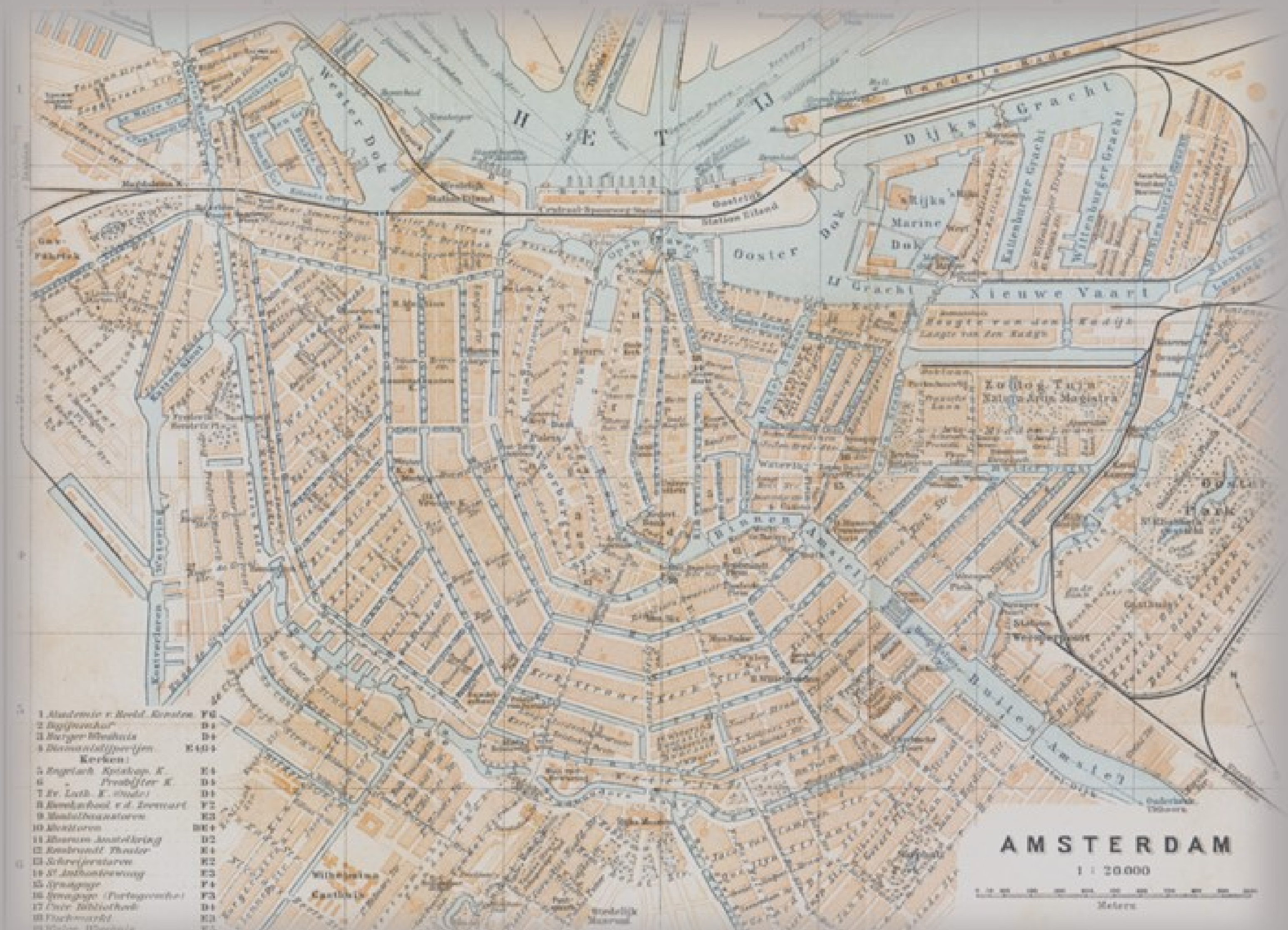
Modelling access and egress mode choice for multi-modal trips
Case study: City of Amsterdam

Author : Reddhi Uday Bhatt



TU Delft

Gemeente Amsterdam

Modelling access and egress mode choice for multi-modal trips

Case study : City of Amsterdam

by

Reddhi Uday Bhatt

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number:	5038995	
Project duration:	February 22, 2021 – November 10, 2021	
Thesis committee:	Prof. dr. ir. S.P. Hoogendorn,	TU Delft, (Chair)
	Dr. Ir. N. (Niels) van Oort,	TU Delft
	Dr. Ir. D.(Danique) Ton,	TU Delft
	Dr. N.M.(Natalia) Barbour,	TU Delft
	Ir. J. (Jos) van den Elshout,	City of Amsterdam
	M.(Marits) Pieters,	City of Amsterdam

Preface

Delft, November 2021

"It is not about the destination, it's about the journey."

This phase correctly captures my journey during my Thesis. It has been indeed has had all its ups and downs. I have been fortunate to have the right people guiding me in this journey, that has made all the difference. I have been always found myself gravitating towards public transport, I believe it stems from my love for travel. So when it came to selecting my thesis topic, I wished to work in the domain of public transport and contribute to make systems more accessible for all. Hence when I got the opportunity to intern at the Municipality of Amsterdam to work on the topic of 'Modelling access and egress mode choice for the public transport' it aligned with all my requirements and was the perfect match.

Firstly, I would like to thank my committee members,
Danique, you have been ever so supportive and understanding. On the days when I felt lost you guided me and brought me back on track. In moments of self-doubt your positivity and encouragement made it better. I am very thankful. To Natalia, I am grateful for all the insights you provided, nudging me in the right direction. I am proud of having such inspiring women in my committee. To Niels, I thank you for all the help you gave when I needed it. I shall never forget important lesson you taught me that is to 'pick my battles', to understand the trade-off all my decisions, it is something I will keep in mind throughout my life. To Serge, I am thankful all the support you provided. I would also like to thank Sanmay for all the help with the data pre-processing, it indeed made my life a little bit easier.
I am thankful to my company supervisors, Jos and Marits, for have being helpful, understanding and supportive. Besides the knowledge you helped me impart related to the study, you also taught me how to organize myself better. Having all of you as my committee has I believe brought out the best in me.
I am thankful to the Municipality of Amsterdam to give me the opportunity work in the public sector and be able to contribute whatever little I can to the society with my technical knowledge.

I am grateful to my friends who helped me get through even the toughest times. Being so far away from my family back home has not been easy but you made 'Delft' home for me. I would like to thank my flate-mates old and new Likhitha, Ranjan, Tanvi, Subhonil and Abhas, you made the house a home. For all the friends I made along the way at the University; Mariana, Merve, Jyotsna, Vivek, Flo, Johan, Ale; I am lucky to have such supportive and like-minded people as a part of my journey who made these years fun, I am sure there will be more to come in the future. My library buddies Arshdeep, Andrain, Philipp, Prateek and Anagha; all the coffee and lunch break conversations have indeed been memorable. Also, I am grateful to my friends back home whose pep-talks were indeed helpful. And everyone who has helped me in any way to get through this. I have made everlasting memories thanks to all of you.

Most importantly, I want to thank my family, who are my pillars of strength. Your unconditional love and unwavering support have helped me achieve everything ever wanted to. I am filled gratitude for this and all the opportunities you have enabled me to have. I am very grateful to my mom who is been my rock through out my life, always encouraging me to do better.I thank my father for all the support he gives me, it has enabled me to have the strength to endure difficult times.And I thank my brother whose positivity kept me going on. To my family I am forever thankful, there are not enough words that can capture how much your love and support means to me. Like all good things this comes to an end my journey at TU Delft comes to an end now, I am gratuitous for the opportunities I have had. Hope for a better tomorrow.

Summary

Improving accessibility in urban regions is an ever-evolving policy goal of the government. To deal with the congestion in urban areas encouraging the citizens to adopt public transport is one of the more popular approaches. To achieve this goal, providing the citizens with a seamless travel experience plays an important role. In order to encourage the adoption of public transport and multi-modal transport, it is useful to model mode choice to understand what factors are affecting mode choice for multimodal trips.

The starting point is to deduce what the current literature suggests as significant variables affecting the access and egress mode choice to and from public transport. Furthermore, the deduction of the factors helps pinpoint which aspects are important from a policy perspective to improve multi-modal transport. Literature suggests umpteen factors that affect mode choice. However, it depends on the context of the research. Furthermore, there has been some research in the domain of access and egress mode choice. However, more often the access and egress mode choice models are modelled separately. Thus, the simultaneous access and egress mode choice can be investigated further. To address the above-mentioned gaps in research, the objective is to provide a conceptual modelling approach that can model access and egress mode choice simultaneously. Furthermore, it is insightful to compare the impact of the factors affecting mode choice for both urban and regional networks. Thus, the framework shall be applied to the train network that connects different regions and the system within the urban network such as a metro network. The main research question is as follows;

How to model access and egress mode choice simultaneously (to and from) public transport systems for regional and urban networks?

As the city of Amsterdam is the case study adopted the train and urban network are analysed using individual models. A discrete choice modelling approach is adopted to maintain compatibility with the existing systems, as the models in practice use a discrete choice modelling approach. Moreover, it is also a popular approach in research.

For the modelling approach, the MNL model is considered to be the starting point. To account for the expected correlation between the alternatives a mixed logit model with random error components. For the generation of alternatives, mode chains are deduced for an origin-destination trip. The assumption made based on the literature review is that people are less likely to make trips more complex than two intra-modal transfers.

The first step in choice modelling is to generate the alternatives and determine the data collection system. Given that the case study is the city of Amsterdam and the context in the Netherlands; the ODiN (On der Weg) database is used as the main data source. It is a travel survey carried out in the Netherlands each year to carry out travel behaviour research. It comprises the socio-demographic data, trip characteristics, mode choice, etc. Hence, it is a form of revealed preference data is used for analysis. From the literature review and data analysis, it is evident that access and egress modes comprise of walking, cycling, car and Public transport (Bus/Tram/ Metro); with train and metro as the main modes. To determine a model that analyses the impacts of the factors affecting mode choice, a structured approach is implemented. The first step is to deduce a base model that comprises only the time variable and the alternative specific constants. The next step is to analyse the hypothesis individually on the base model.

The socio-demographic variables considered to be estimated are car ownership, age, gender, income, urban density, trip purpose, employment status, household composition. Hypotheses are formulated to deduce the impact of the selected factors affecting the access and egress mode choice. The hypothesis is individually tested for the train and metro model. The parameters that are significant in the individual model analysis are added to the combined model. The final model is optimised. Once

the final model is deduced using the MNL modelling approach then it is analysed using the ML (Mixed Logit) model with random error components.

CASE STUDY - CITY OF AMSTERDAM
Based on the 2018 -2019 ODiN data the total number of observations extracted for the model with the train as the main mode is 1187 and for the metro as the main mode is 405. Application of the framework for the given case study leads to certain changes in the choice set constructed. As the number of observations is not enough for all the alternatives it has been re-constructed a bit to deduce the model. It also suggests which alternatives are not preferred to be chosen by the respondents. As there are not enough observations for access and egress using cars they are combined into one alternative with car access- mix mode egress and mix mode access-mix mode egress. The travel time parameters are considered average across the modes for mix modes. In the case of the metro, the number of observations for cycle access and egress as well as car access and egress is limited. Thus, the cycle access -mix mode egress, mix mode access-cycle egress, car access-mix mode egress, mix mode access-car egress and mix mode access- car egress.

The analysis of the base model for the train suggests that the duration of the main mode has the least impact on mode choice. It is the access and egress mode choice that is weighted heavier. However, it must be considered that the main mode duration remains constant throughout all the alternatives. The access is weighted lower as compared to the egress. All the travel time parameters have a negative coefficient suggesting more the duration lower is the utility of the alternative.

In the case of the base model for the metro the main mode is not as significant and is positively correlated. It can be expected as for the trips to and from outside the city of Amsterdam the trip duration could be longer. In addition to that, the metro network is a relatively more dense and high-frequency urban system. Hence, it can be expected that the duration of the main mode is lower as compared to the access and egress in the case of the metro. Further in the time duration of access is weighted higher as compared to the egress. Thus, it suggests that people are more sensitive to the access time to the egress time to the metro. This is the opposite as compared to the trend observed in the train. The expected outcome would be the users being more sensitive to egress as there is an asymmetric mode availability on the egress side. Usually, individuals are expected to have access to the mode they own. However, in the case of the metro system if the users moving to and from the city use the metro as it is a denser network egress might be more walkable and accessible to the destination as compared to the access. Moreover, in the Dutch context, the cycle infrastructure is well developed within the city making it more of a competing mode rather than a complementary mode.

The final model optimization depicts that besides the access travel time and main mode travel time; the age, trip purpose, employment status and car ownership are significant. It is expected so as per the literature as well. The age hypothesis suggests that the respondents belonging to the age category of 18-24 are expected to use public transport for access and egress. The results are significant and positive suggesting that the hypothesis is true. It is an expected outcome as the Dutch students have access to free transport when enrolled for education. In case of the car ownership, it is highly significant and also has a high coefficient as expected. Though the overall trend in the data suggests that about 50% of the population has access to a car. For the trip purpose, it is assumed that the access and egress modes used are walking and cycling. Similarly, for full time employed respondents, it is assumed that active modes are used for access and egress. The coefficient is positive and significant. It is in line with the literature suggesting that professionals with a full-time job are expected to be highly educated and hence make more conscious decisions to choose active modes for access and egress. For the metro model, the final results suggest that only car ownership tends to be the significant variable.

For the ML model, the additional parameters added are the random error components to account for the expected correlation between the alternatives having the same access and/or egress mode. The mix logit model is introduced but leads to the same results suggesting that mode-chain acts as a synthetic model and when making a choice an individual considers the entire trip rather than only a trip leg.

The findings suggest that car ownership is an important factor and in an urban system it might be a competing model. Thus, policies focusing on regulation car ownership and the use of cars can be an effective solution. Furthermore, attributes addressing the station facilities can be introduced to understand the impacts of the station facilities on access and egress mode choice. Furthermore, in the case of trains, it is observed that the socio-demographic variable have a more significant impact. Factors such as having a full-time job have a strong positive correlation with active mode access and egress mode choice. This insight can be used to incentivize to attract users belonging to a similar group. Similarly, in the case of Dutch students having access to free public transport encourages them to adopt more multimodal trips. Hence, innovative incentive schemes, improving the service and facilities of cycling and walking can be improved to attract more users. For further research, it would be interesting to adopt other more hybrid models as well. The results also suggest that in an urban setting in the Dutch context the cycle might be a competing mode for the metro. The trade-off between both can be further investigated. The findings suggest that when the access and egress mode choice is modelled simultaneously, the access main and egress trip legs are not weighted differently within a trip. On the comparison of the train and metro model, it is observed that for the different modes of public transport how access and egress trip legs are weighted differently as well. Hence the policymakers need to implement tailor-made policies for the different form of public transport. This model can also analyse the impact of the travel costs and the travel time and cost models can be compared to deduce what the respondents are more sensitive to.

The framework deduced as a part of the study provides a starting point for modelling access and egress mode choice simultaneously. The overall findings suggest that the type of the network and its characteristics impact the access and egress mode choice.

Contents

1
Introduction

The goal of this chapter to provide a basis for the research. Furthermore, it provides the motivation for the research carried out and the structure of the research.

1.1. Introduction

In urbanized regions, the increased use of cars has led to traffic and congestion problems. To solve this issue, policies encouraging the users to adopt a more integrated network with public transport are necessary. Considering the case of the Netherlands, where the cycling infrastructure is well-developed, cars still have the highest share i.e. 47% (Harms & Kansen, 2018) as compared to other modes. However, with time there has been an increase in the use of cycles overall especially for access and egress to and from the train stations. This indicates the willingness of the users to shift towards more sustainable modes. To integrate the modes and provide a more seamless experience, more insights into multi-modal transport shall be investigated. Hence, the study aims to understand the mode choices of travellers. Usually, the trips that comprise the use of public transport are multi-modal. Thus, improving the door-to-door connectivity and accessibility to public transport is the need of the hour. In order to implement policies to improve accessibility transport models are utilized. Transport models are created as tools that can be used for decision-making for the development of transportation. The transport models are expected to provide insights into the impacts of policies relating to the implications of alternative transport on land-use investments and policies etc. Hence, transport models are usually implemented to deduce policy implications.The classic Four-step models are popular for the purpose of modelling the various choice dimensions Ton et al., 2019. In western European countries such as Sweden, Netherlands and Denmark adopt a four-step transport model. umteen modelling approaches are available but discrete choice modelling approach is often implemented Ton et al., 2019. With the development in research and technology, there are different models such as activity-based models, agent-based models etc. However, the current systems are based on the classical four-step model. Hence, the current study is focused on implementing the discrete choice modelling for mode choice models to maintain compatibility with the existing systems. Thus, the following subsections describe the current trends in the existing modelling approach implemented. Furthermore, how the access and egress mode choice behaviour are modelled is discussed.

1.1.1. Access and egress mode choice

As discussed in the previous section, the world is willing to move towards a more car free economy. To implement such policies understanding of the travel behaviour is necessary. Mode choice is a key step to deduce policy implications (Ortúzar & Willumsen, 2011). Thus, to gain a more realistic perspective of the mode choice involving a multi-modal trip chain, it is essential to understand the mode choices for access and egress travel to and from public transport. To explain access and egress a multi-modal trip is defined. A trip is defined as the travel from an origin to a destination. A trip comprising the use of more than one mode is a multi-modal trip. Walking, cycling, public transport and car are the different modes considered (Fiorenzo-Catalano, 2007). When using public transport as a mode, there are usually multiple modes involved to complete a trip. Thus, a multi-modal trip can be further divided

into three parts i.e. access leg, main leg and egress leg (Fiorenzo-Catalano, 2007). The access leg is the part of the trip from origin to the transfer node to the main leg. Similarly, the egress leg is the transfer from the main leg trip to a destination. In the case of the multi-modal perspective, the various transport systems are expected to be integrated in a way that there is a smooth transfer between the different modes. The following paragraph describes the current approach in research to model access and egress mode choice.
Studies focus on either access mode or egress mode individually or both separately. In the study conducted by Wen et al (2012) focuses on access mode choice only for high speed rail (Wen et al., 2012). A NL (Nested Logit) model with latent classes is estimated to deduce the preferences of the users. The findings suggest that the travelers are cost sensitive. The future research shall be carried out to deduce a more generalized model allowing the integration of access and egress mode choices (Wen et al., 2012). Ton et. al. (2020) investigates the factors affecting access mode choice along with station choice for cycling and walking as access modes. The study focuses on the factors affecting the access and station choice from tram as the main mode. However, for other public transit modes it is expected that the same factors might not be relevant. The mode for access and egress is cycling and walking. MNL (Multinomial Logit Model) model based on the distance of access and station. The study by Azami et. al. (2020) focuses on the region of Orlando to model access and egress to and from public transport.The factors affecting mode choice for the different modes including the micro-mobility modes are deduced (Azimi et al., 2020). Two separate models are implemented for access and egress mode choice. MNL models are implemented for analyses of the model. The findings are limited to the context of the region. Additionally, the transfer-ability is not is not accessed. In this study it is assumed that all the forms of public transport are expected to have the same factors affecting mode choice. Hence, the limitations of the above-mentioned studies is that simultaneous access and egress mode choice is not considered.

A study conducted in China multi-modal choice behavior is modelled for inter-city travel. The main modes comprise of airplane, train, express bus and HSR (High Speed Railway). BMNL (Bayasian Multinomial Logit) model is implemented which is a Bayesian based discrete choice models (M. Yang et al., 2015). This research considers access mode choice and departure mode choice as factors that effect the main mode choice. Thus, it provides insights how the access and egress are weighed depending on the different main modes available. Thus, it is a variable and the alternatives are the main modes, but does not address modelling the access and egress simultaneously. Moreover, it is limited to the context of the location of the case study. The research done by Waerden et. al.(2018) provides insights into the role of main modes and access modes on the decision to chose between car or train as the main mode of travel. Hence, access and egress is not the main focus though it is considered. The study mainly focuses on access mode choice with the main travel mode. The trade off between train and car as the main mode is deduced. Mixed logit (ML) model is estimated to deduce the implications (Waerden & Waerden, 2018). It depicts that the attributes related time and cost are influential. It provides insights into which factors shall be considered to encourage the car users to switch to train. However, various factors such as the trip purpose are not considered. The study by Yang et al (2019) focuses on access and egress mode choice to and from the high speed rail in China. The modelling of access and egress mode choice is done is separate stages. Separate models are analyzed for the business and leisure travelers (H. Yang et al., 2019).

Studies also determine the trade-off between uni-modal and multi-modal trips in the context of the Netherlands. Furthermore, the trade-off between the service quality, travel time and travel costs are deduced (Arentze & Molin, 2013). Insights into how the different stages of the trip such as access time, egress time and in-vehicle time are weighted is provided. However, in this case the public transport is considered as one mode and no differentiation amongst them is not carried out. Hence, most studies focus on access or egress individually. Some consider them as factors for a generalized approach towards public transport. The study by Yang et al (2015) provide a more comprehensive approach towards mode choice modelling for access and egress mode choice for the metro in the city of Nanajing. The focus is on the satisfaction of metro commuters. In this case the focus is only on one mode of public transport. Most studies consider all the modes of public transport as one assuming the that the behavior amongst all the modes of public transport is same, or the focus is only on a particular mode of public transport. This suggests there is not much research in the comparison between the travel behavior between the modes.

Hence, it is evident that most studies focus on either access or egress mode choice separately or only one of them. Most studies focus on access mode and do not consider the egress mode. Some studies discuss the factors affecting the choice of different multi-modal trips. The simultaneous access and egress mode choice is not investigated in depth. Additionally, studies assume that mode choice behavior for all modes of public transport are expected to be similar. Studies do not necessarily compare the different public transport modes (within the same context) to deduce if similar factors affect the access and egress mode choice. Form a societal perspective to encourage the use of public transport, it would be fruitful to gain insights into what are the differences are so as to provide tailor made policies addressing the different modes of public transport. Hence, there is a need for a more generalized modeling approach that integrates access and egress mode choice to determine the factors affecting the access and egress mode choice.Thus, a comprehensive method to model access and egress in a single trip and analyse the required factors in the given context is the existing gap. Moreover, the research only focuses on only a particular main mode of transport rather than comparing the different public transport systems such as the train network and the metro/tram/bus network. Generally, the train network is available at an intercity and intra-city level. In case of metro networks it is available in highly urban areas within a particular city. Thus, within the public transport domain the characteristics of the public transport network varies. As suggested by the current research it is at times assumed that the different modes of public transport would be expected to have similar mode choice behaviour.

1.1.2. State of the art

Focusing on mode choice perspective for public transport access and egress; the current research comprises research that addresses access and egress separately. It is expected that the traveller makes the decision of the entire trip in advance. Thus, a simultaneous approach would provide more insights into factors affecting the access and egress mode choice. It shall provide insights into how the individual experience access, egress and main trip legs within the same trip. Hence, the existing research gap is to incorporate a multi-modal trips for access and egress mode choice in a single model, also considering the different forms of public transport modes. The state of the art is provide a modelling framework that allows for the modelling of simultaneous access and egress mode choice. Moreover, provide insights into the factors affecting the mode choice in a single model. Secondly, provide an approach that is transferable to different public transport modes. Moreover, deduce if the factors mode choice behaviour for the different public transport modes varies or stays the same.
Hence, the results of the research will provide a modelling approach that can model access and egress mode choice simultaneously. Furthermore, implementing the modelling framework analyse regional level networks and analyse the nuances of the travel behaviour. Thus, the added value of the current research is to simultaneously model access and egress mode choice. Furthermore, it is to provide an approach that is transferable to various public transport modes to deduce the different policy implications.

