins	1hr 7mins	45 mins
FARE	$Z 90	$Z 90	$Z 90
CHOICE (X)			

CHOICE SET 7

attributes	 kombi	 zupco	 hitch-hike
Access Time	10 mins	15 mins	10 mins
Waiting Time	10 mins	10 mins	5 mins
Seating arrangement	pakadoma	standing	seating
Travel Time	45 mins	33 mins	1hr 7mins
FARE	$Z 120	$Z 60	$Z 120
CHOICE (X)			

CHOICE SET 8

attributes	 kombi	 zupco	 hitch-hike
Access Time	10 mins	10 mins	5 mins
Waiting Time	10 mins	15 mins	10 mins
Seating arrangement	seating	seating	pick-up trailer
Travel Time	33 mins	1hr 7mins	33 mins
FARE	$Z 60	$Z 120	$Z 60
CHOICE (X)			

Section C: General Attitudinal Survey Statements

In this section, rate how much you agree with the following statements. Indicate with a **X**

In this section, rate how much you agree with the following statements. Indicate with a **X**

Statement	*Disagree*				*Agree*
I worry about losing my belongings, boarding buses	●	●	●	●	●
I worry about losing my belongings, boarding kombis	●	●	●	●	●
I worry about losing my belongings when hitchhiking	●	●	●	●	●
ZUPCO provides a reliable and efficient service	●	●	●	●	●
Kombis provide a reliable and efficient service	●	●	●	●	●
Hitch-hiking is the smoothest way to travel to town	●	●	●	●	●

Section D: Personal Information (Tick where applicable)

Gender: ***tick where applicable***	**Male**		**Female**	

Age in years	
18 – 24	
25 – 34	
35 – 44	
45 – 54	
55 or more	

Employment Status	
Student	
Informal	
Formal in public sector	
Formal in private sector	
Not employed	

Monthly income in $Z?	
10,000	
10,001 – 20,000	
20,000 – 30,000	
More than 30,000	
No income	

Appendix 3: Data dictionary for the City of Harare discrete mode choice data, “CityOfHarare data.csv”

VARIABLE	DESCRIPTION	VALUES
id	Unique respondent ID	1 to 361
TAZ	Trip origin where the survey was conducted	Mabvuku, Dzivarasekwa, Chitungwiza, Warren Park, Budiriro
choiceID	Index for the stated preference choice tasks	1 to 8
Choice	Preferred mode alternative	1: kombi, 2: zupco, 3: car hitch hike
age	Categorical variable for the age of the respondent	1: 18 - 24 years, 2: 25 - 34 years, 3: 35 to 44 years, 4: 45 - 54 years and 5: more than 54 years
income	Categorical variable for the income of the respondent	0: No income, 1: less than 10K, 2: 10K to 20K, 3: 20K to 30K, and 4: more than 30K
employment	Categorical variable for the respondent' employment status	1: Unemployed, 2: Student, 3: Informal, 4: Formal in public sector and, 5: Formal in the private sector
accessT_zupco	Walking time to zupco boarding facilities	5 mins, 10 mins, 15 mins
waitingT_zupco	Waiting time at the zupco boarding facility,	5 mins, 10 mins, 15 mins
seat_zupco	Dummy variable for zupco seating arrangement	0: seat available, 2: standing passenger
ivt_zupco	Duration of the trip from the origin bus stop to the destination bus stop	min: 21, mean: 53, max: 85
fare_zupco	Trip fare for zupco in Zimbabwean dollar	min: Z$60, mean: Z$99, max: Z$180
accessT_kombi	Walking time to kombi boarding facilities	5 mins, 10 mins, 15 mins
waitingT_kombi	Waiting time at the facilities for a kombi	5 mins, 10 mins, 15 mins
seat kombi	Dummy variable for kombi seating arrangement	0: seat vailable,1: pakadoma
ivt_kombi	Duration of the trip from the origin bus stop to the destination bus stop	min: 21, mean: 45, max: 85
fare_kombi	Trip fare for kombi in Zimbabwean dollar	min: Z$60, mean: Z$107, max: Z$180
accessT_car_hh	Walking time to car_hh boarding facilities	5, 10, 15
waitingT_car_hh	Waiting time at the facility for a car to come	5 mins, 10 mins, 15 mins
seat_car_hh	Dummy variable for car seating arrangement	0: seat vailable,1: p/u trailer
ivt_car_hh	Duration of the trip from the origin bus stop to destination bus stop	min: 21, mean: 47, max: 85
fare_car_hh	Trip fare for car_hh in Zimbabwean dollar	min: Z$60, mean: Z$96, max: Z$180
Safety_bus	Answer to "I worry about losing my valuable items each time I use Zupco"	Likert scale from 1 (strongly disagree) to 5 (strongly agree)

Safety_kombi	Answer to "I always worry about losing my valuable items when I commute using kombis"	Likert scale from 1 (strongly disagree) to 5 (strongly agree)
Safety_car	Answer to "I worry about losing my valuable items when I use other people's cars to commute".	Likert scale from 1 (strongly disagree) to 5 (strongly agree)
Reliable_bus	Answer to "ZUPCO provides a reliable and efficient service"	Likert scale from 1 (strongly disagree) to 5 (strongly agree)
Reliable_kombi	Answer to "Kombis are reliable and efficient"	Likert scale from 1 (strongly disagree) to 5 (strongly agree)
Reliable_car	Answer to "Hitch-hiking is a reliable and efficient service".	Likert scale from 1 (strongly disagree) to 5 (strongly agree)
trp_purpose	Destination for the previous frequent trip	1: work, 2: education, 3: leisure or shopping and 4: Other, hospital etc
trp_destn	Categorial variable for public transport weekly usage	Town, Msasa, Belvedere, ...
pT_usage	Categorial variable for public transport weekly usage	1: 1 to 2 times, 2: 3 to 4 times, 3: more than 5 times
accessT	Categorical variable for the walking time to boarding facilities in the last frequent trip	0: None, 1: 1 - 5 mins, 2: 6 - 10 mins, 3: 11 - 15 mins, 4: 16 - 20 mins and 5: more than 20 mins
Mode_used	Categorical variable for the mode used in the last frequent trip	0: bicycle/walk, 1: kombi, 2: zupco and 3: car_hh
sittingArr	Categorical variable for the sitting arrangement in the last frequent trip	0: None, 1: seat available, 2: pakadoma, 3: p/u trailer and 4: standing
IVT	Categorical variable for the duration of the trip to the destination bus stop in the last frequent trip	1: 0 - 15 mins, 2: 16 - 30 mins, 3: 31 - 45 mins, 4: 46 - 60 mins, 5: 61 - 75 mins, 6: 76 - 90 mins and 7: 91 - 120 mins
fare	Trip fare paid in the last trip in Zimbabwean Z$	min: 0, mean: Z$89.50, max: Z$180
egressT	Categorical variable for the walking time to the destination or connecting boarding facilities for a trip to the destination	0: None, 1: 1 - 5 mins, 2: 6 - 10 mins, 3: 11 - 15 mins, 4: 16 - 20 mins and 5: more than 20 mins
gender	Dummy variable for gender	0: Female and 1: male