1.2. Scope and Context

On the basis of exiting research and the research gap the scope of the research and the case study considered are discussed.

1.2.1. Scope

The scope comprises mainly modelling the mode choice for access and egress travel to and from public transport. Moreover, to understand the nuances in the choices for not only national or regional level networks like the train but also urban level public transport networks like bus, tram and metro. Hence, to implement a modelling approach an urban area where it is expected to have a higher number of multi-modal trips. To address these concerns, Amsterdam is the case study selected for the implementation and analysis of the mode choice model. Amsterdam considered as the case study, as the share of multi-modal trips is higher in urban regions i.e. the share of multi-modal trips is 20% in Amsterdam as compared to the Netherlands which is 3 %(Fiorenzo-Catalano, 2007). As the case study is the city of Amsterdam the trips to, from and within the city of Amsterdam are considered for modelling and analysis.

The deliverable is to provide a conceptual framework for modeling a multi-modal trip to and from public transport. Additionally, the requirement of the framework is to access the travel behaviour for mode choice for not only the train network but also the lower level networks such as the bus, tram and metro. As the study is for multi-modal trips only the uni-modal trips are considered outside the scope of study. Discrete choice modelling methods are considered to estimate the model. This approach is adopted to maintain the compatibility with the National Transport models in the Dutch context. In case of the city of Amsterdam the current modelling approach is a dis-aggregate model. With a discrete choice modelling approach hence the other approaches such as activity or Agent Based Modelling is no considered. The explanation and justification of making the choice of focusing on discrete choice models is provided in Chapter 3. Though route choice and mode choice are expected to be correlated, the focus of the research is the modelling of access and egress mode choice. Furthermore, only public transport modes are considered to be the main mode of travel, private modes such as car and bikes are not a part of the scope. To determine the applicability of the model the modelling approach is applied to the train and metro network users.

1.2.2. Context

The Netherlands is known as the logistic gateway to Europe hence the accessibility and the transport are very important to the economy ("Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management", 2010). The population of Netherlands is 16.5 million approximately, out of which about 7 million people reside in the Randstad area, so approximately 50% of the population is concentrated in the same region. The planning of the different ports to support economic activity is considered during the land-use planning process ("Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management", 2010). The growth in the population and the prosperity of the citizens has lead to higher mobility needs. To cater to them and manage the traffic is the need of the hour. In 2017 train use via NS Dutch Railways resulted to more than 18 billion passenger kilometres, an increase of 150 million (0.8 percent) compared to 2016 (Kim, 2018). This depicts a growth in the use of public transport amongst the people. However, it must be noted that there is also a moderate growth in the use of cars.

Approximately 4.5 million trips are made by bus, tram and metro in the Netherlands on a daily basis. Moreover, approximately one million are made by train and 14.5 million by cycles ("Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management", 2010). It must be considered that the Netherlands is an exception when it comes to the high utilization of bikes due to the well developed cycling infrastructure that is available. However, the trips made by bus /tram and metro are relatively lower as compared to the train. This is attributed to the urban density and the type region. In highly urban areas such as Amsterdam the trends are expected to differ from the national average. In terms of access and egress mode choice to and from the railway station the following trends have been observed figure 1.1 and figure1.2 (Kim, 2018). It can be observed that given the context of the Netherlands cycle is the predominant mode for access followed by public transport and walking. Though there is not much difference between the proportion of the modal shares between them. In case of the egress it is observed that walking is the predominant mode. It his can be attributed to the fact that at the activity end there is more uncertainty of the availability of the egress mode on the activity end. The future policy goals imply that a more accessible system is required to attract more public transport users. The approach now being chosen is one in which all the options are utilised both to stimulate accessibility and development opportunities and to limit the negative impacts of (car) mobility, that is to facilitate high quality alternatives for mobility and to stimulate conscious choices. For this the mobility system must be coherent and robust and all the modalities must be sufficiently solid to form fully-fledged alternatives ("Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management", 2010). The current policy goal pertaining to transport is to improve accessibility and reduce the negative impacts of car usage and encourage more conscious choices such as public transport. To do so there is a requirement for a more robust and coherent system.

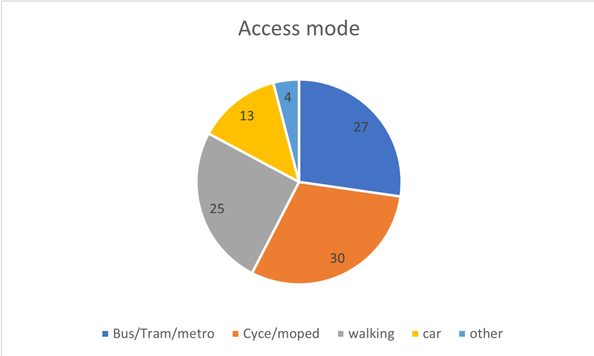


Figure 1.1: Access mode Kim, 2018

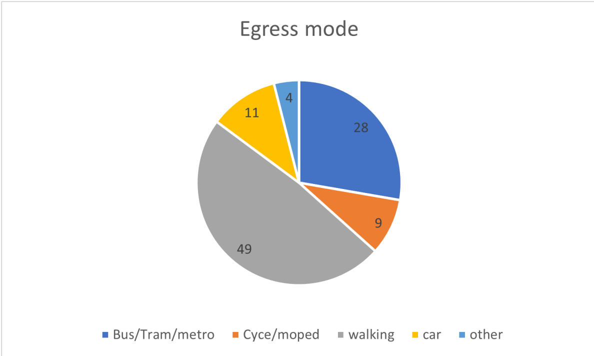


Figure 1.2: Egress mode (Kim, 2018)

1.3. Research objectives and research questions

The objective of the current research The objective of this section is to address the existing research gap.

1.3.1. Objective

The long term policy goal is to make cities more accessible. Understanding the mode choice behaviour and test the policy implications shall contribute to more informed decision making so as to provide a seamless trip. In order to achieve the above mentioned goals, the objective of the research is to improve the understanding of mode choice in a multi-modal trip. The the aim is two-fold; First, is to gain insights into the factors affecting the travel behaviour when access and egress mode choice is modelled simultaneously. Furthermore, provide a state of the art modelling approach to model mode choice travel behaviour more realistically. Secondly, it is to compare the regional and urban level networks and address the the how it impacts the users choice. Furthermore, test various hypothetical scenarios using the model for both networks and analyse the results. So as to achieve the aim the deliverable of a conceptual framework for a modelling approach for an access egress mode choice model. Additionally it is expected to be implemented not only for regional level network i.e. train networks but also for the urban networks such as metro rail network available. The conceptual modelling framework will be applied on the case study of Amsterdam.

1.3.2. Research questions

To address the aim of the research the objective is formulated in the form of the main question. In order to address the main question it is further broken down into sub-questions.
Main question: **How can the mode choice of access and egress be modelled by considering the lower level networks for public transport, in a multi-modal trip simultaneously?**
To address these research questions the following **sub-questions** have been formulated:
1.What are the significant factors affecting the mode choice for access and egress to and from public transport in a multi-modal trip?

The above-mentioned question addresses the significant attributes or variables affecting mode choice based on the literature review of the current research. It is discussed in detail in chapter 2, there are studies carried out in the Dutch and international contexts for deducing the factors affecting the access and egress mode choice to/from public transport. The findings obtained addressing this question is an input for the model i.e. the attributes affecting mode choice and the factors considered to generate the different alternatives.

2. To what extent the Discrete choice modelling approach can be implemented for multi-modal trips?
In order to have a base for the conceptual framework the existing literature will be utilized to determine the methodological approach for modelling. This question addresses the alternatives considered so as to deduce the choice sets based on the current approach. Furthermore, the modelling method that is adopted will be justified. Moreover, The generation of alternatives and the data collection depends on the modelling method as well. Hence, the different steps carried for modelling shall be addressed in this sub-question. The existing literature is expected to help determine the method that shall be implemented to model mode choice. Thus, the findings from the literature review will inform the method that will be applied for the selected case study.

3. How can a multi-modal trip be modelled considering for the context of Amsterdam?
This sub-question addresses the application of the modelling framework for the case of Amsterdam. Sub-questions 1 and 2 address the aspects that have to be considered when modelling multi-modal transport. Additionally, as mentioned it will be useful not only for trains but also can be adapted to lower level networks available on an urban level such as bus, tram and metro. It addresses the structured approach that will be applied to model the mode choice and deduce the impact of the factors affecting mode choice. Hence, a framework to model mode choice in a multi-modal trip chain shall be an expected output of the research.The section 1.5 elaborates on the data pre-processing and analysis for the case study considered.

4. What are the impacts of the factors affecting the access and egress mode choice for train and metro (as main mode)? What are the differences between them?
This sub-question addresses the analysis of the results obtained by applying the modelling framework for the city of Amsterdam. For the different modes of public transport considered in the current research the modes considered are train and metro. It is expected that for the different modes of public transport the factors affecting mode choice behaviour have different implications. Policy hypothesis are tested to check the significant factors affecting the access and egress mode choice will be analysed for the city of Amsterdam. This shall be achieved by analysing the outcomes of the model for a given policy hypothesis. This will deduce if the model provides results in line with expectations. Moreover, the same hypothesis will be tested for both train and metro network to compare the outcomes for both networks.

1.4. Scientific and societal relevance

This section describes the value of the research to science and practice.

1.4.1. Contribution to science

As depicted in the section 1.1.1, there have been various studies focusing on access and egress separately. Additionally, studies focus predominantly on trains or a specific mode of public transport. However, the a simultaneous model can be investigated at a greater depth. Moreover, the comparison the lower and higher level networks has not been ventured into. Additionally, the comparison between the different public transport modes i.e. bus, tram, and metro with a train can be investigated further. The scientific contribution is to prepare a conceptual framework to model multi-modal trips more realistically. The societal relevance is to understand the mode choice implications and use it as an input to develop a more seamless travel experience. The objective of transport models is to simulate the required situation as realistically as possible to investigate the expected implications of the different transport policies that are required to achieved. The contribution of the research is the provision of a modelling approach that can be implemented to analyse the access and egress mode choice behaviour for different modes of public transport simultaneously. Furthermore, provide an approach that allows for analysing various factors affecting mode choice based on the hypothesis to be tested. Moreover, the model is expected to be transferable to the different contexts for different countries.

1.4.2. Contribution to practice

The long term goals of the Dutch government comprise of providing more optimised network so as to encourage people to use public transport. In order to achieve that goal it is necessary to improve the coordination between the urban transport networks ("Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management", 2010). There are multiple aspects associated with the improvement of the coordination is technical which is associated with managing the schedule and the time table so that the existing systems complement each other. The other aspect is understanding the requirements from a behavioural perspective for more informed decision making. There is an expected increase of 1.1 % from 2017 to 2023 annually. It is also stated that the travel by bus, tram and metro are less dynamic as compared to rails and no policies are set Kim, 2018. Hence, understanding the implications of the difference between the train and the lower level urban networks can contribute to more informed policies. Furthermore, the Amsterdam policy 2030 focuses on making Amsterdam more attractive and accessible. Though the public transport network has been developed since 1990 there is not much difference in the modal choices that have remained the same. A behavioural model shall be helpful to direct policies that encourage public transport use. The modelling approach that is obtained as the output of this research works as a tool for decision makers on a strategic level to determine relevant factors that affect the mode choice behaviour. It can aid the decision maker to deduce if a particular aspect is worth venturing into in depth. In addition to that it contributes to the understanding of the preferences of the people towards the access, egress and main legs of the trip. This can help pin point which are the relevant aspects of the trip to focus on. Hence, obtain insights what as per the perspective of the traveller can help improve the accessibility to and from public transport.

1.5. Methodology and the Structure of Thesis

As depicted in the figure 1.3 the structure of the thesis and the methods used to address the research questions formulated is as follows;

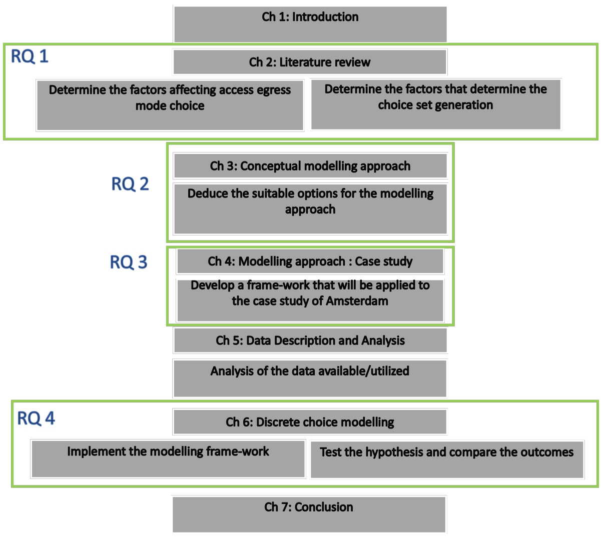


Figure 1.3: Structure of thesis

Chapter 2: Literature review

This chapter addresses the current approaches applied for access and egress mode choice modelling. In addition to the mathematical approach used the other steps involving the methods of data collection, generation of alternatives and the factors affecting mode choice. Thus, the literature review address the inputs required to prepare a conceptual framework to model access and egress mode choice. The findings of this chapter address the sub-question-1.

Chapter:3 Modelling approach
In this chapter the conceptual framework that is deduced based on the the findings of the literature review is described in detail. Furthermore, it addresses the exact modelling approach that can be adopted to any context ideally and all the required inputs are explained in depth. The findings of this chapter addresses the sub-question-2.

Chapter: 4 Modelling approach : Case study of Amsterdam
Chapter 4 describes how the conceptual framework deduced in chapter 3 can be applied onto the Case study of Amsterdam. Introduction to the case study is provided. Additionally, the details regarding the data used, the assumptions made and the steps taken to estimate the model specifically for the case of Amsterdam is explained. Hence, the findings of this chapter provide a structured approach to the simultaneous access and egress mode choice modelling for the different modes of public transport. This chapter addresses sub-question-3.

Chapter:5 Data Description and analysis
In this chapter descriptive analysis of the data utilized to estimate the model is explained. Additionally the pre-processing of the data to generate the alternatives and the attributes selected are discussed. The findings of this chapter proved the inputs to estimate the model.

Chapter:6 Discrete choice modelling
The results of the model estimated are depicted in this chapter. Furthermore, the analysis of of the results provides the insights that are gained from the model are discussed in detail. This chapter addressed the sub-question-4.

Chapter:7 Conclusions and Recommendations
The findings from the research, conclusion and recommendations based on the research done is explained in this chapter.

2

Literature review

2.1. Introduction

The goal of the literature review is to provide the insights into the current state of art. An introduction to the application of transport models is described. This chapter also provides the foundation of the conceptual model, given the current research for modelling of multi-modal transport. The methodological approach is discussed wherein the data used and the modelling method is described as well. The sub-question 1 is answered in this chapter. The sub-question 1 addresses the variables affecting the access and egress mode choice are addressed. The second sub-question is also partly addressed in this chapter that provides to determine the modelling approach that can be applied for the case study.

For literature review search engine such as google scholar and science direct were used. The key words used to find the relevant research were ’mode choice’ ’modelling mode choice’ ’access and egress mode choice’, ’factors affecting access and egress mode choice’. To filter the more suitable studies more recent research carried out within the past 5-10 years was considered. Additionally, as the case study is Amsterdam research carried out in the context of European region and the Netherlands were given preference.

2.2. Transport models

This section describes the application of transport models in practice. It also provides a basic understanding of the classic four-step model which is a popular modelling method used. Moreover, as the context considered is Western Europe is the main focus as the case study is more specific to his region.The focus is on public/private transport modelling of passengers. The national model also deal with the freight transport comprising of road, rail and inland transport. However, as the current research focuses on the travel behaviour of public transport passengers, freight transport is out of scope.

The purpose of a transport is to create a simplified version of the real world situation focusing on the elements that are significant for a particular problem (Ortúzar & Willumsen, 2011). The traditional model is a 4 step model; it comprises trip generation, trip distribution, mode choice, and assignment. Mode choice is the 3rd step in the modeling approach and requires input from trip generation, trip distribution (Ortúzar & Willumsen, 2011). However, the following module of assignment i.e. route choice has a feedback loop that is required to be considered. Thus, it is important to note that the route choice impacts the mode choice as well (Ton et al., 2020).
The focus of the current research is on travel demand domain. The travel demand is an extremely qualitative and differentiated aspect (Ortúzar & Willumsen, 2011). The current research addresses the multi-model passenger transport domain. As the name suggests multi-model transport has the characteristic that more than one mode is used to transport from one location to another.

One of the main objectives of a transport model is to deduce the feasibility of the investment made in the transport sector. Investment in the transport sector has a range of benefits from various aspects such as economic, environmental, and social perspectives (Transport for London, 2018). Transport models provide a more consistent evidence base for data analysis to deduce if a particular investment will be fruitful (Transport for London, 2018). Furthermore, they help create a more unified framework to access the policy implications (Rich et al., 2010).

2.3. Data and modelling approach

The review of the literature relevant to mode choice is searched on science direct, google scholar. The key words used are access egress mode choice', 'mode choice', 'modelling access and egress mode choice'. The table 2.1 depicts the data collection methods and the data used to apply mode and station choice models.

Table 2.1: Data and modelling approach

Author	Focus of study	Scope/Data collection system	Country	Main statistical approach
Krygsman, S., & Dijst, M. (2001)	The research focuses analysis of multimodal trips on the access and egress stage separately.	RP (NTS)	Netherlands	Multivariate analysis
Chakour and Eluru (2014)	To analyse train commuter behavior	SP	Canada	Latent segmentation
Arentze, T. A., & Molin, E. J. E. (2013)	The focus of the research is to address how the travelers trade off the travel costs ,travel time and inconvenience on mode choice	SP	Netherlands	MNL
Shelat, S., Huisman, R., & van Oort, N. (2018)	Analyze passenger profiles of the Bike – Train choice	RP (OViN Data)	Netherlands	Latent class cluster Analysis
Van Kampen et al. (2020)	Impact of the socio-economic characteristics, neighborhood characteristics on the station choice	RP (OViN Data)	Netherlands	MNL
Yang, M., Zhao, J., Wang, W., Liu, Z., & Li, Z. (2015).	Metro commuters satisfaction on the access and egress commuter journey	SP	China	Binary logistic regression model
Wen, C. H., Wang, W. C., & Fu, C. (2012)	Analyzing high-speed rail access mode choice.	RP	China	Latent class Nested logit model
Brands, T., De Romph, E., Veitch, T., & Cook, J. (2014)	Model Public transport route choice	RP	Netherlands (Amsterdam)	NL (Route choice) OmniTRANS (Simulation)
Azimi, G., Rahimi, A., Lee, M., & Jin, X. (2020)	The study investigates factors affecting the access and egress mode choice	RP (GFTS, GIS data)	USA (Orlando, Florida)	MNL (Access and egress separately)
Li, X., Tang, J., Hu, X., & Wang, W. (2020)	The aim of the study is to address the travel behaviour and the subjective factors that impact the decisions of the travellers.	SP	China	BMNL
Waerden, P. Van Der, & Waerden, J. Van Der. (2018).	The aim of this study is to address the factors that contribute to the decision for mode choice for access for medium and long distances.	SP	Netherlands	ML with panel effects
Yang, H., Feng, J., Dijst, M., & Ettema, D. (2019)	The aim of the study is to address the access and egress mode choice to/from High speed rail	SP/RP	China	MNL
Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2019)	Provide a mode choice model considering a more comprehensive mode choice set	RP (MPN Survey)	Netherlands	MNL, panel effects, MMNL
Ton, D., Shelat, S., Nijënstein, S., Rijsman, L., van Oort, N., & Hoogendoorn, S. (2020)	The aim is to investigate the factors that affect the joint decision of access mode choice and station choice for tram stops.	RP (GFTS Data)	Netherlands (Hague)	MNL

The table depicts which method of data collection is utilized for modelling access and egress mode choice. The modelling approach represents the statistical approach applied to model mode choice. The following subsection to discuss the methods of data collection and the modelling approach used in literature.

2.3.1. Data collection method

As depicted in the table 2.1, **revealed preference** and **stated preference** data collection is the most frequently used methods. The following paragraphs describe the above-mentioned methods and the pros and cons of each method.

The revealed preference surveys for travel behaviour research comprise of origin and destination passenger surveys. This data can be collected in the form of on board surveys, collected at station or conducted by the government. It suggests that the actual observed behaviour is captured in this case. In case of stated preferences the repentant is presented with background information and hypothetical alternatives are presented to the respondent. Based on what is the aim of the research the data collection approach varies. As shown in the table 2.1, in case of studies pertaining to modelling approaches or empirical analysis revealed preference data is used. Revealed preference data collection allows for the analysis of existing alternatives. As the name suggests it is the data revealed through the choices people make given that the options are already available to them. On the other hand stated preference data is collected in the form of an experimental survey setup (Train, 2002). The stated preference survey allows for the analysis of hypothetical situations. Subjects such as the perceptions of the users and attitudes towards situations can be captured.

However, both methods have their pros and cons. The revealed preference data collected reflects the actual choices people make. But, the disadvantage is that it is only applicable to the existing situations. Thus, it limits the choice situations and alternatives. In case of situations where the demand of the new product can not be examined using a revealed preference survey as such data is not available for such situations (Train, 2002). For a stated preference survey the advantage is that the respondent can be presented with hypothetical choice situation that don't exist. Additionally, the researcher has the freedom to design experiments to create as much variation as they require (Train, 2002). However,the drawback of obtaining responses from stated preference survey is that the respondent might not actually follow through. Moreover, there is also a chance of having biased responses based on what is understood by them and might not actually be applicable in reality (Train, 2002). Hence, both methods in a way, complement each other. As a consequence a combination of both methods is an ideal approach. As depicted in the table it is observed that that based on the objective of the study both reveled and stated data is used for research. In certain cases it a combination or additional data acquired by simulation. The approaches used in literature is discussed in chapter 3.

2.3.2. Modelling approach

The discrete choice modelling approach will be adopted so as to maintain the compatibility with the existing systems. As it is evident from the table2.1, that for the research purposes also Discrete choice modelling methods are more popular. It can be attributed to the fact the existing transport models (in the Dutch context) are discrete choice models are applied. The following paragraphs describe the different mathematical models used in literature for modelling access and egress mode choice

MNL (Multinomial Logit) models are adopted by a majority of the studies to analyse access and egress mode choice behaviour Azimi2020.Hence, the MNL model is considered suitable for deducing the access and egress mode choice. In the given study by Azimi et al., the access and egress are modelled separately. This model is chosen as it is considered that the alternative mutually exclusive. Moreover, the computation time is lower due to the closed-form mathematical structure of MNL models (Young et al., 2018). To deduce the mode choice for metro commuters, the access and egress mode choice MNL model is adopted for the four modes (i.e. subway, bus, taxi, car) considering car as the reference mode (H. Yang et al., 2019). Furthermore, study by Ton et al. to deduce the determinants for the use of active mode for access and egress mode choice where the MNL model is used to determine the significant variables as it has a more efficient computation time. Once the significant variables are used to determined, an MMNL (Mixed Multinomial Logit) is applied as the model fit is better (which is tested using the likelihood ratio test).

The study by Arentze and Molin implements a ML (Mixed multinomial Logit) framework to model travellers preferences towards multi-modal networks (Arentze & Molin, 2013). In the above mentioned study, to account for the correlated terms between the main modes a error component is added to the utilities of the alternatives having the same main mode. Thus, a 'shared error component' is added to the utility. This enables to account for the common unobserved attributes amongst the different modes (Arentze & Molin, 2013). Similarities on the level of access and egress stages are not taken into consideration. To have a parsimonious approach, only the important sources of covariance are considered in the model (Arentze & Molin, 2013). The added advantage of the error component model structure is that it provides more flexibility as compared to hierarchically nested model structures (Arentze & Molin, 2013).

To analyse the access mode choice behaviour to the high speed railway in Taiwan, a latent class NL (Nested Logit) model is adopted. Contrary to the popular choice of MNL models this approach is implemented as the latent class can provide further insights into the number, sizes and the characteristics of the segments (Wen et al., 2012). The latent class MNL model exhibits the IIA (Independence of Irrelevant Alternatives) property within the segments. The modelling was done for latent class MNL models and latent class NL models. As the goodness of fit is better for the likelihood ratio test at 5% significance level (Wen et al., 2012). In the research conducted by Waerden and Waerden, to deduce the significant attributes that affect the decision of the access and main mode choice, an ML (Mixed Logit) model is considered. The investigation of the individuals choice is carried out using a mixed logit (ML) model with panel effects. Such a model structure takes into consideration random taste variation in the population and the fact that a decision-maker can make more than one decision (Waerden & Waerden, 2018).

In the study by Li et al. the Bayesian-based model is adopted, as has higher accuracy. The advantage of this method is that it allows for complete uncertainty of the parameters through posterior distribution (Li et al., 2020). The added advantage of this method is that it avoids over-fitting issues and numerically intensive likelihood function maximization (Li et al., 2020). Some studies adopt the binary logistic regression model. In the study by Krygsman and Dijst the aim is to analyse the factors that impact the choice of multi-modal trips and it is analysed using the binary logistic regression model. Also, in the study by M. Yang et al. a similar modelling approach is implemented to determine the factors affecting the metro commuters satisfaction. In this study, the model for each predefined mode-chain, a model is estimated and the factors impacting the choice of the mode chain is analysed. Such a modelling approach estimates the variables by an iterative likelihood procedure Yang2015. In the study by Ton et al., station choice behaviour of cyclists is analysed wherein the MNL modelling approach is adopted to estimate the attributes that impact the station choice.

Thus, it can be observed that there are various approaches considered however MNL is the most prominent approach as it is simple and is computationally efficient. In the case of a situation that there is a correlation between the alternatives to allow more flexibility more advanced models such as the ML models are preferred. However it is computationally more intensive to implement ML models. However, the real-world validation is not possible at times so it is difficult to justify the use of more complex models (Young et al., 2018). To sum it up, many for the various combinations of travel choice dimensions such as mode choice (in this case) the extensions or variations of the MNL and the NL (Nested Logit) model are used (Ton, 2014). However, to account for the correlation within the trip chains for mode choice ML would be an ideal method of modelling mode choice.

2.4. Generation of alternatives

The choice set or the alternatives that are said to be available to the passengers are generated to analyse the observed behaviour and to predict the future choice behaviour Fiorenzo-Catalano, 2007. It is a complex process, the kind of complexities that require to be dealt with for the generation of choice sets for multi-modal transport model is further elaborated. To deduce the choice sets from an individuals perspective the individual is expected to have knowledge which is what is actually considered. Based on the observed choices and certain algorithms an objective choice set is generated. To generate these choice sets in addition to the observed choices, the behavioral criteria shall be considered. In the current research the focus is on mode expected to be available to the passenger Fiorenzo-Catalano, 2007. To determine the alternatives of the choice sets , the first step is to address the complexities to understand the challenges and the constraints that need to be considered when generating the choice sets. At the different stages of the modelling process there different kind of complexities to be dealt with. As depicted in the figure 2.1, three main aspects of the different complexities are encountered when modelling a multi-modal network. The first point deals with the choice set generation, i.e. the range and the the umpteen possibilities of the different combinations. Ideally it a multi-modal transport models is expected to predict the utilization all the modes possible. However, it is not feasible to consider them all, many combinations are highly unlikely (van Eck et al., 2014).

The second aspect is the mathematical complexity. Planning multi-modal trips have multiple choice dimensions associated with it, making it a challenge to determine the underlying behavioural traits in a tractable way (van Eck et al., 2014). For the third complexity it suggests at peak hours the use of the physical infrastructure and network loads effects the experienced travel time differently. The transfer related attributes are also required to be modelled separately.

The multi-modal modelling approach is an extension of the classical model (van Eck et al., 2014). The mode-chains are distinguished as separate modes. A filtration method is adopted to access the realistic options for the choice set generation. The size of the route choice set trips depends on the size of the network and the type of trips considered. The size of the objective choice set depends on the or is proportional to the network density(van Eck et al., 2014). Whereas the size of the subjective choice set depends on knowledge of the traveller of the available alternatives. Thus the challenging part is to generate the choice sets that represent the alternatives that are more likely to be available to the individual. To determine the mode-chains, the following definitions are considered;

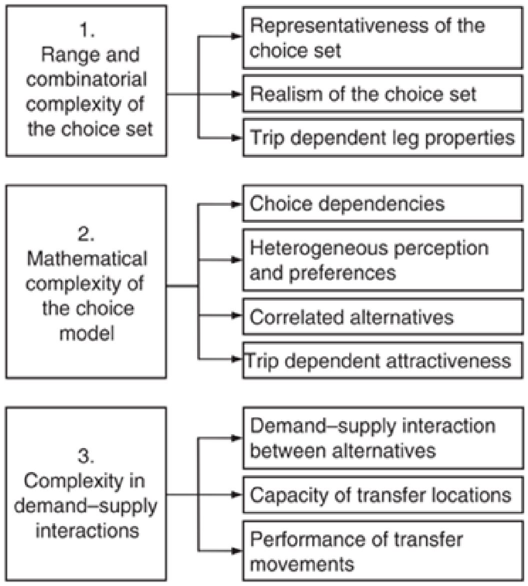


Figure 2.1: Modelling complexities (van Eck et al., 2014)

- Main trip part: It is the part of the multi-modal trip performed covering the highest possible distance comprising of one or more legs (Fiorenzo-Catalano, 2007).
- Trip leg: The part of the trip where a single mode is used without transferring is considered to be a leg. However, a walking leg with a mechanized mode leg is also considered a single leg (Fiorenzo-Catalano, 2007).
- Trip: It is considered to be a sequence of travel modes and transfer nodes connecting a given OD pair(Fiorenzo-Catalano, 2007).

Hence, to generate the different alternatives to fulfill the objective of the current research, literature pertaining to mode choice are reviewed and the choice sets generated are studied in more depth review the current practices. In order to determine mode chains studies adopt the frequently occurring combinations based on the given context. The most frequently occurring combinations are walking cycling car and bus (in descending order) of use as access modes. Walking is also preferred to access the different public transport modes. However, the car is mainly used to access train stations (Krygsman & Dijst, 2001). In the case of the egress side of the trips, there is a lower availability of owned vehicles. As a consequence of that, the findings suggest that the egress is dominated by walking and public transport modes followed by cycle (Krygsman & Dijst, 2001), in case of the Netherlands. In the case of the main modes, the most dominant mode is the train followed by the bus, tram and then metro StephanKrygsman. This trend is observed in 2,3,4 stage multi-modal trips (Krygsman & Dijst, 2001). Depending on the focus of the study the access and egress modes considered are more generic i.e. car, bus, tram, metro. However, some studies where the focus is on TNC (Transportation Network Companies) adoption the access and egress modes considered for the different alternatives are Walk, TNC/Taxi, micro-mobility, drive alone, carpool and wheelchair (Azimi et al., 2020). Moreover, depending on the data the pre-specified mode chains are considered. In the study by (M. Yang et al., 2015), the metro commuter satisfaction the following mode chains were considered. Based on the data collected and the availability of the observations the following alternatives; walk-metro-walk, walk-metro-bus, bicycle-metro-walk, bicycle-metro-bus, bus-metro-walk, bus-metro- bus, car-metro-walk (M. Yang et al., 2015). Similarly, in the study by Brands et al. the following mode chains were considered as alternatives for public transport mode choice assignment; walk – public transport– Walk, Bicycle –public transport – walk, car – public transport – walk, walk –public transport – bicycle, walk – public transport – car.

In the study by Arentze and Molin they develop different experiments for varying distances where the mode alternatives differ for the different distances. The experiments differ in terms of the mode alternatives available in the choice-set and the length of the imaginary trip (5, 20 and 65 km). The first experiment focuses on the short distance (5 km) and a choice between slow mode (bicycle) and uni-modal fast modes (car and bus) which are reasonable travel options for this distance. The second experiment focuses on the medium distance (20 km) where a car, public transport and multi-modal transport trips are possible modes for the trip. The third and fourth experiment assumes PT alternatives only (car and multi-modal are not part of the choice-set) for medium-distance (20 km) and long-distance trips (65 km) respectively. The public transport model, these experiments allow us to examine in more detail preferences related to service attributes (20 km) and modes in access and egress stages (65 km). When implementing the model on a case study the data is also considered to have enough observations in each alternative.The access and egress mode choices provided comprised of the subway(metro), bus, car, walking and cycle Yang2019. Thus, considering public transport as the main mode the possible alternatives ideally considering a set of access and egress modes as car, walk, cycle, BTM (Bus-Tram-Metro).

Hence, to generate the choice sets the as observed from literature two intra-modal transfers are considered when specifying a trip. Given the context of research, the access and egress modes used in combination with the main mode can vary. However, the generic access and egress modes such as walking, cycling, car and public transport modes are frequently used. In addition to data depending on the observed trends and data collected based on the observed behaviours also provide the input for the choice set generation.

2.5. Factors affecting access and egress mode choice

To determine the relevant variables to be considered when modelling access and egress travel behaviour the existing literature is reviewed and the factors that can be considered are classified and determined in this section. In a traditional four-stage model, mode choice is a classic step. It depicts the mode choice of individuals on a more aggregate level (Ortúzar & Willumsen, 2011). Hence, it plays a crucial role from the policy-making perspective (Ortúzar & Willumsen, 2011). Thus, to realistically model the user's choices, the attributes impacting the individual's mode choices must be considered. The mode choice factors are classified into three parts; the first is the characteristics of the trip maker (it comprises age, income, availability of car license, etc) (Ortúzar & Willumsen, 2011). Second is the characteristics of the journey such as trip purpose, time of day, group size, etc. The third is the characteristic of the transport facility travel time, distance, in-vehicle time, etc (Ortúzar & Willumsen, 2011). Moreover, these sections can be sub-divided into more classifications. The division of the different categories for mode choice can be made into three parts i.e. multi-modal travel variables, socio-demographic variables and residential environmental variables.The study by Yang et al (2015) carried out to determine the satisfaction of metro commuters divides the factors into the personal attributes and the journey details (M. Yang et al., 2015). Some studies consider the personal characteristics as a category and define the attributes per particular mode (H. Yang et al., 2019). However, in the case of mode choice involving active modes (i.e. cycling and walking), the factors affecting access and egress mode choice can be categorised as; characteristics of individual characteristics, household characteristics, built environment, season and weather characteristics, work conditions, and trip characteristics (Shelat et al., 2018; Ton et al., 2019). Similarly, in a multi-modal trip, the factors are allocated to categories of variables, such as socio-demographic variables, LoS (Level of Service) parameters, trip characteristics, land-use and built environment factors, and station characteristics (Chakour & Eluru, 2014). Thus, based on the approaches considered in literature; the following classification is made for the factors affecting the access and egress mode choice.

- Individual characteristics
- Household characteristics
- Built environment

- Work conditions
- Trip characteristics
- Station characteristics
- LOS (Level of service)

The following subsections describe the significant variables for the provided classifications based on the current state of art. Moreover, the variables belonging to the various categories;

2.5.1. Individual characteristics

Individual characteristics comprise factors such as age, gender, car ownership, income, level of education etc. Car ownership is a significant variable (Chakour & Eluru, 2014), if the car is available it affects the choice of public transport as the main mode of travel (Hensher & Reyes, 2006; Mark et al., 1993). Moreover, it is an obvious observation that the availability of cars impacts the choice of the car as the main mode and the share of taxis is very low that it is not considered significant in the result of the study conducted (Mark et al., 1993). Depending on the mode and the variables such as the travel time and distance, socio-demographic variables such as gender and more importantly age give a deeper insight into the travel behaviour (Van Kampen et al., 2020). Hence, age and income also have an impact on the mode choice but it depends on the context (H. Yang et al., 2019). Thus, it can be concluded that the significance of the socio-demographic variables is more context dependent.

2.5.2. Household characteristics

Size of the household, household composition, income of the household , the number of vehicles in a household etc. are included in this category (Shelat et al., 2018; Ton et al., 2019). The household characteristics affect the probability of selecting a more complex trip (Hensher & Reyes, 2006). It also implies that is is more likely to be a trip involving the use of public transport (Hensher & Reyes, 2006). The household size and composition are significant and have a negative correlation to the use of public transport (Hensher & Reyes, 2006). It is expected as larger household with children affect car ownership, also the household income determines if people have accessibility to more expensive trips, i.e. (long) trips by car or train.

2.5.3. Built environment

The urban density and the high density of the mix of activities have an influence on the demand for multi-modal transport (Mark et al., 1993). The urban form at the macro level indicates that business travellers who are more sensitive to time tend to choose a car instead of the bus to access built HSR (High speed Railway stations) stations in the suburban areas with low urban density (H. Yang et al., 2019). Especially in the utilization of active modes the built environment has a more crucial role (Ton et al., 2019; Van Kampen et al., 2020). Whereas when the mode of egressing is walking, the commuters using the motorized modes for access are more concerned about the walking environment (M. Yang et al., 2015).

2.5.4. Work conditions

Full-time employment, paid and unpaid job status is included in this category (Mark et al., 1993). Studies suggest that employment affects the choice of the access and egress mode (Shelat et al., 2018). A permanently employed person is expected to have the location as per the job and is more probable to use non-motorized modes (Mark et al., 1993). Additionally, despite having the car available, the group of people living in urban areas choose the transit – bicycle mode as the preferred mode to reach job locations in urbanized areas (Shelat et al., 2018). Moreover, in the Dutch context highly educated professionals are the largest group selecting the bike–train combination (Shelat et al., 2018) In addition to that, the provision of incentives discouraging cars induces the use of public transport (Ton et al., 2019).

2.5.5. Trip characteristics

Trip characteristics comprises the access egress travel time, distance, travel costs, trip purpose, size of the travel group (**Ton2019a**). Trip purpose or the motive to travel plays a significant role in the choice of the access and egress mode (Mark et al., 1993). Hence, it is deduced that the mode choice preferences for access and egress are affected by trip purpose (H. Yang et al., 2019). The trip purpose can be classified into mainly business and leisure purposes (H. Yang et al., 2019). Business travellers value time higher and are inclined to pursue shorter travel times. Whereas, leisure travellers are more willing to accept longer travel times (H. Yang et al., 2019). The highest share of travellers in the Dutch context is the work commuters and educational purposes (Givoni & Rietveld, 2007). The results show that the VOT (Value of time) is strongly linked to trip purpose (Mark et al., 1993). The willingness to pay is higher in the case of business or leisure travel, as compared to a systematic travel purpose such as work (Mark et al., 1993). Travel distance is an important variable for access and egress mode choice. Higher the walking distance lower is the chance of selecting a slow mode such as walking for access. The access and egress station choice is a decision affected by the distance negatively, hence the Dutch population is willing to go to a further station (Krygsman & Dijst, 2001). However, travel time is is expected to be more appropriate as compared to travel distances (Young et al., 2018). Thus, travel time is expected to be the most influential attribute (Ton et al., 2019). It can be further broken down into waiting time and in-vehicle time (Arentze & Molin, 2013; Waerden & Waerden, 2018). Waiting time is perceived more negatively as compared to the in-vehicle time (Arentze & Molin, 2013). Additionally, the travel costs such as the parking are influential attributes (Waerden & Waerden, 2018; Wen et al., 2012). Furthermore, from a station choice perspective cyclists prefer to use the distant but larger station as compared to a close-by station. This is so a traveller prefers to cycle further to avoid transfers (Jonkeren et al., 2019). The findings suggest that the travel group size and the moment of travel are relevant for mode choice (Ton et al., 2019).

2.5.6. Station characteristics

Factors such as the availability of bike parking facilities are influential for access and egress station choice (Krygsman & Dijst, 2001). Bicycle parking facilities that are perceived as having a higher quality have a stronger impact on station choice (Young et al., 2018). Additional services at larger stations are also expected to have a positive correlation with the station choice. One of the findings indicates that the presence of bike parking facilities during access results in a higher willingness to cycle further. Consequentially, walking distance is weighted more negatively as compared to cycling (Ton et al., 2020). Hence, when walking the closer station is preferred by the passenger as compared to using cycle as a mode. Intercity and sprinter status of the station also has an impact on the station choice for access and egress. In terms of the station location based on the usage station leading to lesser transfers and IC (inter-city) stations are favoured even if the distance is longer (Van Kampen et al., 2020).

2.5.7. LOS (Level of service)

The factors such as the quality of the service provided, punctuality, frequency of travel etc are considered. The assessment of bus lines portrays that passengers consider the service quality of public transport but also access and egress part of the trip in their decisions (Brand et al., 2017). Additionally, the station attributes such as the parking availability and the seat availability have an impact on mode choice (Chakour & Eluru, 2014). The frequency of the mode. The on-time performance of public transport is an important factor affecting the choice of public transport in terms of user satisfaction (M. Yang et al., 2015). The evidence indicates that station utility decreases as the access journey becomes further or longer, as the rail leg journey time increases, when the journey involves more transfers or has a higher fare, and when service frequency is reduced (Young et al., 2018).

2.6. Summary

To create a framework and estimate a model the input required is the data, choice set or alternatives and the attributes that shall be estimated. The literature review provides insights into how the modelling process is carried out and what is done as per the current state of the art to determine the inputs for modelling.

It is evident that for compatibility sake it is suitable to adopt the discrete choice modelling methods. additionally , the classic four step method is considered. The aim of the literature review is to deduce the parameters that would affect the mode choice behaviour. However, there are studies addressing mode choice modelling world over and also in different contexts. The following table depicts the different factors that can be considered as for the mode choice modelling.

- Findings suggest that the influential factors are as follows;

Table 2.2: List of attributes from literature

Category	Probable Attributes	Study
Individual characteristics	Employment, age, gender, license, income, ownership of cars, availability of free student travel, frequency of travel using various modes.	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Ton et al. (2019), Ton and Shelat (2020)
Household characteristics	Household size, number of vehicles per household	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Ton et al.(2019)
Trip characteristics	Distance, travel Time, Purpose, frequency of travel via different modes, origin and destination locations.	Mark R. et al. (1993), Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014),Yang et al. (2015), Ton and Shelat (2020), Van Kampen et al. (2020)
Level of service variables	Frequency, number of stops, in vehicle time, waiting time	Mark R. et al. (1993), Debrezion et al.(2007), Yang et al. (2015), Ton and Shelat (2020), Brand, et al. (2017)
Built environment	Area or region, Urban density	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Van Kampen et al. (2020),Ton and Shelat (2020)
Working condition	Working hours, Travel compensation	Van Kampen et al. (2020), Ton et al. (2019)
Weather characteristics	Month of travel	Ton et al. (2019)
Station choice	Origin and destination station, area region, type of stations, parking facilities	Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Van Kampen et al. (2020), Ton et al. (2019), Ton and Shelat (2020)

- Ideally the alternatives generated consider car, walking , cycling ,BTM as the complementary modes for access and egress resulting to the following combinations for access and egress to and from public transport. Most studies have a more generalised approach. The process of deducing the choice set depicting the alternatives is a complex process. In order to obtain feasible alternative assumptions based on the observed behavior is an essential input.

- Most commonly used approach for modelling is the MNL model. However, to get a better model fit it would be preferred to adopt more advanced models such as the ML model. Simultaneous modelling of mode choice is expected to lead to alternatives that are expected to be correlated. Hence in this case it would be effective to take into account the correlations that exist between the different alternatives. Furthermore the limitations of the MNL model can be overcome by the implementation of more advanced models. However it must be taken into account the trade off using more complex models.

3
Conceptual framework

3.1. Introduction

The goal of this chapter is to use the findings from the literature review chapter as input, to prepare a framework that shall be applied on a conceptual level. Furthermore, it forms the basis of the framework that shall be applied to the case study of Amsterdam. The sub-question 2 is addressed in this chapter.

3.1.1. Data Collection

As discussed the chapter 2, the popular methods for data collection are revealed and stated preference surveys. In this section the data used in literature is discussed in depth.
Revealed preferences comprise the observed behaviour, such data is usually collected in the form travel surveys. In the context of Netherlands a National Travel survey is carried out on a yearly basis. Revealed and stated preferences are used by various studies to deduce the mode choice behaviour. As depicted in table 10, most studies selected are based on travel surveys such as travel diaries. NTS (National Travel Survey), OViN (Onderzoek Verplaatsingen in Nederland) and ODiN (On Der Weg in Nederland), MPN (Mobiliteitspanel Nederland) are national travel surveys carried out every year in the Netherlands. Study Ton et al. adopts the MPN surveys that are household, personal travel diaries as. It is a longitudinal household survey that is carried out to investigate the changes in the travel patterns of individuals including a younger population of teenagers (Ton et al., 2019). It is a 3-day survey where the respondent is expected to record the trips, mode, distance travelled. Moreover, the household survey comprises is related to the availability of the modes and the ownership of vehicles. Whereas the personal survey also captures the mode preferences and attitudes towards them (Ton et al., 2019).

ODiN or OViN is a National travel survey carried out every year and captures a one-day travel trip. The aim of this survey is to provide insights into travel behaviour. It is a one-day travel survey that comprises trip characteristics, mode choice, distance, time etc. It does not represents individuals. It is conducted by the bureau of statistics (CBS). It was previously OViN and since 2018 it is renamed as ODiN. The study (Shelat et al., 2018) uses the is data to analyse the bike-train users and their behaviour. NTS is an older version of the current ODiN travel survey carried out. The objective of this survey is to determine the travel behaviour characteristics of the Dutch population. In the current survey format, children below the age of 12 are not considered as a part of the survey sample. The study (Van Kampen et al., 2020) use the OViN data as well to determine the effects of socio-demographic variables on the station. However, to deduce the alternatives available, the origin and destination pincode corresponding to the first, second, third and the fourth closest station in every departure postcode is extracted from the OViN dataset and the stations present within the 10km radius are selected.

For deducing the factors affecting the simultaneous access mode and station choice GFTS i.e the General Transit Feed Specification data is used. GFTS data is utilized to generate the different routes to deduce different route alternatives for the routes between the different stations. The best routes algorithm is adopted to deduce the routes between the stations based on the number of transfers, transit time etc (Ton et al., 2020). It is observed that for studies comprising route choice as a part of the scope utilize simulation softwares to generate the routes. In the study Azimi et al. the LYNX data collection system is utilized, it is an on-board survey collecting the data of the passengers in conjugation with the GIS (Geographic Information System) data for the origin and destination location for home addresses (Azimi et al., 2020). The socio-demographic information has been extracted as well. Furthermore, trip information the origin and destination location, access and egress mode etc. was also collected. Similarly, for the study by Chakour and Eluru, an on-board survey on the rail for commuters was carried out to investigate the access and egress mode choice. Data associated with trip characteristics, socio-demographic variables such as the individual characteristics, boarding and alighting stations, departure time etc is collected. To deduce the access and egress time and distances the google map algorithm was utilized to obtain more realistic values. A similar approach is used to deduce the values for the chosen and non chosen alternatives. Simulation software such as GIS is adopted to deduce the origin locations next to streets and roads were geo-coded finding corresponding GIS data (Chakour & Eluru, 2014).

In the case of stated preference, the questionnaire is designed so as to address the perceptions of the users. To answer specific questions related to the perceptions of the stated preference survey is usually conducted. In the study to deduce the metro commuter satisfaction with respect to the whole trip including the access and egress experience, the survey is designed and circulated within the region (H. Yang et al., 2019). In the study to deduce the trade-off of distances (Arentze & Molin, 2013), 4 experiments have been carried out as all modes were not feasible for all alternatives. In this case, a panel is selected in the Netherlands to fill the survey as per the traveller’s preferences.

Hence, overall for modelling purposes, the use of Revealed preferences and simulation is more common so as to realistically model what is actually observed. The stated choice experiments provide more insights into the different factors affecting mode choice based on particular situations. Moreover, the trade-off between the different factors is also analyzed. Thus, it is observed that the revealed preference surveys are the most preferred options for data collection. Depending on the scope of the study the revealed and stated preferences are adopted. In most studies travel surveys are used, they comprise of the socio-demographic data, data related trip purpose etc. The travel surveys are revealed preference data as the purpose is to record the travel behaviours of the passenger. The ODiN data-set is described in detail in the following chapters. It is preferred as the data source as observed behaviour is required to model the choices made. Moreover, the case study is Amsterdam, for the existing traffic model used by the Municipality of Amsterdam the data source is the ODiN travel surveys. Hence, making it a more compatible option as compared to other data sources.

3.2. Modelling approach

In the current research the aim is model mode choice behaviour simultaneously for access and egress mode choice. Thus, it is necessary to generate feasible alternatives and deal with the complexities associated with the choice set generation. The following section depicts the steps taken into consideration to generate the alternatives.

3.2.1. Generation of Alternatives

To determine the alternatives for the modes studies consider the mode specifically only, but in order to model simultaneous mode choice mode chains shall be constructed to generate alternatives. As discussed in chapter 2, to generate a choice set the feasible alternatives shall be considered. It is assumed that the individual is aware of all the alternatives and makes the choice accordingly. The complementary modes to and from public transport are car, bicycle, walking and BTM (Bus-Tram-Metro). As mentioned in Chapter 1, the framework shall be transferable to bus, tram and metro networks. Hence, the model will be tested for the train and metro network available in Amsterdam.

As shown in the figure 3.1 for a multi-model trip, walking is inherently a part of all trips made. It may be walking to the public transport stations, walking up to the car or cycle parking, transferring between the different modes. Hence, it can be said to be a universal component of any trip. The multi-modal trip composition is complex consisting of a series of different travel legs connected by walking legs (Fiorenzo-Catalano, 2007). To simplify it the trip can be considered to have the main trip part, and the

part connecting the main mode and the origin or destination node is considered as the access and egress trip legs respectively.
As the approach applied here is an extended form classical model, the mode chains are pre-defined

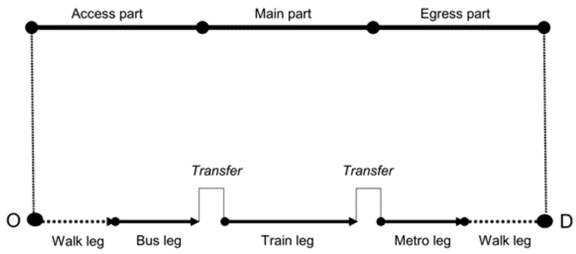


Figure 3.1: Access - Main - Egress trip composition (Fiorenzo-Catalano, 2007)

(Fiorenzo-Catalano, 2007) to generate the alternatives. Thus, the mode chains such as walk-train-walk etc. form an artificial mode itself. Though it provides a way to integrate the access and egress modes, issues are expected to arise from the implementation as there is an overlap expected between the alternatives. If the number of transfer nodes increases to more than two the number of alternatives explodes (Fiorenzo-Catalano, 2007). Thus, such an assumption creates a limitation for the number of alternatives that can be generated is limited. However, the aim is to consider the commonly considered mode chains (Fiorenzo-Catalano, 2007).

Hence, the generic modes can be used to determine the mode chain and the number of intra-modal transfers considered is 2. However it is expected to vary with the context. As the scope consists of Amsterdam, the multi-modal trips to, from and within Amsterdam are considered as a part of the data-set. As the idea is to model multi-modal trips the modes can be classified and filtered from the data-set in the following manner; As the case study is focused on an urban region. The generic modes i.e. walk,

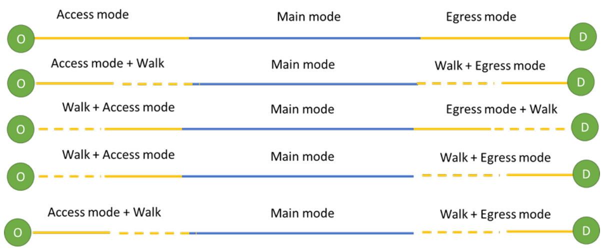


Figure 3.2: Filtration of trips

cycle, public transport (Bus-Tram- Metro (BTM)) and car are considered the complementary modes to public transport. However, for all practical purposes, it is viable to have all alternatives with a substantial amount of observations.
Ideally, the following are the alternatives generated;

- Walk- PT-Walk
- Walk-PT- Bicycle
- Walk -PT- BTM
- Walk- PT- Car
- Bicycle - PT-Walk
- Bicycle -PT- Bicycle
- Bicycle -PT- BTM

- Bicycle -PT- Car
- BTM - PT-Walk
- BTM -PT- Bicycle
- BTM -PT- BTM
- BTM - PT- Car
- Car- PT-Walk
- Car -PT- Bicycle
- Car -PT- BTM
- Car - PT- Car

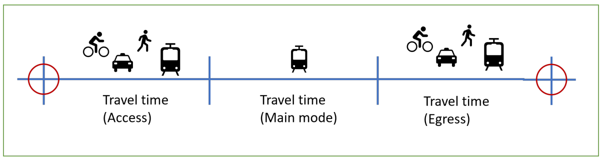


Figure 3.3: Alternatives

As depicted in the figure3.3, to generate the alternatives the main modes are considered based on the limitation of the data as only the boarding and alighting train stations are provided with a model for trains and to deduce the transfer-ability of the method to lower-level networks the metro stops are deduced.Hence, additional data sources and data extraction process is required to deduce the values for the chosen and non-chosen alternatives and to address the shortcomings of the data available. Description of the data pre-processing is provided in the following chapters.

3.2.2. Derivation of travel-time attribute

Based on the literature review and the requirements of the Municipality of Amsterdam, the attribute that varies with the alternative chosen is travel time. Travel time is intuitively more appropriate, as the time taken to get to a station is likely to be more important than the distance travelled (which may be unknown) (Young et al., 2018). It is to be noted that the public transport catchment area depends on the total trip time and not only the access and egress time (Brand et al., 2017; Krygsman & Dijst, 2001). In the case of travel time, the access and egress travel time is considered as the duration of the main mode is expected to remain constant throughout the different alternatives. However, the travel time adding the main mode will be considered for the whole trip. Thus, travel time (for the whole trip, access time, egress time) is the only variable that varies as per the selected alternative. As discussed in Chapter 2, there are umpteen factors extracted from the literature, in the interest of time and to depict how the model works the most relevant factors are considered. For the given case study the other attributes that shall be considered are described in chapter 4.

3.3. Application of the model

As suggested in the literature review chapter discrete choice modelling methods are preferred as they are more often used for research. Moreover, in practice the classic four-step modelling approach that widely implemented for modelling purposes. Though there are more advanced methods such as activity based modelling and agent based modelling. Activity based models are the more advanced version of the four-step model as they add the link between the activities carried out by the individual to the travel. It allows for the incorporation of individual and household level attributes. The challenge of this method is detailed data requirement. Additionally, as the choice facts are linked, secondary effects may occur. Especially, the activity categories are more sensitive to the demographic variables.

Agent based modelling approach is a microscopic framework that allows for the integration if different transport models as location choice, mode choice, car ownership, land use models etc (Kagho et al., 2020). However, it is utilized for research purposes. The challenges with this method is the data requirement. As detailed data is r pertaining human behavior is required and so it is an expensive process. Also, it is more computationally intensive. In addition to that there are no known cases of re-produceability of this method (Kagho et al., 2020). Thus, due to such constraints of advanced modelling methods Discrete choice modelling methods are preferred in practice.

In the following section the modelling approach that will be adopted and implemented is discussed in detail. Furthermore, it describes which discrete choice model will be applied and the procedure to implement the modelling technique.

3.3.1. Discrete choice models

As discussed in chapter 2, there are various methods used in the existing literature. In this section the methods that are considered to be suitable for the current research are discussed in depth in this section. The introduction the utility maximization that is the decision rule of utility maximization is explained in the following paragraphs. As the selected methods to analyse the models are the MNL modelling and ML modelling, the MNL model is explained briefly. Additionally, the more advanced model such ML (Mixed Logit) model is discussed.

The discrete modelling approach is utilized for modelling is based on the random utility theory. Based on the data the individual is expected to make a choice from a set of alternatives that is said to maximize their utility. The utility is determined by deducing the attributes that affect the utility and affect the choice of the individual. Moreover, the utility comprises the attributes that are the observed components, and the second part comprises the unobserved component which is treated as the random component. Hence, the utility is expressed in the following manner (Ortúzar & Willumsen, 2011; Young et al., 2018).

$$U_{jq} = V_{jq} + \mathcal{E}_{jq}$$

J is the selected alternative by the q individual Here, the U_{ni} is the utility of alternative, V_{jq} is the utility measured based on the attributes and E_{ni} is the unobserved utility. The expression for V_{jq} is as follows;

$$V_{jq} = \sum_{k} \theta_{kj} x_{jkq}$$

In the case of the MNL model, the parameters are calculated based on the maximum likelihood principle. When carrying out the estimation of parameters that are the betas, the MNL model iteratively finds the combinations of beta's that makes the given data most likely (Young et al., 2018). It is the simplest form of a model. It is generated assuming that the random residuals are IID (Independently drawn from distribution with same variance) Gumbel distributed (Ortúzar & Willumsen, 2011)). Thus, the choice probabilities are calculated in the following manner;

$$P_{jq} = \frac{exp(\beta V_{jq})}{\sum_{Aj \epsilon A(q)} exp(\beta V_{jq})}$$

As depicted in Chapter 2, the most commonly used approach is the MNL model. Hence, it is as a starting point to start modelling, due to the closed form and the relatively low computation time (Ton, 2014). However, there are certain assumption with MNL models that does not lead to accurate results always. One is that the model assumes that the unobserved components of the utility of the different alternatives are independent of each other. Hence, showcases the independence from irrelevant alternatives (IIA). Thus, this ends up being the weakness of the model as in reality this behaviour is not exhibited. Thus, the trade off for the simple structure and the low computation time is that the realism in the model is compromised (Train, 2002).

To address this weakness various different versions of the MNL model such as the Mixed MNL model is applied (Ton et al., 2019). Furthermore, there are NL (Nested Logit) models are also applied often. The NL model relaxes the assumption that the alternatives are independent and groups the correlated alternatives a-priori. This is expected to account for the unobserved correlation between the different alternatives within the same nest. Hence, the modelling step has two steps. One assigns a marginal probability of an individual choosing a particular nest at the upper level that applies to the group of alternatives within the same nest. The second one is at the lower level that predicts the probability of choosing an alternative within the nest.

ML (Mixed Logit) with error component structure has a flexible error structure and is theoretically able to reproduce the same structure as both the CNL (Cross Nested Logit) and NL (Nested Logit) models (Train, 2002). It has an added advantage in that it is able to incorporate heterogeneity and heteroscedasticity that is expected to be present in the given sample (Ton, 2014). The CNL model is an extension of the NL model. It allows for the correlation for not only within the nests but also across the alternatives in different nests. In addition to the parameters existing in an NL model, α is introduced in the equation. It results in a medium computation effort (Ton, 2014).

More complex models such as the ML model have a better modal fit. The ML (Mixed Logit) model has an added error component (Young et al., 2018). It counters the limitations of a standard logit model as it allows for random taste variation substation patterns and correlation in unobserved factors, unrestricted (Train, 2002). However, it is computationally intensive as it is a simulation-based model. As the computer speed and understanding of simulation models improves the utilization of the ML models has also increased. Hence, they are more complex models (Young et al., 2018). ML models can be estimated for various behavioural specifications. The ML model can be defined based on the functional form for the choice probabilities. ML probabilities are the integral of standard logit probabilities over a density of parameters. The choice probability is described in the following manner;

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d(\beta)$$

HereL_{nib} is the logit probability evaluated at parameter b , fb is the density function.

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{nj}(\beta)}$$

V_{ni} is the depends on parameters b. In case the utility is linear in B then V_{nib} =b'_{xni}. The ML has the following form;

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

ML is the mixture of the logit function for the different and the f_b as the mixing distribution (Train, 2002).

Hence, it can be interpreted that MNL models are faster to compute and hence can be used to decide the significant variables. Furthermore, more complex models such as the ML model can be used for the estimation of the model using the significant variables to obtain a better modal fit (Young et al., 2018). However, in case of complex models, it is necessary to compare their predictive performance with simpler models so as to access the trade-off between the complexity and performance of the model (Young et al., 2018). Moreover, the ML models capture more realistic substitution patterns and taste heterogeneity. To conclude it is preferred to carry out numerous iterations using a simpler model as it is more time efficient. Further, optimise the model by implementing a more complex model and use the parameters such as the likelihood ratio test, AIC and BIC values to justify the model performance.

3.3.2. Assessment of model performance and Validation

Model performance is often accessed analysing the parameters parameters such as the likelihood ratio test, adjusted Mcfadden value, but is only useful to compare models estimated using identical samples (Young et al., 2018). One way to validate the model is to carry out the log-likelihood test. The likelihood ratio test can be carried out to determine if the model is a better fit statistically. It is useful to validate the model to a certain extent but is also useful to compare the different models. It is depicted as follows. Where in the LL_u is the log-likelihood of convergence of the unrestricted model and the $LL_b r$ is the log-likelihood of the restricted model.
The other way to validate the model is more extensive method is by using similar but independent data by splitting the data into equal parts and hold out sample data. The sample that is held out is used to measure the performance (Ton, 2014; Young et al., 2018). This method is expected to deduce the predictive accuracy better.

3.3.3. Implementation of the model

In order to estimate the model Biogeme (software package) is used to estimate the models (**Ton2019a**), (Van Kampen et al., 2020). It adopts the likelihood estimation techniques to estimate the coefficients. Using the choice sets deduced, an iterative procedure is implemented to obtain the final model. As the MNL model is computationally efficient, so the first step is to analyze the model using MNL formulation to deduce the relevance. It is an iterative process wherein the different variables. As travel time is an important variable that is then introduced as access and egress variable separately. The following steps are carried out to estimate the model; The first step pertains to the input of the different alternatives based on the observed behaviour extracted from the data. Based on the the current approach in literature and the requirement of the modelling the different alternatives are the mode chains with an access- main - egress leg. The whole mode chain combined forms a trip. As suggested in literature , travel-time is an alternative dependent parameter and is significant. Hence, to have a more structured approach and understand the weight associated with the each trip leg i.e. access, main ,egress is estimated. It is expected that the travel time values have a significant impact on the mode choice, further it is the only parameter that varies for each alternative except for the main mode.

The next step is to introduce the socio-economic variables by formulation of hypothesis and test the significance of the parameters. The value of the socio-demographic variables is constant irrespective of the alternative selected as the attributes associated to the individual remain the same. In the following chapters the process of modelling is explained in more detail. Such an approach is adopted to deduce the impact of the independent variable that is travel-time which varies as per the mode choice. Furthermore, the objective of the model is to be able to deduce the impact of certain policy implications based on the socio-economic variables. Additionally it is expected that not all the socio-demographic variables will have a significant impact on all alternatives. Thus, it provides a more structured way of modelling and obtain more meaningful insights. Adding all the variables in one model is usually preferred in case of a stated preference survey. However, in this case randomly adding all the variables at once makes it difficult to deduce the underlying behavioural trait.

Finally, to deduce the final model a ML model will also be analyzed and compared with the MNL model. Furthermore, as the models are analyzed for train and metro as the main mode there is an additional comparison made between the different modes of public transport and how the policy implications remain constant or differ in both cases.

3.4. Summary

The following is the flowchart of the methodological approach to implement the model for the case study of Amsterdam.

As depicted in the figure 3.4, the first step is to pre-process the ODiN data to form mode-chain. The mode-chains form the alternatives. From the literature and the requirement of the city of Amsterdam four generic modes are considered for access and egress. It is a combination of private and public modes i.e. walk, cycle, car and bus-tram -metro (considered as one public transport alternative). Once the mode chains are deduced, the travel time for all the chosen and non chosen alternatives divided into access, main and egress travel-time. It should be noted that the waiting time and in -vehicle

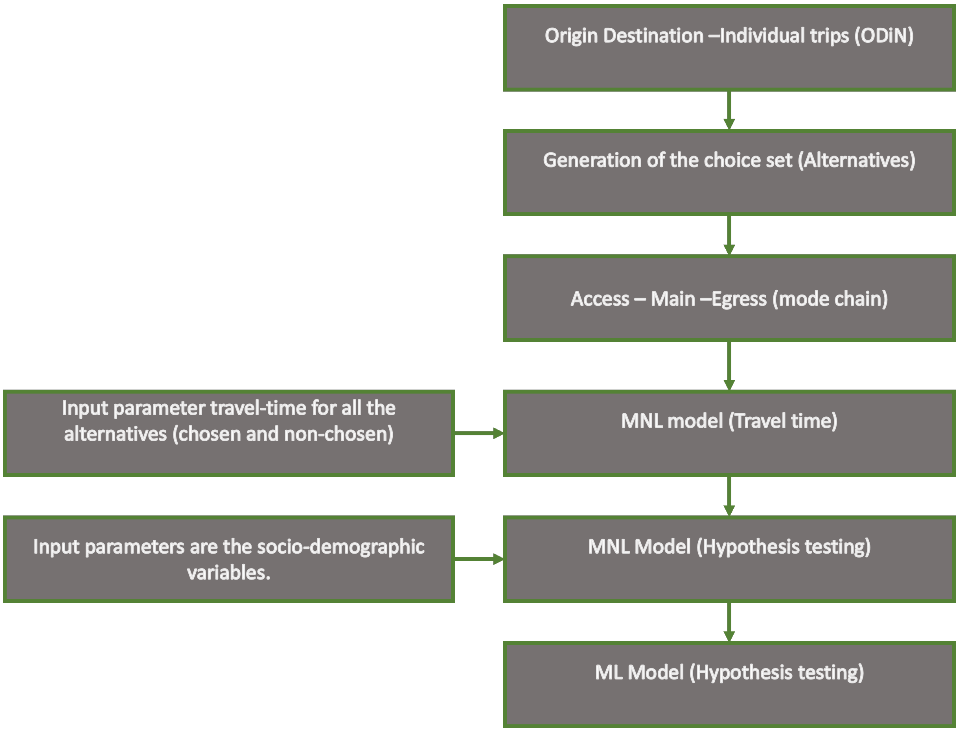


Figure 3.4: Flowchart of the method

time is assumed to be added in the total value itself. A base model is deduced using travel time as the only parameter using MNL modelling theory. The hypothesis for the selected socio-demographic variables will be tested individually. The significant variables will be considered for the final model to be analysed. Once, the MNL model is deduce a mixed logit model with the random error components to the correlation between the different alternatives will be estimated. As the model is to be applied to the train and metro network and the results of both models will be compared.

4

Modelling Approach : City of Amsterdam (Case study)

4.1. Introduction

The objective of this chapter is to explain the method implemented to model and analyse the data. A more detailed explanation of the method applied for the case study of Amsterdam is described. The sub-question 3 is addressed in this chapter. The results and the analysis of the data is explained in the following chapter 5.

4.2. Case study : City of Amsterdam

4.2.1. Public transport networks in Amsterdam

The number of inhabitants in Amsterdam is about 850,000. Including the regions surrounding Amsterdam the number of inhabitants is approximately 1,350,000. The area covered $250km^2$ (Brands et al., 2020). There are mainly there services existing in the region of Amsterdam for the public transport namely the train, bus, tram and metro. As it is a highly urbanized region there are more services available. The train network comprises 10 station within the city. The metro network has 52 station and 4 lines forming a network that is 41 km long. The bus network can be divided into the bus and night bus network consisting of 43 lines (32 regular bus / 11 night bus). The tram network has 15 lines and 500 stations forming a rail network of 80.5 km long [1].

4.2.2. Model used by City of Amsterdam

The transport model of the City of Amsterdam is in principal based on the NMS (National Modelling System) but is modified to better accommodate the policy issues at an urban level. It is a dis-aggregate model that consists of the socio-demographic characteristics of the users as well.
Focusing on the mode choice perspective, the destination choice and the transport mode are modelled simultaneously. The level of service associated with the alternative depends on the time, cost and distance associated with the alternative selected. The objective of the model is to predict the combined choice of destination, travel period and means of transport. In the case of public transport, cycling and walking are considered to be complementary modes of transport. Moreover, the tour frequency and mode choice are modelled in a similar way to the NMS (National Modelling System) model.

Multi-modal trips are modelled for all public transport trips, by assuming that there are 8 possible PT trips in the mode choice for PT. Car-bike, car-walk, bike-car, bike-bike, bike-walk, walk-car, walk-bike and walk-walk are the access and egress combinations for this model. The public transport assignment gives the different level-of-service variables for all eight combinations.

[1] https://amsterdammap360.com/amsterdam-tram-map: :text=Amsterdam %20Tram%20Maptext=Amsterdam%20tram%20is%20a%20transit,(80%2C5%20km

4.3. Data

In this section the data-set selected for analysis i.e. ODiN survey described in detail.

4.3.1. Data collection

Predominantly, there are two forms of data collection i.e., Revealed and stated preference. Both are analyzed based on choice models Keuchel and Richter, 2011. The limitation of revealed preference data is that the alternatives have a high possibility of being strongly correlated Keuchel and Richter, 2011. Thus, experimental designs allow for more control over the covariance structure of the attributes of the choice alternatives. However, there are many attributes there will end up being biased results (Keuchel & Richter, 2011). Thus, it is necessary to consider the limitations of the different forms of data collection. Additionally, as the case is the city of Amsterdam and as discussed in the previous chapters the ODiN data is used in the NMS model and the traffic model of Amsterdam is based on this data. Hence, it is considered to be the most suitable for the current research. For the sake of compatibility to the existing model used by the Municipality of Amsterdam, ODiN data is used to deduce the various sub-models and calibrate the data, it would be the most feasible to utilize the ODiN data.

Starting from 1978-2003 it was known as the Travel Behavior Survey (OVG) and was conducted by the Central Bureau of Statistics (CBS) [2]. For the year 2003 the survey was conducted by Rijkswaaterstraat known as AVV (Traffic and Transport Advisory Service, part of Rijkswaterstaat). Later on the MON (Mobility Research Netherlands) from 2004-2009. Then it was taken over again by CBS and known as OViN (Relocation Survey in the Netherlands) up to 2017. Since 2018 the relocation survey has changed and is known as ODiN (Onderweg in Nederland)[2]. The main difference between the OViN and ODiN is that the sample of the ODiN data does not comprise respondents younger than 6 years old [2]. The following paragraphs describe the process of data collection for the ODiN are carried to collect data:

Sampling

The target groups are pre-specified as per the income, age, and migration background. The different response probabilities and sample fractions are drawn such that that the final response is more balanced. This is an additional process as compared to OViN. Thus, the objective is to make the sample more representative [2].

Data collection

The fieldwork to collect the responses is done via an internet questionnaire only. Whereas in OViN it was carried out via mixed modes. To deduce the impact of the mixed responses there is a simulated questionnaire circulated and the difference between the responses is deduced. Additionally, the respondents are incentivized to respond by promising a prize for a few lucky winners. Further, it is designed to be adaptable to be filled using tablets. It is a location-based survey wherein the relevant travel journey information is requested via different locations. The respondents are assigned a day to fill in the survey. Moreover, the holiday trips are also recorded as a part of the survey which was not the case with OViN. The aeroplane rides are deleted. In the case of the distances covered by train, the values are deduced using a standardized distance table depicting the travel distances between the different stations [2].

Processing

The data collected is processed so as to remove all the previous errors. One of which is the correction of the over speeds. In the case of two public transport journeys using the same public transport vehicle, the intermediate walking trip is removed. Thus, all the inconsistencies in the data are corrected and complete surveys are considered. Series of work trips are collected [2].

Weighting

On average the ODiN/OViN data represents 0.3% of the population, the weighting factors are applied to correct for the over and under-representation of the population. There are individual weights, trip

[2] https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksbeschrijvingen/onderweg-in-nederland–odin—plausibiliteitsrapportage-2019

weights and household weights. As the starting point of the ODiN data also comprises pre-weighting, the initial weights are corrected for the respondents that finish the survey in 7-14 days. The correction for the non-responses among the holidaymakers is applied [2].

Analysis
The focus of the analysis is the transport performance to determine the Dutch mobility patterns and travel behaviour [2].

4.3.2. Data description

The structure of the data comprises trips made by the individual, each row depicting one trip leg made using a particular mode of transport. Each column depicts a particular variable. Hence, multiple records form a trip. Thus, there are 203 variables. They depict the record number of trips, aggregation of the different variables (Trip motive is aggregated for the overall trip as well). Urban regions such as Amsterdam, Rotterdam and The Hague area have more detailed responses. Certain questions are made addressing particular trip makers (questions associated with car ownership). The data comprises umpteen variables that address the individual characteristics, socio-demographic characteristics and travel characteristics (Shelat et al., 2018). It can be further divided into different categories. This is depicted in the table4.1. Based on what is relevant to the research the variables are depicted. The variables are associated with a person, displacement, ride, and weight factor. The factors associated with a person are the socio-demographic variables, household variables, vehicles ownership etc. In case of displacement and ride, data associated with the transport mode, purpose, distance travelled in each trip etc. Weight is addressed by weighting factors.

As the focus of the study is to analyze multi-modal trip to and from public transport, the data must be pre-processed to form trips that comprise public transport and the main mode and the access and egress legs. The current data-set comprises data from 2018 and 2019. However, it must be noted for the future that in 2018 the data is considered from august 2018 as it is expected to be a change in the mode choice and travel behavior due to the operation of the NZL (Noord-Zuidlijn) line.

The study conducted by Brands et al., the impact of the NS line is accessed based on the analysis of the smart-card data. The findings from this research suggested that the there is a shift n the over all usage of Public transport, there is an increase of approximately 4% increase in the total ridership in terms of the journey (Brands et al., 2020). Despite the modifications carried out on the existing bus and tram lines, there is a positive impact on the metro ridership (Brands et al., 2020). The figure 4.1 depicts the change in the ridership of the the modes used before and after the introduction of the NZL line. It is observed that with the increase in the metro journeys there is a decrease in the bus and tram journeys. Furthermore, the share of the combination of bus,tram and metro increases to a larger extent. The expected ridership of the NZL line in the year 2030 is 121000 passengers per day [3]. It

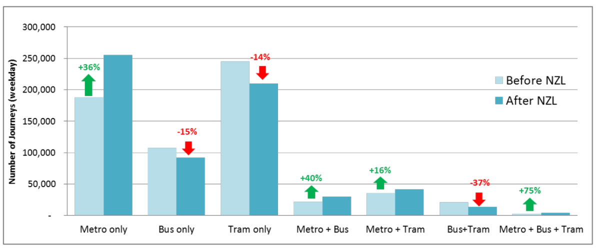


Figure 4.1: Change in ridership by modes(s) (Brands et al., 2020)

aligns with the policy goals of the city of Amsterdam to make Amsterdam more accessible and reduce the car congestion in the urban areas. Thus, it is expected that there is a substantial impact in the mode choice of the users. Hence, to keep the data set more consistent the ODiN 2018 data is considered from the month of August.

[3]https://www.ams-institute.org/urban-challenges/smart-urban-mobility/noordzuidlijn/

Table 4.1: List of attributes based on available data

Category	Probable Attributes (based on data available)
Individual characteristics	Employment, age, gender, license, income, ownership of cars, availability of free student travel, frequency of travel using various modes, main mode of transport displacement
Household characteristics	Household size, number of vehicles per household
Trip characteristics	Distance, travel time, travel purpose, frequency of travel via different modes, origin and destination locations.
Level of service variables	Additional data required
Built environment	Area or region
Working condition	Working hours, car provided by the employer
Weather characteristics	Reporting month
Station choice	Origin and destination station, area, region, train stations used

4.4. Extraction of trips and filtration

The data is extracted in such a way that the individual trips are converted into a multi-modal trip chain. Hence, the data is pre-processed to convert it from individual trips to origin–destination trips. As the data set is available in SPSS the first mode of filtration was SPSS, all the observations with trips comprising more than one trip was filtered out. Furthermore, the trips comprising of public transport as a part of the trip chain was extracted from that. To divide the trips into the different alternatives of the pre-specified mode chains. The transfer nodes are 2, so the number of trip legs are 3. Hence ,to filter out the data the multi-modal trips having 3 trip legs, observations with more than 2 trip legs are extracted. The data-set consists of a variable namely 'Ant-rit' depicting the number of trip legs carried out by the respondent. Thus, all the observations having more than 2 trip legs are filtered in SPSS using that variable.

From the filtered data the observation of each trip leg is to be formulated into a complete trip for further analysis. This process is carried out in MATLAB . The first step comprises dividing the data into socio-demographic data and trip-related data. The focus is on trip-related data to obtain the sequence of the mode-choice in the trip and the origin and destination pin code. The trip related variables comprise trip purpose, trip origin and destination pin-code, number of trips made, a mode used for the trip etc.

The data is filtered are manually to extract the different mode chains using the filter and sort function. Two models will be analyzed using the data set. One model is the train model, which comprises of train as the main mode. The main mode is fixed to train/metro and the access and egress mode are varied. Also, if walking is the part of the access or egress trip for motorized modes or cycles then walking is considered a part of single access or egress leg. Whereas, if walking is the only mode of transport to and from the station then it is considered as a separate alternative. Public transport modes if walking is part of the trip for PT access or egress trip it is considered as one mode. The tables 4.4 and 4.4 depicts the number of observations extracted. To give an idea of the aggregation levels and the extent of data used from the data set is depicted.

Trips (ODiN 2019)	**Number**
Total trips in the Netherlands	150497
Trips within the Amsterdam area	8484
Multi-modal trips	1950
Filtered trips	1213

Trips (ODiN 2018)	**Number**
Total trips in the Netherlands	67296
Trips within the Amsterdam area	4055
Multi-modal trips	921
Filtered trips	439

From the total of 1652 (filtered) trips extracted from ODiN 2018-2019, the observations with missing data are removed, resulting into a total of 1187 observations for the train model and 405 observations for the metro model. The objective of having a similar process implemented for train and metro is to indicate the transfer-ability of the method for the lower-level network such as the metro. It is expected that the metro trips are lower as compared to the train as the main mode as the metro stations are available only within the city of Amsterdam. In the case of the train stations, the network is larger.

Alternatives considered

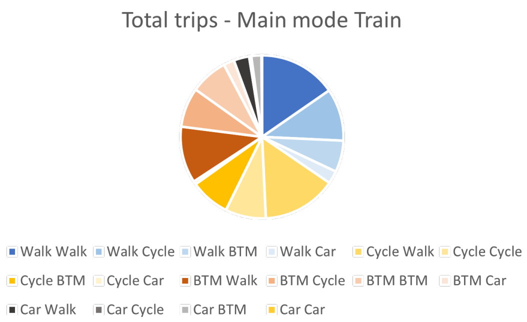


Figure 4.2: Distribution across alternatives (Train)

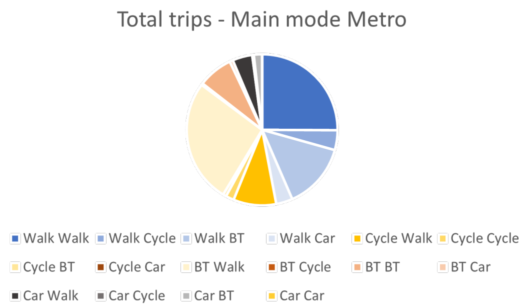


Figure 4.3: Distribution across alternatives (Metro)

As shown in the figure 4.2 the distribution across the alternatives suggests that the alternatives with walk as the access has the highest share. Where as the alternatives having car as the access or egress mode has the least share. The number of observations within the car alternatives is not prominent. In case of the metro as the main mode as depicted in figure 4.3 the alternatives with walk as access and egress has the highest share. In this case the alternatives consisting of cycle and car as the access or egress mode is lower as compared to other modes. Such a distribution is in line with expectations.

According to the study carried out by Krygsman and Dijst slow modes such as walking and cycling are preferred as access modes for about 80% of all the access stages. Moreover, walking is also a connecting mode of transport. Thus, walking forms a continuous and flexible access mode that requires comparatively lesser amount of infrastructure (Fiorenzo-Catalano, 2007). Furthermore, in the case of the egress side, asymmetric availability is expected. Walking is the more dominant mode in that case. Additionally, there are monetary costs associated with public transport services. Thus the distribution of the choices across the alternatives is plausible (Brand et al., 2017).

The distribution of the choice of the data set is also representative of the choices described in the literature as well in the Dutch context. In the study by Shelat et al. it is suggested that 82.6% of the trips are carried out 82.8% trips are carried out using the train and the rest comprise bus tram and metro as the mode of transport. As suggested in the literature the availability of cycle parking facilities is an integral part of the choice of the cycle as the preferred mode and metro stations are usually not equipped with such facilities.

4.4.1. Additional Data

The data only comprises postal pin codes of origin and destination. Thus, the centroid of the postal pin-codes extracted from GIS (Geographical Information System) is used to deduce the origin and destination nodes. In the case of the train model, the origin and destination train station are specified in the data. Hence, the access and egress travel distance and time (for the chosen and non-chosen alternatives) are extracted using the google API algorithm. The algorithm performs calculations for future events. Hence, a weekday is considered and the values are extracted off-peak, so the time fixed for deducing the travel time (Ton et al., 2020). The google API deduces the travel time (and travel distance) for the modes namely transit (public transport), driving, walking, and cycling. The origins and destinations are specified in the form of coordinates or addresses. A similar approach is implemented to generate the choice sets for deducing the how further individuals are willing to cycle to access the station by Van Kampen et al. However, in the case of the metro stations observations the stations are determined. The coordinates of the metro station locations are extracted using GIS. Based on the reported access and egress time in the data set the time the probable station location is deduced. A loop is added to the google API algorithm. The logic for all the origins and destination locations as per the chosen mode the station that has the travel time closest to the reported time is considered.

4.5. Modelling process

As suggested in the chapter 3, the theoretical framework extracted from literature is applied for the case study of the city of Amsterdam. The same framework is applicable on the mode-choices comprising train as a part of the trip chain and metro as well. As described in chapter 2, there are umpteen number of parameters that affect the mode choice as per literature. However, in the interest of time and the available data the a selected factors are analyzed.

To have a more structured approach certain hypothesis are formed for the selected factors and tested individually onto a base model. The base model comprises of the travel-time attributes along with the alternative specific constants. The utility function for the base model with the travel-time parameters is as follows;

$$V_{base-model} = \beta_{TA} * traveltime_{access} + \beta_{MT} * traveltime_{main-mode} + \beta_{TE} * traveltime_{egress} + ASC_{AM-EM}$$

Where, β_{TA}, β_{TM}, β_{TE} are the estimated variables for access time, main mode travel time and egress travel time respectively for the chosen alternatives. ASC_{AM-EM} depicts the alternative specific constant, AM is the access mode and EM is the egress mode.

In chapter 5 the detailed results of the iterations carried out to deduce the model finalized as the base model is explained. The final model is a compilation of the different hypothesis. Due to the limitation of time and data, for a selected attributes hypothesis are formulated. The attributes are added to the relevant alternatives. The coefficients are estimated and checked if they are significant at 90% and 95% interval. The following equation depicts the utility function for estimating the model testing the hypothesis.

$$V_{Hypothesis-test} = V_{base-model} + \beta_{attribute} * Attribute$$

$\beta_{attribute}$ is the coefficient that is estimated for the selected attribute for testing the particular hypothesis. For the attributes that are significant a model with the combined attributes is estimated as follows;

$$V_{Combined-model(MNL)} = V_{base-model} + \beta_{attribute_{s1}} * Attribute_{s1} + \beta_{attribute_{s2}} * Attribute_{s2} + \ldots\ldots$$

It is optimized considering all the significant factors.

The study by Wen et al. the modelling comprises of two steps where in in the initial step MNL model is estimated and then the more advanced model of Latent class model is implemented. Moreover, in the study carried out by Arentze and Molin to capture the error terms are added to correlation between the alternatives when they have same main modes and so are expected to have similar unobserved characteristics. Furthermore, once the model is optimized for the MNL model, the Final optimization is carried out by implementing the error component ML model. The ML model is implemented to as it is a more complex model which can account for the correlation between the different alternatives having the same access and egress mode.

$$V_{Combined-model(ML)} = V_{Combined-model(MNL)} + \epsilon_{AM} + \epsilon_{EM}$$

Where, ϵ_{AM} is the error component for access mode and ϵ_{EM} is the error component for the egress mode for the particular alternative. There are multiple approaches that can be applied using an ML model. In the current study the error-component structure is considered to be the most suitable approach. This method suggests that the error components are added to the alternatives that create the correlations among the different alternatives (Arentze & Molin, 2013; Train, 2002; Wen et al., 2012). In case of standard modelling approach such as MNL the unobserved portion of the utility due to the IIA property the model does not account for correlations amongst alternatives (Arentze & Molin, 2013; Train, 2002). Hence, the ML model shall account for accounting for the taste heterogeneity amongst the alternatives having the same access and egress modes. Furthermore, the ML model is estimated using the Halton draws. It is method that adopts a procedure and takes intelligent draws (Arentze & Molin, 2013; Train, 2002).

Hence, the process comprises estimation of a base model , then a combined model by implementing the MNL model. For a given mode the same mode is repeated for access /egress, thus, to capture the correlation across the different alternatives having the same access and egress modes, the mode across the alternatives is same it is expected that there is a correlation, ML model is estimated for the combined model.

4.5.1. Model performance parameters

To access the performance of the different iterations of the base model certain model performance parameters are considered. To determine the performance of the model the parameters such as the Log-likelihood, Rho-bar square, AIC (Akaike information criterion) and BIC (Bayesian information criterion) is adopted. They help analyze the performance of the overall model. The final value of log-likelihood depicts the log-likelihood of the estimated model [4]. It indicates the goodness of fit of the model. Higher the value better is the fit of the model. It is estimated by $-2(L^i - L^*)$, where L^i is the initial log-likelihood of the model and L^* is the final log-likelihood of the final log-likelihood of the model. The Rho bar square value is estimated as follows;

$$\bar{\rho}^2 = 1 - \frac{(L^i - K)}{L^*}$$

Where K is the number of parameters [4]. Higher the value better the model expected to perform. AIC value indicates the Akaike information criterion (AIC) is an estimator of prediction error and thereby the relative quality of statistical models for a given set of data. The performance of the model is considered better with a lower value. Bayesian information criterion (BIC) is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model. Usually, the lower value of BIC is preferred when selecting the model.

Furthermore, to deduce if the estimated parameter is significant or not certain parameters are considered. In the case of the estimated betas, the impact of the parameter is the t-value and the p-value. The t value depicts the significance of the parameter. The coefficient depicts the relation of the factor with respect to the parameter within the utility function. The p-value depicts at level is the factor significant. Hence, for the estimated beta i.e. β_k, the the t -statistic (t_k), the standard deviation is σ_k, $t_k = \frac{\beta_k}{\sigma_k}$ [4]. There is also a robust t value specified. The difference is that the standard deviation is calculated as the k^{th} diagonal entry or the robust value obtained from the variance covariance matrix i.e. σ_k^R. Hence, in that case the robust value $t_k = \frac{\beta_k}{\sigma_k^R}$ [4]. The p-value specifies the significance of the parameter at a calculated as$2(1 - \Phi(t_k))$, where $\Phi(\cdot)$ is the cumulative density function of the uni-variate standard normal distribution. Whereas for the robust value the calculated as $2(1 - \Phi(t_k^R))$, where Φ is the cu-

[4]https://transp-or.epfl.ch/pythonbiogeme/documentation/pythonfirstmodel/pythonfirstmodel.html

mulative density function of the uni-variate normal distribution [4]. The robust p-value is considered for accessing the significance of the model.

4.6. Selected attributes for Analysis

As suggested in section 4.5, the individual hypothesis is tested related to the attribute that is to be tested is added to the base model. The subsection discusses which attributes are considered to apply the model on the case study of Amsterdam.

4.6.1. Attributes

As depicted in Chapter 2. There are umpteen variables that are significant. However, there are limitations pertaining to data availability and time constraints few relevant factors are selected to be tested. The important factors based the current literature review are deduced in the following paragraphs.

In case of individual characteristics, their impact are context dependent. Car ownership is a significant variable in relation to mode choice (Arentze & Molin, 2013; Krygsman & Dijst, 2001). Moreover, age and gender in many cases are significant (Van Kampen et al., 2020). In the study (Azimi et al., 2020) the of the GFTS (General Transit Feed Specification) and the GIS (Geological Information Systems) data is utilized to make separate models for access and egress. The findings suggest that car ownership, age, income are significant, access and egress distance (Azimi et al., 2020). Similarly, in the study by H. Yang et al. the access and egress mode choice preferences were collected via a stated preference survey for the HSR (High speed Rail) corridor in Shanghai China. Two separate models were analyzed for the access and the egress stage for business and leisure travellers (H. Yang et al., 2019). The findings suggest that car ownership, age, income , access and egress distance are significant. In the case of the business travellers age, income affects the choice of the users (H. Yang et al., 2019).

For the Household characteristics, factors such as the household composition determines the size of the household and the travel group. Households comprising of children can be expected to prefer private modes as it would more convenient for them to make complex trips (Hensher, 2005). Many studies suggest that the mode choice behaviour is highly impacted by trip purpose (H. Yang et al., 2019). In association to that the employment status and the work hours impacts the attitudes towards mode choice. The built environment and the urbanity class in a way depicts the accessibility and attitudes towards modes (Krygsman & Dijst, 2001).

The influence of the trip characteristics is significant as well. The analysis of train commuter behaviour is carried out by latent segmentation method (Chakour & Eluru, 2014). Wherein simultaneously two segments are considered; one is the station first and access mode second and the segment 2, with the access mode first station second. Findings suggest that the travel time is a significant indicator, a better level of service at the station is influential (Chakour & Eluru, 2014). Additionally, the built environment plays an important role in the business and leisure travellers are less sensitive to the egress time as compared to the access time (H. Yang et al., 2019). A stated preference survey is carried out to investigate travel costs and travel time trade-off that the travellers make when choosing the modes of travel available (Arentze & Molin, 2013).

The data regarding the LoS (Level of Service) and Station choice parameters are available, but not within the ODiN dataset which is considered to be the main data source for the modelling process. However, for future implications and understanding station choice, it would add value to have these attributes added using an external data source. Thus, the following attributes are considered;

- Car ownership
- Gender
- Age
- Household size

- Trip purpose
- Employment status
- Urbanity class

The inputs from the discussion with the experts at the Municipality of Amsterdam is also taken into consideration to deduce the hypothesis. Furthermore, as the aim is to provide a framework that allows to deduce the impact of the access and egress mode choice. Hence, the framework allows to test the various hypothesis associated with the different variables that are extracted from literature study.

4.7. Summary

The following figure depicts the process of estimation schematically. As depicted in the figure 4.4, the

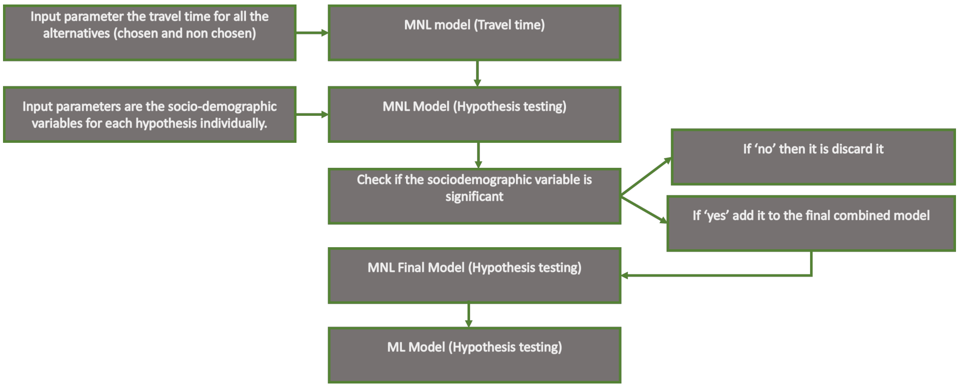


Figure 4.4: Model Estimation process

process for the modelling and analysis of the parameters for the city of Amsterdam is carried out in the following manner. The base model is estimated comprising the travel time parameters (access, main, egress) and the alternative specific constants.

For the next step the hypothesis tested individually for each attribute. If the outcome of the estimated parameter if significant is is added to the combined model. The estimation of the model is carried out using the MNL modelling approach. Once a final combined model is estimated, random error components are added to capture the correlations between the alternatives and a ML model is estimated.

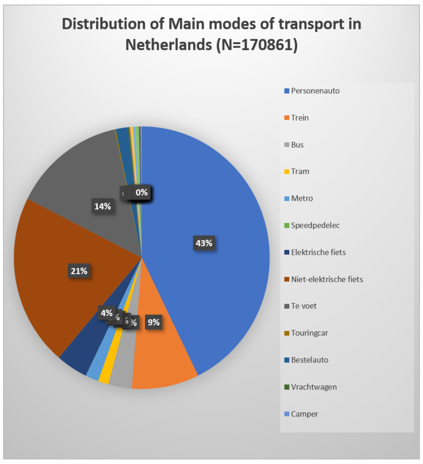
5 Data Description and Analysis

5.1. Introduction

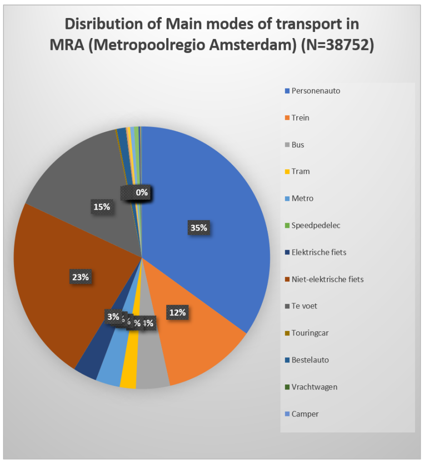
The objective this chapter is to depict data analysis and pre-processing of the data used. Furthermore, the data is pre-processed to generate the choice set as per the decided alternatives. And the insights that can be extracted from the existing data set is explained. The modelling process, analysis and results of the models estimated are discussed in Chapter 6. In this chapter the assumptions and the data analysis for the Case study application is discussed. Furthermore, as depicted in chapter 1 the aim is to apply the modelling framework for not only the regional network but also the urban networks . Thus, there will be two models formulated ; one is the Train model and the other is the Metro model.

5.2. Descriptive analysis

The analysis of the ODiN data is done for the year 2019 to understand and obtain the overall picture of the mode choice behaviour within the data set. The analysis provides the insights into the what the data suggests in terms of the mode choices of the individuals.



(a) Netherlands



(b) Metropolitan Region of Amsterdam (MRA)

Figure 5.1: Distribution of Main modes of transport

The figure 5.1a depicts the distribution of main modes for all the trips in the Netherlands. It show-cases that car has the highest share as the main mode. This followed by cycles. It is expected given the context of Netherlands, due to the availability of well developed cycling infrastructure. The figure 5.1b depicts the main mode distribution across trips for the metropolitan region of Amsterdam. This shows a similar trend to the Netherlands. However the share of cars reduces slightly as compared to the case of the Netherlands.

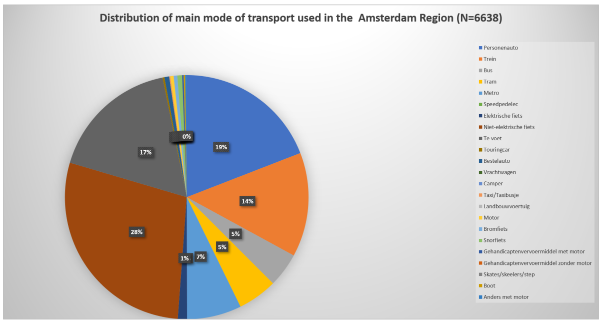
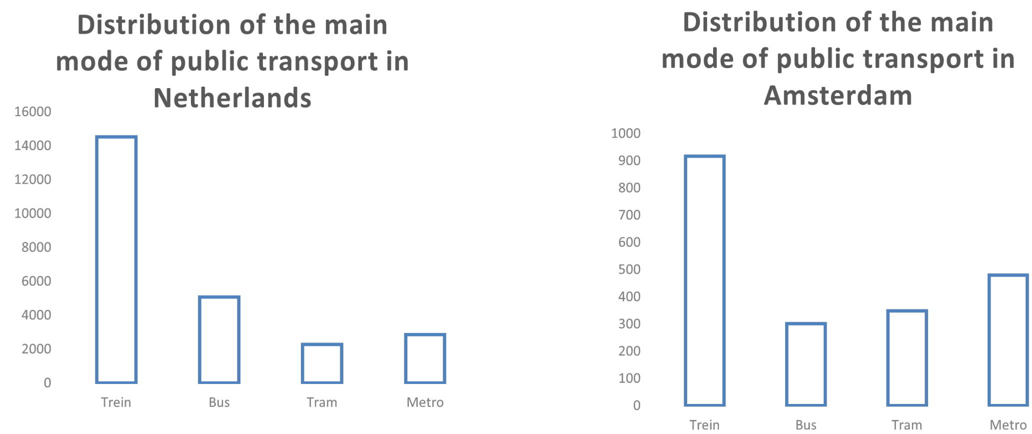


Figure 5.2: Distribution of main mode of transport used in the Amsterdam Region

As shown in figure 5.2, in case of Amsterdam the share of cycles is the highest followed by car and other modes of public transport. Hence, it is evident that in densely populated urban areas with more access to public transport the share of public transport is higher.



(a) Netherlands

(b) Amsterdam

Figure 5.3: Distribution of the Main mode of public transport

From the figure 5.3a, It is observed that amongst the various public transport modes train, followed by bus,metro and tram is the most preferred form of public transport selected to travel. However, as shown in figure 5.3b, in the case of Amsterdam the second preferred option is Metro and not bus. It is expected as the metro services are mainly available in the urban regions i.e. in Amsterdam and the Rotterdam. Hence, the locations at which the metro option exists is very limited.

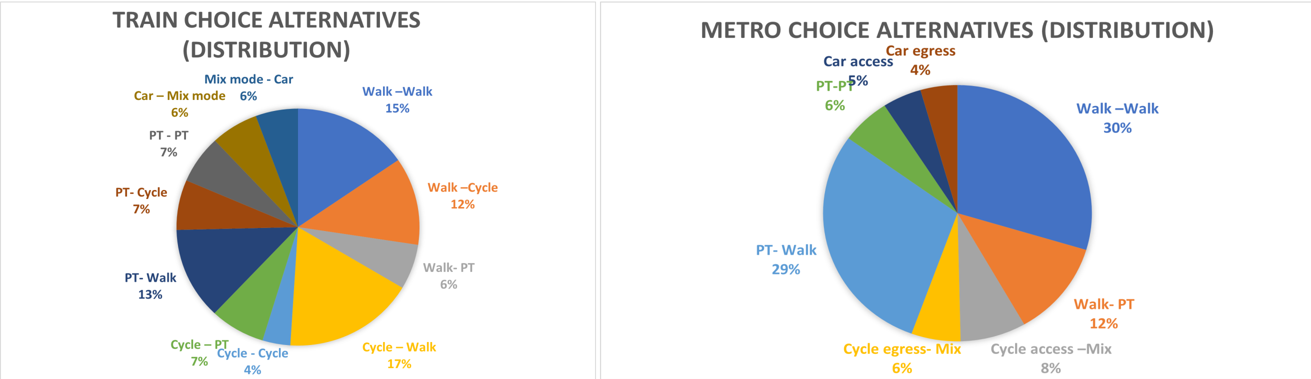
5.3. Data processing and filtration

As depicted in chapter 4, the data is filtered from ODiN data. Ideally it would be possible to have all the mentioned alternatives, but given the data limitations certain alternatives are combined into one alternative.

5.3.1. Alternatives considered

Due to the constraints in the data availability and in come cases the mode- chain formulated is not feasible the alternatives are combined to have a more even distribution of the observations across the alternatives.

For the train model, the car access and egress alternatives are combined into individual alternatives as the data set comprises of lesser number of observations. Hence as depicted in figure 5.4a, to deal with that the alternatives having car access are combined into one alternative and car egress are combined to another alternative. To address the mix-mode access or egress average travel time obtained from the four modes is considered. Similarly for the metro model, as shown in figure 5.4b the



(a) Train (b) Metro

Figure 5.4: Distribution across alternatives

number of observations is even lower and not distributed equally across the alternatives. Hence, the alternatives having cycle as access or egress mode are combined and a similar process is applied to the car alternatives.

5.3.2. Google API vs ODiN travel time values

The values of travel time extracted using the google API algorithm are considered, to maintain the uniformity for the chosen and the non-chosen alternatives. In the original data-set the travel time values of the chosen alternatives are available. Moreover, the exact addresses of the origin and destination of the individual is not available in the data-set but only the postal pin-codes of origin and destination location are available. Hence variations from the original data-set are expected with respect to the estimated values are expected. Furthermore, it the travel time reported by the respondents, there is a chance of misinformation as well. Thus, to counter these limitations the google API algorithm is utilized to calculation the access and egress travel times.

Whereas, for the main mode travel time calculations are extracted from the data set itself as it is expected to stay the same across all the alternatives. To determine the deviation between the reported values and the google API values obtained, the difference between them is calculated. The squares of the difference were calculated and the median of difference between the values is estimated. Thus, as per the calculations the deviation of the train alternative for access and egress travel times is +/-7 minutes. In the case of the metro the probable station locations are based on the reported access and egress travel times. However, it might not always be accurate. The the deviation as compared to reported values is +/- 6 minutes. The choice of the median is made as the variation is expected to be high especially for slow modes such as walking. As the actual address is not used and the centroid data is used there is a chance the the estimated walking time would be longer than the actual travel time experienced by the traveller.

5.3.3. Assumptions

Two transport nodes considered to determine the pre-specified mode chains. This assumption is made to limit the number of alternatives and obtain feasible number of them. Hence, all the possibilities are not considered. This would lead to having an inadequate amount of data for the different alternatives. However, considering the literature the most commonly used number of transfer nodes is two (Fiorenzo-Catalano, 2007). As a consequence more complex trips are not considered.

The travel time estimated by the application of the google API algorithm is off-peak. The travellers experience the travel time during the peak hours and off-peak hours differently (van Eck et al., 2014). The results obtained from this model is not applicable to peak hour traffic conditions. Thus, in order to deduce the behaviour at peak hours, the model can be applied to the data pertaining to peak hours.

The difference between the reported time vs the time extracted using Google API is expected to have an impact on the values. The reported values for the metro station are expected to show deviation as compared to the values estimated using the Google API algorithm. As explained in the previous section, to generate the values for the non chosen alternatives, and the unavailability of data of the selected station can at times lead to the selection of the incorrect station in case of the metro model. Hence, it is expected that there are discrepancies in the values deduced.

In case of the alternatives wherein the mix-modes are used, it is expected that it does not capture the pure form of the alternative due to the limitation of the data used. Hence, ideally it would be preferred that the each mode chain as specified is a separate alternative. However, it is also indicative of the reality as it is line with expectation and literature that the alternatives with lower observation are already less likely to be selected by the passengers. In order to have a comprehensive set of alternatives all are expected to be considered.

5.4. Analysis of the selected attributes

As mentioned in chapter 4, a few relevant factors are selected to test on the Train and Metro model. Hence, the generic formation of the utility is as follows;

$$V_{Hypothesis-test} = V_{base-model} + \beta_{attribute} * Attribute$$

$\beta_{attribute}$ is the coefficient that is estimated for the selected attribute for testing the particular hypothesis. For the attributes that are significant when tested individually, they are added to the to the combined model.

For each attribute selected , there are sub-categories. Based on distribution of the data over the alternatives selected the hypothesis is formulated. It is to be noted that the same hypothesis is applied for the train and metro model. The section 5.4.1 the choices made by the individuals are depicted.

5.4.1. Hypothesis testing: Train model and Metro model

To understand the distribution of the various categorical variables and the distribution across the different alternatives a cross-tabulation is done for all the selected variables across the alternatives. This would be an integral input to test the variables considered for the modelling purposes and analysis of policy implications. In this section a graphical representation shows the distribution across the categories of the selected attributes for train and metro users.

Gender

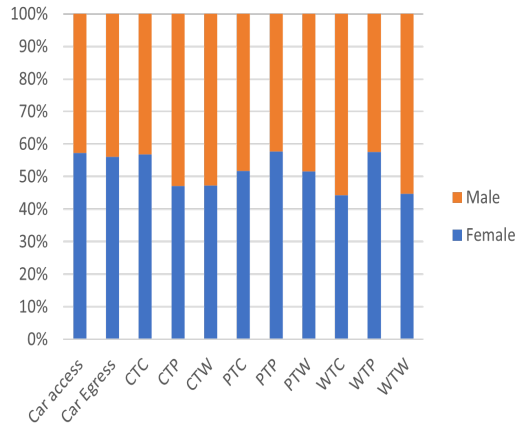
The gender variables have only two categories i.e. male and female. As it is depicted in the figures 5.5a 5.5b, amongst the alternatives, the distribution of both genders is more or less equal across the alternatives. Thus, it can be deduced the gender might not have much of an impact on the access egress mode choice in the Dutch context. This is evident for both train and metro users. The hypothesis formulated suggests that one particular gender favours a particular mode.

Car ownership

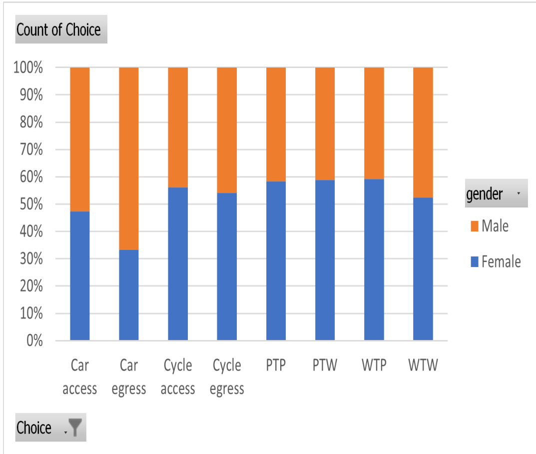
As shown in figure 5.6 an obvious observation is that the alternatives with the car access and egress have a higher percentage of car owners which is almost 70-80% of the respondents. However, it is interesting to note that despite the respondents selecting the other alternatives have a higher percentage of car ownership they still choose to use other access and egress modes. This behaviour can be attributed to the fact the main mode in all the case is public transport. As depicted in figure5.6b metro users the car ownership is relatively higher as compared to the train users (refer figure 5.6a).

Education

The education variable is divided into different categories from no training to highly educated professionals.As depicted in figure 5.7 It is observed that across all the alternatives Highly educated professional is a common underlying observation. Highly educated individuals have the higher share in both cases (ref5.7a,5.7b which is approximately within the range of 50-70%, but is it is more so in case of train

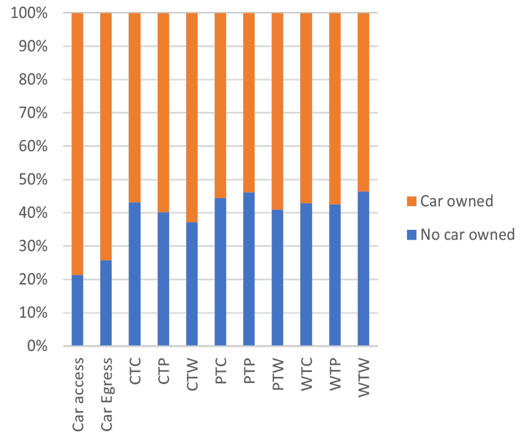


(a) Train

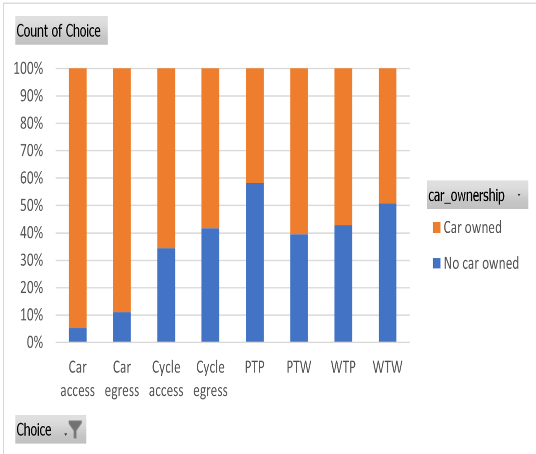


(b) Metro

Figure 5.5: Gender



(a) Train



(b) Metro

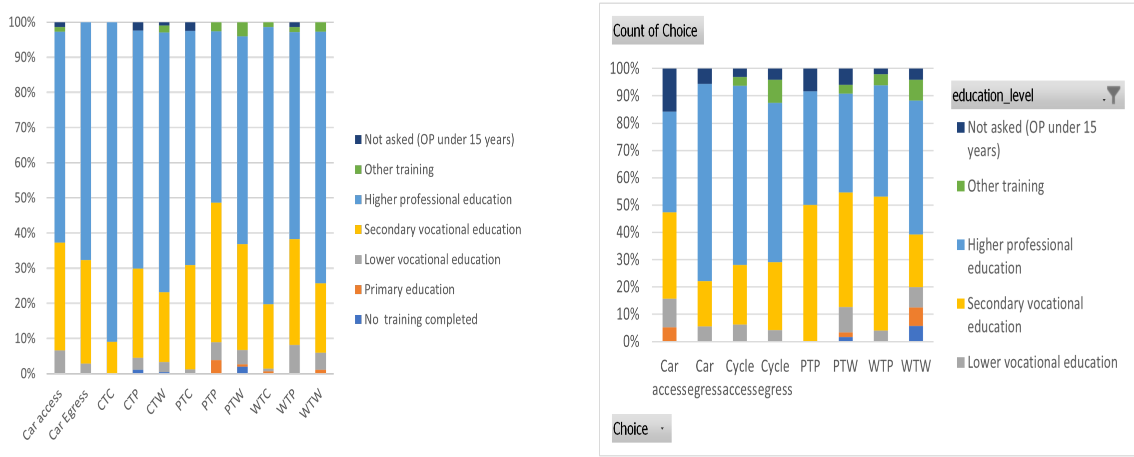
Figure 5.6: Car ownership

users as compared to metro users.

Age
The age categories are divided with a range of 5 years.From the figure 5.8, it can be observed the dominant age groups range from 18-35 years across all the alternatives, so the youth population is well represented. However, for the car access and egress alternatives, the age group is relatively older. The age group 18-29 years is dominant for the Public transport alternatives, this can be attributed to the fact that the Dutch students have access to free public transport. Whereas the young professionals might be more likely to choose active modes such as cycling and walking for access and egress.

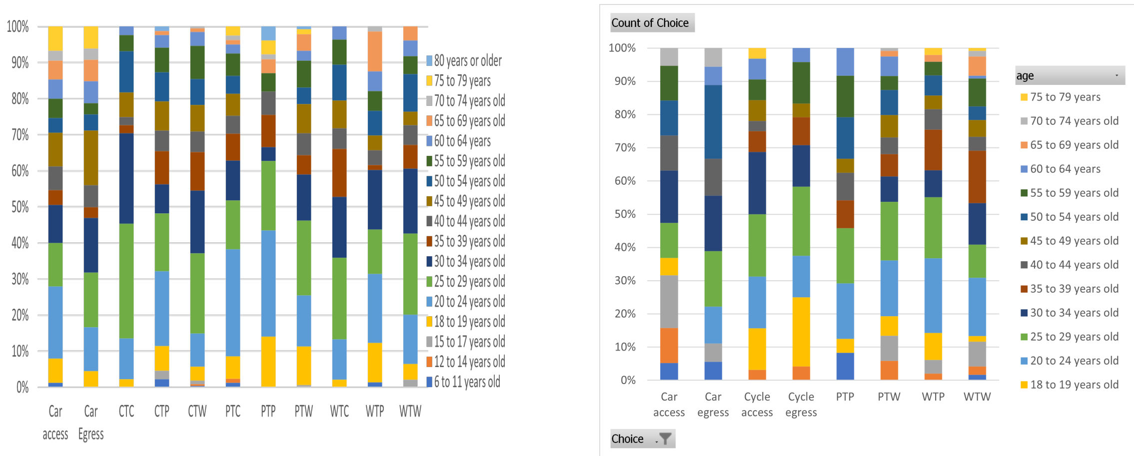
Income
The income parameter depicts the standardized disposable household income. This variable is adjusted for the difference in the household size and composition. It is a measure of prosperity for a household. it is noticeable from the figure5.9 that each category is more or less equally represented.

Trip purpose
Considering the different purposes to travel addressed in the given data-set, the pre-dominant purpose of travel is 'work'(refer figure 5.10) which is in the range of 30-60%. It is observed that the choice of alternatives public transport for access and egress the purpose of education is relatively more prevalent.



(a) Train (b) Metro

Figure 5.7: Education Level



(a) Train (b) Metro

Figure 5.8: Age

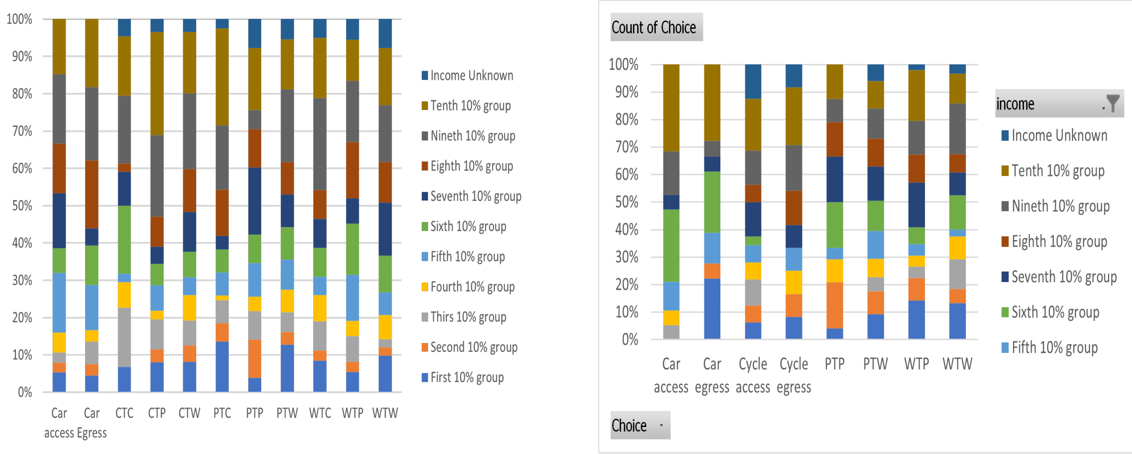
This is in line with expectation as free transport is provided to Dutch students. Whereas, for the people who work the more dominant access and egress modes are active modes (i.e. walking and cycling). The alternatives comprising car access and egress have a higher percentage of visit and recreational purposes as compared to the other alternatives.

Urbanity class

The variable urbanity class represents the residential area of the respondent. As portrayed in figure5.11, the most dominant category is very urban or highly urban, It is an expected outcome as the case study is Amsterdam. Furthermore, for the car access and egress alternatives, it can be observed that the urbanity class is 'less urban'. It is also a plausible outcome as denser areas have not only more public transport access but also the accessibility via active modes is expected to be better.

Household composition

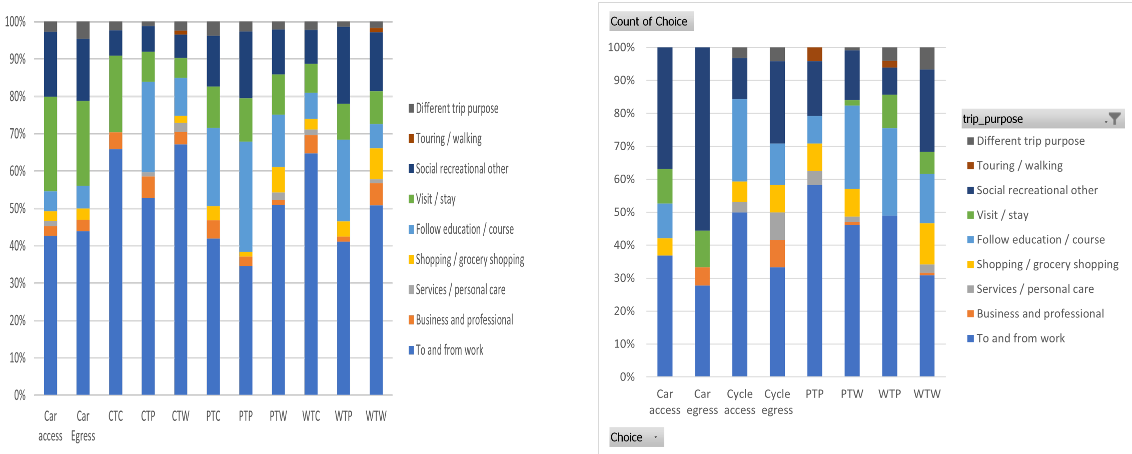
As depicted in figure 5.12, there is variation observed across the alternatives concerning the household composition. Having a child seems to impact the choice of the respondents. Couples and single households without children tend to prefer more combinations of active modes as alternatives. Whereas in the car access alternative the respondents have a dependent i.e. children. Similarly, it seems plausible to travel in a group in public transport as well and so the share of a household composition with more than 2 members is depicted.



(a) Train

(b) Metro

Figure 5.9: Income



(a) Train

(b) Metro

Figure 5.10: Trip purpose

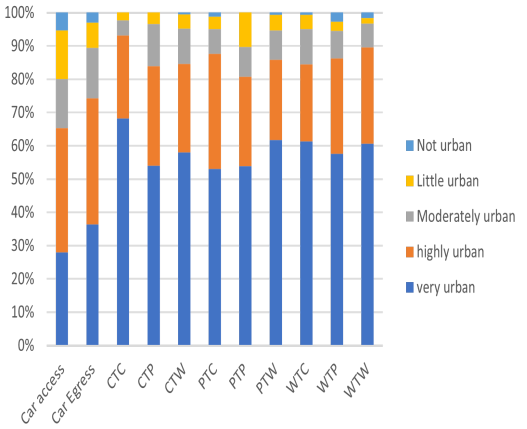
Employment status
As shown in figure 5.13,larger share of the individuals have a full time job which is approximated within the range of 40-80%. It is more so in case of the train as compared to the metro users.

5.4.2. Sub-categories of the variables

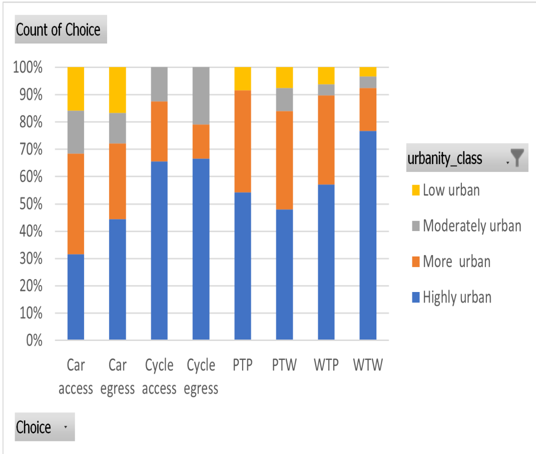
As per the observations made the variables are divided further into categorical variables and added according to the hypothesis tested as per the hypothesis formulated that will be discussed in the next section. The following figure depicts how the data is distributed amongst the different sub categories for alternatives with train and metro as the main mode.

5.4.3. Hypothesis formulation

Considering the insights from the data analysis and the re-categorized data the hypothesis is formulated. The attribute is introduced as a dummy variable except for the level of education. The level of education is added as an ordinal variable.
For the car ownership the hypothesis suggesting that owning a car makes it more likely to choose the alternative that consists car for access and egress.
Hypothesis tested: The respondent is likely to use a car for access or egress if the household owns one.

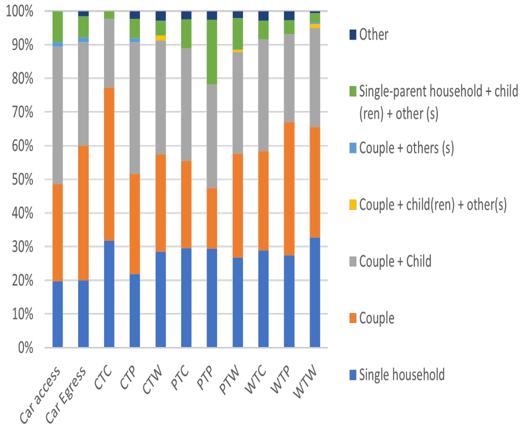


(a) Train

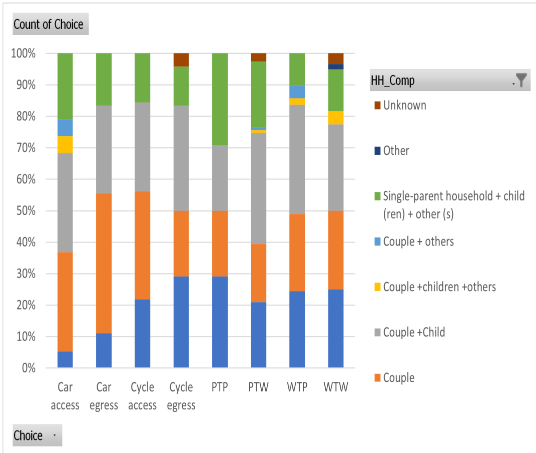


(b) Metro

Figure 5.11: Urbanity class



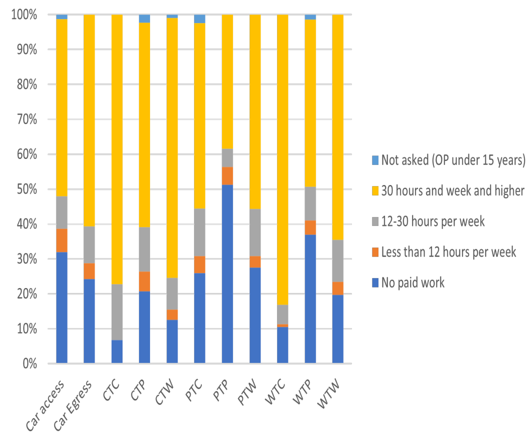
(a) Train



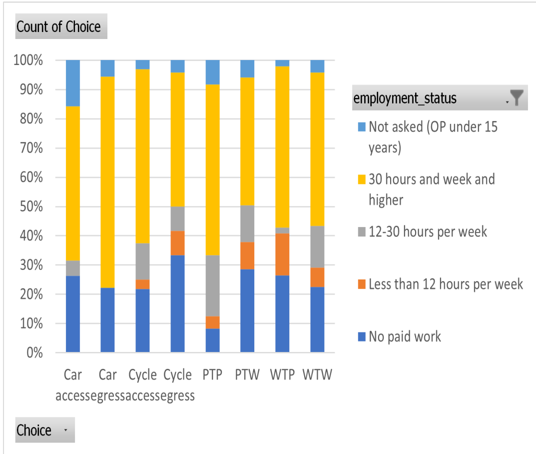
(b) Metro

Figure 5.12: Household composition

For the subcategories pertaining to the age group the highest share is of the 25-34 years (refer table5.1). However, as suggested in multiple studies that dutch students obtain access to free public transport thus incentivising them to use public transport.
Hypothesis tested:The age-group within the range of 18-24 are highly likely to select the alternatives with public transport mode for access and egress.
This assumption is made considering that Dutch students have free public transport when pursuing education. This variable is added to the alternatives comprising public transport for access and egress.
As suggested the table 5.1, the higher share of individuals have the motive to travel for work or business. As depicted in figure 5.10b,5.10a it can be observed that the the major share of the users use active modes such as walking and cycling for access and egress. Furthermore, in the Dutch context professionals travelling for work purposes are expected to select active modes for access and egress (Shelat et al., 2018; Ton et al., 2019). The hypothesis is as follows;]
Hypothesis tested: Respondents having their trip purpose of work are more likely to use active modes for access and egress.
The household composition has family as the highest share amongst the categories. It is expected that individuals with families would prefer to chose modes that allow for a group to travel hence alternatives with car and public transport as access and egress modes.
Hypothesis tested: Couple with others or children are more likely to opt for public transport or car as it is expected that it's easier to travel in groups.



(a) Train



(b) Metro

Figure 5.13: Employment status

Employment status depicts that the highest share of the individuals have a full time job (refer table 5.1). As depicted in figure 5.13a,5.13b it is evident that the individuals with a full time job select active modes (i.e. walking and cycling) for access and egress especially for the individuals selecting trains as the main mode as compared to the metro users. Additionally, the individuals with full time jobs are more likely to use active modes for access and egress as they have a fixed routine to follow (Shelat et al., 2018). The hypothesis is formulated as follows;
Hypothesis tested: Respondents with full time jobs prefer to use active modes (walking and cycling) for access /egress.
As shown in figure 5.7a, the highest share of the individuals are highly educated. Especially, for alternatives comprising of cycling as the access and egress mode followed by walking. Where as, for metro users as shown in figure 5.7b a similar trend is observed for cycling. Hence, the hypothesis posited is as follows; ***Hypothesis tested: highly educated individuals are more likely to use active modes i.e., walking and cycling for access and egress.***
As suggested in the data (refer table 5.1) higher share of the respondents live in highly urban areas. It is expected as the main mode of transport is a public transport mode. Moreover, urban areas are expected to have more access to public transport. Living in dense urban regions encourages the adoption of multi-modal trips as well (Krygsman & Dijst, 2001). Thus, it can be expected that the individuals that lived in highly urban areas are more likely to chose public transport for access and egress. The hypothesis is as follows;
Hypothesis tested: Higher the urban density is more prone to select the alternatives with public transport access.
The gender is more or less equally distributed across the alternatives. So it can be expected that the impact of gender as high as expected. However to test it the following hypothesis is formulated;
Hypothesis tested: Men are more likely to use cars for access and egress
It can be expected that individuals earning a high income can afford more expensive modes of transport such as cars and can be expected that they are more likely to use cars for access and egress. Hence, the hypothesis is follows;
Hypothesis tested:Respondents having a higher income have more chance of using a car for access and egress.
However, it can be observed from the table 5.1 that the dominant classification is the middle-income individuals.

5.5. Summary

Based on the data analysis the key takeaway is as follows;
In the Netherlands, the dominant mode of transport are cars i.e. 43%. However in case of the region of Amsterdam the share of public transport and cycles increases and the cars decreases by 7%. In the city of Amsterdam Train and metro dominate as the main modes of public transport.

In case of car ownership the distribution is almost the same about 60% of the respondents own cars for both the modes. However, it is necessary to note that the car ownership depicts that the household owns cars. In case of the age category the distribution is similar for both modes, both cases the cases the dominant age group is 25-34 years old. In terms of house hold composition the family category has the highest share. For the trip purpose , work related trips are the dominant ones especially for the train as compared to the metro. The employment status echos that the work related trips are mainly made using public transport modes and having a full time job is makes it even more likely make such multi-modal trips. It is observed that the level of education the train users is higher as compared to the respondents using metro. The urban density of the residence is very urban in both cases. Overall, it can be inferred that a similar trend is observed for both modes of public transport.

Variable	Categories considered	Frequency (Train)	Frequency (Metro)	Percentage (Train)	Percentage (Metro)
Car ownership	Car owned	712	238	60	59
	No car owned	475	167	40	41
Age	> 18 years	17	45	1	11
	18-24 years	262	90	22	22
	25-34 years	406	105	34	26
	35-49 years	235	80	20	20
	50-64 years	201	67	17	17
	≥65 years	66	18	6	4
Household composition	Single	330	91	28	22
	Couple	366	99	31	24
	Family	491	215	41	53
Trip purpose	Work/Business	674	171	57	42
	Shop/Services	58	41	5	10
	Education	149	76	13	19
	Recreation/Visit	306	117	26	29
Employment status	No pay	275	121	23	30
	Part-time (upto-12 hours)	43	30	4	7
	Part-time (12-30 hours)	124	45	10	11
	Full-time	745	209	63	52
Education level	No training	13	30	1	7
	Primary	7	11	1	3
	Lower-vocational	43	28	4	7
	Secondary-vocational	294	131	25	32
	Higher-professional	830	205	70	51
Urbanity class	Highly urban	660	241	56	60
	More urban	338	109	28	27
	Moderately urban	116	31	10	6
	Low urban	58	24	5	6
	Not urban	15	0	1	0
Gender	Male	590	183	50	45
	Female	597	222	50	55
Income	High	281	124	24	31
	Medium	434	149	37	37
	Low	472	132	40	33

Table 5.1: Sub categories of selected variables

6

Discrete choice model

6.1. Introduction

The goal of this chapter is to showcase the process of the estimation of the model. The previous chapter provided insights into the the data and the pre-processing done generate the choice set/alternatives. Furthermore, the results pertaining to the hypothesis testing and the final model analysis is depicted in this chapter. The first section showcases the results obtained from the base model estimation and the analysis of them. The section.. compares the results obtained by the hypothesis testing for metro and train. The section... comprises the final model obtained for the model with all the significant parameters.

6.2. Modelling process

As suggested in the chapter 4, the theoretical framework extracted from literature is applied for the case study of the city of Amsterdam. The same framework is applicable on the mode choices comprising train and metro as the main mode. To have a more structured approach certain hypothesis are formulated for the selected factors and tested individually. The final model is a compilation of the different hypothesis. It is optimized considering all the significant factors. Furthermore, once the model is optimized for the MNL model, the Final optimization is carried out by implementing the error component ML model. The ML model is implemented to as it is a more complex model which can account for the correlation between the different alternatives having the same access and egress mode.

6.2.1. Base model estimation process

To deduce the base model an iterative process is carried out to to obtain the travel time parameters. As the research focuses on mode choice, the first iteration comprised of the mode specific access and egress constants; the following equation depicts the structure of the utility function;

$$V_{model-1} = \beta_{access-mode} * travel - time_{access} + \beta_{main-mode} * travel - time_{main-mode}$$
$$+\beta_{egress-mode} * travel - time_{egress} + ASC_{accessmode-egressmode}$$

The second iteration depicts the of having generic parameters for access and egress mode choice. The utility function is as follows;

$$V_{model-2} = \beta_{access} * travel - time_{access} + \beta_{main-mode} * traveltime_{main-mode}$$
$$+\beta_{egress} * travel - time_{egress} + ASC_{accessmode-egressmode}$$

In case of the model 3 that is chosen, an additional parameter is estimated for the alternatives having a mix mode access and egress. In case of the train model car alternatives are grouped into one.Thus, the structure of the equation remains similar to the model 2 except for the for the alternatives having mix mode components β_{access} is replaced by $\beta_{mix-modeaccess}$. Similarly, β_{egress} is replaced by $\beta_{mix-mode-egress}$.

$$V_{model-3} = \beta_{access} * travel - time_{access} + \beta_{main-mode} * traveltime_{main-mode}$$

$$+\beta_{egress} * travel - time_{egress} + ASC_{accessmode-egressmode}$$

To deduce which model fits better a model with mode-specific constants and betas associated with the access and egress travel time per mode is estimated. To access the model performance and select the base model the Log-likelihood ratio test, rho bar square, AIC and BIC are considered (Ton et al., 2020). The value of the former parameters shall be higher and the latter values are preferred to be lower (Ton et al., 2020).

6.2.2. Train Model

Base model estimation (Train)
In order to deduce the mode based on the indicators decide how to create the base model. Note, that for all the cases walk-walk alternative is the reference alternative.
As depicted in table 6.1, the model number 3 fits has the highest value for the LL test and Rho bar square value. The AIC value is the lowest but the BIC value is not the least but close to the least value. As, most parameters indicate **Model 3** as the best performing model, it is considered as the base model. The alternatives associated with the car are clubbed into two specific alternatives wherein the travel time is considered to be the average travel time of all the modes used hence it is separately accounted for as a travel time beta for the model 3. The final LL value of the model is higher than the model with generic travel time betas. Although the AIC value is slightly higher it provides a more behaviorally plausible model. Thus, the final model is a blended version of the model with generic travel time betas across the modes with the exception of the alternatives with the car as the access or egress mode.
Thus, to have separate travel time parameters for cases in which mix mode access and egress are

Model	Model 1 Mode specific parameters	Model 2 Generic parameters	Model 3 Generic parameters with Mix mode
No of parameters	20	13	15
$\mathcal{L}(\beta_0)$	-2846.302	-2846.302	-2846.302
$\mathcal{L}(\hat{\beta})$	-2592.392	-2586.371	-2577.425
$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})]$	507.819	519.861	533.65
$\bar{\rho}^2$	0.082	0.087	0.088
AIC	5224.784	5198.743	5188.954
BIC	5326.368	5264.772	5265.142

Table 6.1: Base Model performance (Train)

considered. Similar approach is applied in the case of the Metro model.
The following table depicts the estimation of the base model
Base model estimation (Train)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC$_{caraccess}$	-1.33	0.337	-3.95	0.00
2	ASC$_{caregress}$	-1.45	0.484	-3.00	0.00
3	ASC$_{cycle-cycle}$	-2.36	0.428	-5.53	0.00
4	ASC$_{cycle-PT}$	-1.57	0.335	-4.69	0.00
5	ASC$_{cycle-walk}$	-0.0878	0.202	-0.44	0.66
6	ASC$_{PT-cycle}$	-1.73	0.384	-4.50	0.00
7	ASC$_{PT-PT}$	-1.66	0.313	-5.32	0.00
8	ASC$_{PT-walk}$	-0.388	0.186	-2.08	0.04
9	ASC$_{walk-cycle}$	-0.726	0.356	-2.04	0.04
10	ASC$_{walk-PT}$	-1.28	0.306	-4.19	0.00
11	β_{access}	-0.0216	0.00738	-2.93	0.00
12	β_{egress}	-0.0385	0.0195	-1.98	0.05
13	$\beta_{main-mode}$	-0.00601	0.00271	-2.22	0.03
14	$\beta_{mix-modeaccess}$	-0.0580	0.0144	-4.02	0.00
15	$\beta_{mix-modeegress}$	-0.0667	0.0255	-2.61	0.01

Summary statistics
Number of observations = 1187
Number of excluded observations = 0
Number of estimated parameters = 15

$$\begin{aligned}
\mathcal{L}(\beta_0) &= -2846.302 \\
\mathcal{L}(\hat{\beta}) &= -2579.477 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 533.650 \\
\rho^2 &= 0.094 \\
\bar{\rho}^2 &= 0.088
\end{aligned}$$

Studies usually consider the confidence interval of 90%. The base model estimation table6.2.2 depicts that the access egress and main mode travel time are not weighted the same. The value of access travel time is relatively lower as compared to egress. It is expected as on the access side the respondent is certain of the mode as it would be available at the access end of the trip. The trip duration of the main mode comprises the in-vehicle time only. The coefficient of travel time of the main mode is lower as compared to the travel time access and travel time beta parameter. It is observed that the coefficient of travel time of the main mode of the trip has the least weight. Thus, it provides insights into how the individual weights the different segments of the trip. In the case of travel time betas associated with the car alternatives the travel time of egress has a lower value. Furthermore, the value of the coefficient associated with the average travel time across the modes is lower as compared to the generic travel time value based on the actual choice made by the individuals. The following parameters are deduced for the different models specified the steps carried out to estimate the base model.

6.2.3. Metro Model

A similar approach to the train model is adopted to estimate and select the base model for the metro model. The process of estimation as explained in section6.2.1. In case of the metro the best performing model is model 1 in terms if all the parameters. However, the betas estimated for each of the mode is insignificant (refer appendix..). Furthermore, to have a transferable framework for all modes of transport it is only viable to follow a similar modelling process. Moreover, as discussed in chapter 4, as the alternatives pertaining to cycling and car are grouped into one alternative. In that case the average travel time is considered across the modes to estimate the mix mode betas for the cycle and car alternatives. For the model with generic access and egress parameters performs the worst as per the values of all the parameters considered. Model 3 which as the generic parameters and mix-mode parameters performs better than model 2. Hence, the model 3 is the base model, the estimated parameters are depicted in table6.2.

BASE MODEL METRO

In the case of the metro model, the base model is deduced using the same framework to the train. The

Model	Model 1 Mode specific parameters	Model 2 Generic parameters	Model 3 Generic parameters with mix mode
No of parameters	20	10	14
$\mathcal{L}(\beta_0)$	-842.174	-842.174	-842.174
$\mathcal{L}(\hat{\beta})$	-553.339	-604.594	-583.262
$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})]$	577.67	475.16	517.925
$\bar{\rho}^2$	0.319	0.270	0.291
AIC	1146.678	1229.188	1194.524
BIC	1126.755	1269.227	1250.578

Table 6.2: Base Model performance (Metro)

table below depicted, the iteration of the base model. Note that that the parameter estimation results are attached in the appendix A.

Base model estimation (Metro)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC$_{caraccess}$	-3.70	0.754	-4.91	0.00
2	ASC$_{caregress}$	-5.16	0.709	-7.28	0.00
3	ASC$_{cycleaccess}$	-3.16	0.507	-6.24	0.00
4	ASC$_{cycleegress}$	-2.45	0.459	-5.35	0.00
5	ASCPT-PT	-3.44	0.390	-8.82	0.00
6	ASC$_{PT-walk}$	-1.35	0.338	-3.98	0.00
7	ASC$_{walk-PT}$	-2.11	0.355	-5.93	0.00
8	$beta_{access}$	-0.0654	0.0153	-4.28	0.00
9	$beta_{car-mix-access}$	-0.0180	0.0111	-1.62	0.10
10	$beta_{car-mix-egress}$	-0.114	0.0626	-1.82	0.07
11	$beta_{cycle-mix-access}$	-0.145	0.0326	-4.45	0.00
12	$beta_{cycle-mix-egress}$	-0.0448	0.0235	-1.91	0.06
13	$beta_{egress}$	-0.0468	0.0176	-2.66	0.01
14	$beta_{main-mode}$	0.133	0.0853	1.56	0.12

Summary statistics

Number of observations = 405

Number of excluded observations = 0

Number of estimated parameters = 14

$$\mathcal{L}(\beta_0) = -842.174$$
$$\mathcal{L}(\hat{\beta}) = -583.262$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 517.824$$
$$\rho^2 = 0.307$$
$$\bar{\rho}^2 = 0.291$$

The table above depicts the parameters to analyze the performance of the model. In the case of the metro, the weights associated with the access mode is higher as compared to the egress modes. Moreover, the travel time beta associated with the main mode is positive and insignificant. It can be attributed to the fact that the metro network is only available in the city of Amsterdam. Furthermore, the trips to, from and within Amsterdam are considered. For the respondents travelling to and from outside Amsterdam, the journey duration is expected to be longer. Furthermore, the metro network is a high frequency dense urban network. Thus, it is plausible that the value of travel time of the main mode is insignificant and positive. The betas for the access travel time have a lower weight as compared to the egress travel time.

6.2.4. Analysis

6.2.5. Train model and Metro model: Hypothesis testing

The hypothesis is tested by adding each variable at a time to deduce if the variable is significant parameter and then added to the final model. As the same hypothesis are tested on the metro and train both are depicted in a comparative format. The coefficient and the robust p-value estimated for the train and metro for each individual hypothesis tested on the base model.

Car ownership

Hypothesis tested: The respondent is likely to use a car for access or egress if the household owns one.

The expected outcome is that the value of car ownership is significant, positive.

Mode	Value of parameter	Significance (Robust p- value)
Train	0.885	0
Metro	1.68	0.01

As shown in the table the car ownership parameter is added to the alternatives comprising of the car used as an access or egress mode. The coefficient is positive, high, and significant. It is an expected outcome and is in line with the literature. Car ownership is depicted to be a significant variable that affects mode choice. It is necessary to consider that this study addresses the access and egress mode choice to and from public transport. For metro, The value of the coefficient for car ownership is significant and positive and high. Hence, it implies that car ownership has a significant relationship to the choice of car as the mode for access and egress. Also, the value is high, it can be attributed to the fact that for the trips carried out by respondents from outside Amsterdam it is necessary to reach the city to access the metro services. Dutch context car ownership is expected to have a significant relationship with the adoption of multi-modal trips as well (Krygsman & Dijst, 2001).

Age

Hypothesis tested:The age-group within the range of 18-24 are highly likely to select the alternatives with public transport mode for access and egress.

This assumption is made considering that Dutch students have free public transport when pursuing education. This variable is added to the alternatives comprising public transport for access and egress.

Mode	Value of parameter	Significance (Robust p- value)
Train	0.818	0
Metro	0.247	0.31

The age group of 18-24 has a significant and positive coefficient when added to the public transport alternatives. It is in line with literature in the Dutch context. Additionally, the value of the travel time for students is lower as compared to working professionals. Hence, the factors such as longer waiting and vehicle time don't hinder there. The age is a significant parameter, when it comes to the choice of public transport as the access or egress mode. However, depending on the cultural context the relevant age group changes. In a study done in China HSR (High-speed rail) is more likely that older people would select the public transport mode choice (H. Yang et al., 2019). However, in a case on studies carried out in Dutch context, it is expected that the respondents that belong to the age group of 18- 24 are more likely to select public transport as the mode for access and egress as they have the advantage of having public transport free as students (Shelat et al., 2018), (Krygsman & Dijst, 2001). However,The parameter of age is not as significant in the case of the metro. It is a positive coefficient; however, it is not a significant parameter even though it is positively correlated. Furthermore, studies depict that age is not always an important factor affecting the adoption of multi-modal transport choices but in this case, it is.

Household composition

Hypothesis tested: Couple with others or children are more likely to opt for public transport or car as it is expected that it's easier to travel in groups.

Mode	Value of parameter	Significance (Robust p- value)
Train	0.245	0.04
Metro	0.0284	0.90

The relation, when added as a test, is a significant and positive coefficient when added to the car and public transport alternative. It is considered that households comprising of families with children are expected to have a preference to be able to choose alternatives allowing them to travel in a group. Furthermore, the studies carried out in the American context suggest that people with larger families and more responsibilities are more likely to select fewer complex trips and prefer cars. However, studies here suggest that the household composition comprising more than 2 members are likely to choose PT and cars. It is an insignificant parameter when combined with other variables in the final model. Having PT subscription associates positively with PT and cycle usage (**Ton2019a**).For the metro model, the value of the coefficient is associated with positive but not significant. However, when a model with all the socio-demographic variables is added the household composition doesn't stay a significant parameter. It can be attributed to the fact that the main mode of transport is the train which is a mode of public transport itself. However, studies suggest that the household with children have a negative correlation with the adoption of multi-modal transport (Krygsman & Dijst, 2001).

Trip purpose

Hypothesis tested: Respondents having their trip purpose of work are more likely to use active modes for access and egress.

Mode	Value of parameter	Significance (Robust p - value)
Train	0.597	0
Metro	0.164	0.6

The hypothesis suggests that for the trip purpose of work the respondents prefer to use active modes i.e., walking and cycling for access and egress to and from the train station Shelat et al., 2018, Ton et al., 2020. The coefficient is positive and significant. Furthermore, it remains significant for the optimized model with all the parameters. Thus, the trip characteristic of the trip purpose is an essential variable that affects the mode choice for access and egress. This observation is echoed in other studies as well. It is necessary to consider that in the context of the Netherlands with a relatively well-developed cycling infrastructure the choice of the cycle for access and egress is expected. Furthermore, in the final model, it is a significant variable impacting the mode choice for access and egress.
For the metro model The value of the coefficient is associated with positive but not as significant. It can be attributed to the facilities available at the train station as compared to the metro station are more; eg: cycle parking facilities etc.

Employment status

Hypothesis tested: Respondents with full time jobs prefer to use active modes (walking and cycling) for access /egress.

Mode	Value of parameter	Significance (Robust p - value)
Train	0.707	0
Metro	-0.322	0.13

It is a variable added to the alternative comprising active modes i.e., walking and cycling as access and egress modes. The variable is dummy coded. It can be expected to be correlated to the trip purpose to a certain extent. However, the coefficient of full-time workers is positive and significant not only in the initial model but also in the final model with all parameters. Having a full-time job suggests that a fixed timetable is expected to be followed and hence the value of time is higher as well. This is in line with the studies carried out in the Dutch context. Full-time workers are highly likely to adopt the cycle-train combination (Shelat et al., 2018). For the metro model is negative but not as significant. There are studies suggesting that a part-time job has a more positive correlation with cycling as a mode than a full-time job (Ton et al., 2019). However, this also depicts that base on the mode of public transport adopted the preference changes. It can be argued that the availability of cycle parking facilities at the train stations makes it more viable for the commuters to use cycles to access and egress the train station as compared to the metro station.

Education level

Hypothesis tested: highly educated individuals are more likely to use active modes i.e., walking and cycling for access and egress.

Mode	Value of parameter	Significance (Robust p-value)
Train	0.246	0.01
Metro	0.178	0.14

This parameter is modelled in an ordinal suggesting it would be expected with the increase in the level of education the chance of selecting an active mode of transport is higher. It is in line with literature that suggests that highly educated professionals are more likely to select an active mode of transport. In the case of the train model, the coefficient obtained for this variable is significant and positive. However, in the case of the final model, it is not as significant. Literature suggests that highly educated professional with full-time jobs is more likely to select active modes for transport (Shelat et al., 2018). It is expected that the higher the level of education more likely are the users to adopt active modes for access and egress purposes. In case of the metro model the level of educations has a positive but insignificant relation.

Gender

Hypothesis tested: Men are more likely to use cars for access and egress.

Mode	Value of parameter	Significance (Robust p-value)
Train	-0.173	0.34
Metro	0.711	0

Gender is dummy coded in the model. The hypothesis suggests that males are expected to choose alternatives with the car with access and egress modes. The results suggest that gender is not as significant and is negatively correlated. However, it is in line with what literature depicts. Especially in the Dutch context gender is not a significant variable that affects mode choice. Furthermore, it also suggests that males are more likely to select active modes. Additionally, as it is observed in the descriptive analysis, men and women are equally represented. Furthermore, men are less likely to use car alternatives for travelling to the station (Chakour & Eluru, 2014).

Income

Hypothesis tested:Respondents having a higher income have more chance of using a car for access and egress.

Mode	Value of parameter	Significance (Robust p-value)
Train	-0.232	0.22
Metro	0.158	0.73

The variable is dummy coded in the model. Income is an insignificant factor and positively correlated with the adoption of cars as an access/egress mode to and from the station. It is in line with the studies in the Dutch context suggesting that high-income working professionals chose more sustainable modes of transport. Studies suggest that for cases even when cars might be available a more conscious choice is made by the individuals (Ton et al., 2020). In case of the metro model, the income is having a negative and insignificant coefficient suggesting it does not play a crucial role in the choice of car as the preferred mode for access and egress. There is a possibility that due to the limited availability of metro services outside Amsterdam car seems to be the more comfortable option.

Urban density

Hypothesis tested: Higher the urban density is more prone to select the alternatives with public transport access.

Mode	Value of parameter	Significance (Robust p-value)
Train	0.156	0.20
Metro	-0.104	0.66

It is expected public transport accessibility is higher in dense urban areas. The urban density variable depicts the level of urban density of the residential area of the respondent. Thus, the hypothesis suggesting that the variable of high urban density is added to alternative select public transport as the access mode. The value of the coefficient is not as significant but is positive. It is expected as the case study of Amsterdam is considered in this case thus it is expected that the residential area of the respondent is relatively more urban. For the metro model, the coefficient of urban density is negatively correlated with the use of public transport. It is necessary to note that the public transport access is considered is bus and tram, the train is not included in this model. Hence, there is a possibility as bus networks are available in suburban areas as well, hence it is highly possible that the area is not urban. However, studies support that in urban areas the use of public transport is higher. In addition to that, studies suggest that in the Dutch context the Urban density has a significant impact on the choice of multi-modal trips (Krygsman & Dijst, 2001), (Shelat et al., 2018).

6.2.6. Combined MNL model estimation

The final MNL model is estimated in the case of train and metro by introducing all the significant parameters as tested individually and then optimized. The travel time parameters, even when not significant at p- value 0.05 or 0.01, are still kept as a part of the model to understand the influence of the time on the various access and egress modes.

Combined model estimation (Train)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	β_{age}	0.841	0.146	5.76	0.00
2	ASC$_{caraccess}$	-1.39	0.396	-3.50	0.00
3	ASC$_{caregress}$	-1.46	0.536	-2.72	0.01
4	ASC$_{cycle-cycle}$	-2.33	0.428	-5.44	0.00
5	ASC$_{cycle-PT}$	-1.72	0.338	-5.11	0.00
6	ASC$_{cycle-walk}$	-0.0626	0.202	-0.31	0.76
7	ASC$_{PT-cycle}$	-1.89	0.388	-4.87	0.00
8	ASC$_{PT-PT}$	-1.36	0.336	-4.05	0.00
9	ASC$_{PT-walk}$	-0.544	0.190	-2.87	0.00
10	ASC$_{walk-cycle}$	-0.710	0.358	-1.98	0.05
11	ASC$_{walk-PT}$	-1.45	0.310	-4.66	0.00
12	$\beta_{car-ownership}$	0.912	0.208	4.39	0.00
13	β_{access}	-0.0211	0.00731	-2.88	0.00
14	β_{egress}	-0.0384	0.0197	-1.95	0.05
15	$\beta_{main-mode}$	-0.00620	0.00274	-2.27	0.02
16	$\beta_{mix-modeaccess}$	-0.0585	0.0139	-4.19	0.00
17	$\beta_{mix-modeegress}$	-0.0652	0.0251	-2.60	0.01
18	$\beta_{employment}$	0.575	0.187	3.08	0.00
19	$\beta_{trip-purpose}$	0.379	0.186	2.04	0.04

Summary statistics

Number of observations = 1187

Number of excluded observations = 0

Number of estimated parameters = 19

$$
\begin{aligned}
\mathcal{L}(\beta_0) &= -2846.302 \\
\mathcal{L}(\hat{\beta}) &= -2539.503 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 613.598 \\
\rho^2 &= 0.108 \\
\bar{\rho}^2 &= 0.101
\end{aligned}
$$

Alternative specific constant The alternative specific constants depict the preference of the population. The reference alternative is the walk-train-walk. The alternative specific constant is insignificant and has the highest value suggesting that the preference to use the cycle-train-walk is the same as walk-train-walk. The public transport- train- cycle has the highest negative value suggesting that is the least preferred alternative. It is expected that the passengers prefer to utilize other modes for the egress side of the trip over cycle due to asymmetry caused due to the availability of vehicles on the egress side of the trip. Furthermore, in the case of public transport, the level of service indicators such as punctuality, frequency of the service are influential factors as well.

As compared to the individual model wherein more parameters are significant; the socio-demographic variables significant in the final model are car ownership, age, trip purpose of work and full time employment. The value of the travel-time bets i.e. β_{access},β_{egress},$\beta_{main-mode}$ depict that all the trip legs are not not valued the same. Similar to the values in the base model the access and egress are weighed higher than the main mode travel time. The parameter for egress us weighed slightly higher than access. All the parameters are negatively correlated depicting that travel time causes dis-utility to the choice. Car ownership has highest positive weight indicating the if the individual's household owns a car there is a higher possibility to utilize it for access and egress. Age is also a significant factor in the combined model. It is line with expectations that for the youth i.e. the age-group 18-24 that consists of students are very likely to adopt public transport modes for access and egress. Hence, it depicts that the incentivization of free public transport for collage going students encourages that particular segment of the society to adopt public transport usage for access and egress. In terms of trip purpose of work and fully employed individuals are more likely to utilise active modes for access and egress

to and from the train station. It is to be notes that the weight of the employment status i.e. full time employment is higher as compared to the purpose of travel indicating that full time employment is more influential adoption of active modes for access and egress.
The utility function for the combined model (Train) is as follows;

Alternative(walk-walk) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{reference}$

Alternative(walk-cycle) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{walk-cycle}$

Alternative(walk-PT) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{age} * Age_{18-24years}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{walk-PT}$

Alternative(cycle-walk) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{cycle-walk}$

Alternative(cycle-cycle) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{cycle-cycle}$

Alternative(cycle-PT) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $ASC_{cycle-PT}$

Alternative(PT-walk) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $\beta_{age} * Age_{18-24years}$ + $ASC_{PT-walk}$

Alternative(PT-cycle) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{trip-purpose} * Trip - purpose_{work}$ + $\beta_{employment-status} * Employment - status_{Fulltime-job}$ + $\beta_{age} * Age_{18-24years}$ + $ASC_{PT-cycle}$

Alternative(PT-PT) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{age} * Age_{18-24years}$ + ASC_{PT-PT}

Alternative(car access- mix egress) = $\beta_{access} * travel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{mix-mode-egress} * averagetravel - time_{egress}$ + $\beta_{car-ownership} * Age_{car-ownership}$ + $ASC_{car-access}$

Alternative(Mix access- car egress) = $\beta_{mix-mode-access} * averagetravel - time_{access}$ + $\beta_{main-mode} * travel - time_{main-mode}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{car-ownership} * Age_{car-ownership}$ + $ASC_{caregress}$

Combined Model estimation (Metro)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	$\text{ASC}_{car-access}$	-5.94	0.790	-7.52	0.00
2	$\text{ASC}_{car-egress}$	-5.37	0.784	-6.84	0.00
3	$\text{ASC}_{cycle-access}$	-2.58	0.381	-6.76	0.00
4	$\text{ASC}_{cycle-egress}$	-1.80	0.363	-4.98	0.00
5	ASC_{PT-PT}	-2.82	0.258	-10.93	0.00
6	$\text{ASC}_{PT-walk}$	-0.895	0.167	-5.34	0.00
7	$\text{ASC}_{walk-PT}$	-1.42	0.179	-7.91	0.00
8	$\beta_{car-ownership}$	1.66	0.637	2.61	0.01
9	β_{access}	-0.0680	0.0167	-4.07	0.00
10	$\beta_{car-mix-access}$	-0.0226	0.0113	-2.00	0.05
11	$\beta_{cycle-mix-access}$	-0.144	0.0320	-4.51	0.00
12	$\beta_{cycle-mix-egress}$	-0.0293	0.0172	-1.70	0.09
13	β_{egress}	-0.0293	0.00840	-3.49	0.00

Summary statistics

Number of observations = 405

Number of excluded observations = 0

Number of estimated parameters = 13

$$\begin{aligned}
\mathcal{L}(\beta_0) &= -842.174 \\
\mathcal{L}(\hat{\beta}) &= -595.233 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 493.882 \\
\rho^2 &= 0.293 \\
\bar{\rho}^2 &= 0.278
\end{aligned}$$

Alternative specific constants The reference alternative is walk-metro-walk. The ASC of the alternative with car access and mix mode has the most negative value depicting that it is the least preferred alternative as per the choices made by the population. The less negative value is of the public transport-metro–walk alternative. Hence it is the second most preferred alternative selected by the population. The travel time parameter for the car access with mix mode egress and the main mode are not significant at 90% interval and hence the optimized version of the model does not comprise these parameters. However as the objective is to model access and egress mode choice the travel time parameter for car egress can be replaced with the generic parameter value estimated i.e. β_{egress}. In case of the metro model the car ownership is the only significant socio-demographic variable, with a very high value. The access, egress and the main mode trip legs equally. The beta estimated for the main mode is not significant (p-value greater than 0.10) and shows a positive correlation with time. Consequently on the further optimization of the model the beta for main mode is removed from the model. The egress parameter weighs lower than the access mode almost by a factor of 2. The reason behind obtaining such a value can be attributed to the fact that as the metro is only available in the city, for the individuals visiting from outside the city need to travel longer to reach the metro station location. The only significant socio-demographic variable is the car ownership. It is indicative that car is convenient to access the stations especially for the trips outside the city of Amsterdam. The utility function for the combined model (Train) is as follows;

Alternative(walk-walk) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + $ASC_{reference}$

Alternative(walk-PT) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + $ASC_{walk-PT}$

Alternative(cycle access-mix egress) = $\beta_{access} * travel - time_{access}$ + $\beta_{cycle-mix-mode-egress} * averagetravel - time_{egress}$ + $ASC_{cycleaccess-mixegress}$

Alternative(mix access-cycle egress) = $\beta_{cycle-mix-mode-access} * averagetravel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + $ASC_{mixaccess-cycleegress}$

Alternative(PT-walk) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + $ASC_{PT-walk}$

Alternative(PT-PT) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + ASC_{PT-PT}

Alternative(Car access-mix mode egress) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * averagetravel - time_{egress}$ + $\beta_{car-ownership} * Age_{car-ownership}$ + $ASC_{Caraccess-carmixmodeegress}$

Alternative(mix mode access-car egress) = $\beta_{access} * travel - time_{access}$ + $\beta_{egress} * travel - time_{egress}$ + $\beta_{car-ownership} * Age_{car-ownership}$ + $ASC_{carmixmodeaccess-caregress}$

6.2.7. Mixed logit model

The random error component is added to the utility suggested in the base model. There are multiple nesting structures accessed to deduce the correlation between the different overlapping modes.The mixed logit model with the random error components is tested on the model. The random error components are added for alternatives having the same mode for access and egress. This is done to capture the correlation across the different alternatives having the same modes for access and egress. The values of the random error components is insignificant and hence the MNL model insights are applicable. It also suggest that the multi model trip is perceived as a synthetic mode itself rather than having the individual access, main and egress mode components. The error-component structure is as follows;

$$U_{ML-model} = V_{MNL-Model} + \epsilon_{access-mode}$$

$$U_{ML-model} = V_{MNL-Model} + \epsilon_{egress-mode}$$

As shown in the equation above one only comprising access and egress only are nested.

$$U_{ML-model} = V_{MNL-Model} + \epsilon_{access-mode} + \epsilon_{egress-mode}$$

Where as in the equation depicted above both access and egress modes are accounted for. On the model estimation it is deduced that for both train and metro the error-component values are not significant. Hence, it implies that the different alternatives are treated independently despite having modes that are overlapping amongst the alternatives. It is not necessarily in line with literature. In the study conducted by Arentze and Molin, the alternatives with the same main modes exhibit a correlation. Furthermore , multiple iterations suggest that the none of the error components are significant and hence the MNL model is considered to be the used for analysis as the final model results.

Note that that the parameter estimation results are attached in the appendix D.

6.3. Discussion

Model performance

The conventional four-step model is often used for modelling various travel choice dimensions. The Dutch National model, the Swedish model are a few examples of the application of the four-step model using discrete choice models. The decision rule applied is utility maximization. Frequently used models used in practice comprise MNL and NL models. Due to the assumptions and the simplicity of such models they often hinder the realistic depiction of choice dimensions. Furthermore, there is increasing evidence that the choices are heterogeneous. Hence new and more complex models can be applied. The base model only comprises the travel time parameters. For the train and metro, it is observed that adding parameters leads to a better model in terms of the goodness of fit. The AIC and BIC criterion also suggest that the value minimizes for the final model. In the case of the train model comparing the base model with the additional parameters is better as compared to the base model. It is expected as the goodness of fit improves with the increase in the number of parameters. In this approach, a simultaneous approach is adopted to mode choice. It is to be noted that there are not enough observations across all the alternative lading to the alternatives that combine the different alternatives.

TRAIN

The travel time parameters in case of the train suggests that the value of the beta estimated for the main mode travel time in the train is -0.00602, where as in the study conducted by Arentze and Molin the beta estimated for train as the main mode in a multi-modal trip is -0.060 (Arentze & Molin, 2013).

The access and egress are weighed higher than the main mode travel time (Arentze & Molin, 2013). The access time is weighted as -0.0211, -0.0585, where as it ranges from -0.073 to -0.110 as per the study by Arentze and Molin. For the coefficients of egress modes the values are weighted slightly higher than access time. The estimated values of the egress results are -0.0384, -0.0652. The values from the study suggest the a range from -0.069 to -0.130 Arentze and Molin. The values are slightly higher than the ones estimated in this study, it can be due to the nature of data collection of stated preference hence at times the values are over-emphasized due to the bias. Age has a significant and positive correlations as it can be confirmed by literature as well. The age group of 18-24 comprises of university going students and hence due to the PT subscription are very likely to use public transport (Krygsman & Dijst, 2001; Ton et al., 2019; Van Kampen et al., 2020). Car ownership is a significant and positive variable the value is 0.912 and is significant. Trip purpose of work has a significant and positive correlation with work purpose with a coefficient of 0.379, the mode choice determinants estimated in the study conducted by Ton et al. it is also depicted that the trip purpose of work has a positive coefficient with the mode choice of public transport, the value of the coefficient is 6.51, 2.29 and 2.09. Thus, the working professionals are expected to adopt public transport and active modes for travelling for work. The coefficient for full time employment is 0.575 and significant. Full-time employed individuals are more likely to adopt multi-modal trips as well (Krygsman & Dijst, 2001).

METRO

the value of the travel time parameters in case of metro as the main mode is positive and insignificant. In the study by Arentze and Molin the parameter for public transport modes is 0.016 for short trips. However, it is insignificant so it is removed from the final model. The access and egress travel time coefficients are not weighed higher than the main modes but are significant. The travel time for access is weighed higher than the egress travel time. It is contrary to the expected norm wherein it is expected that egress is weighed higher due to uncertainty of availability of modes. However, it can be explained by the scale and availability of the metro. For the individuals commuting from outside the city the journey is longer. The travel within the city is more compact and the users expect to use other modes for the shorter distance. Consequently, the individuals are more sensitive to the access travel time. The only socio-demographic variable that impacts the mode choice is the car-availability. Hence, it is in line with expectations. The other attributes tests might not be as significant as in the case the trips to , from and within the city are considered. Within the city with a well developed cycle infrastructure it might imply that it is a competing mode rather than a complementing mode. Additionally, for individuals who own cars, its is a more convenient option have a less complex trip.

6.3.1. Comparison of Train and Metro model

Comparing the base model

The figure 6.1 showcases the comparison of the base model's with travel time estimates for the train

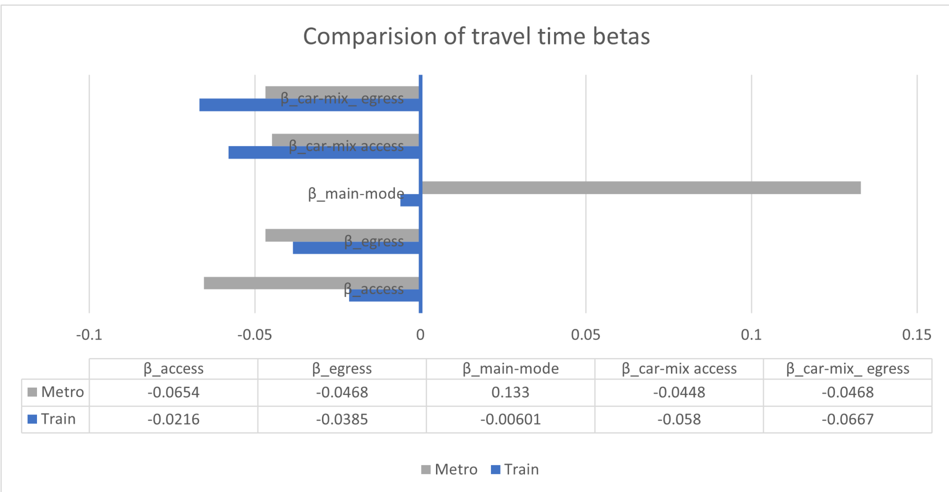
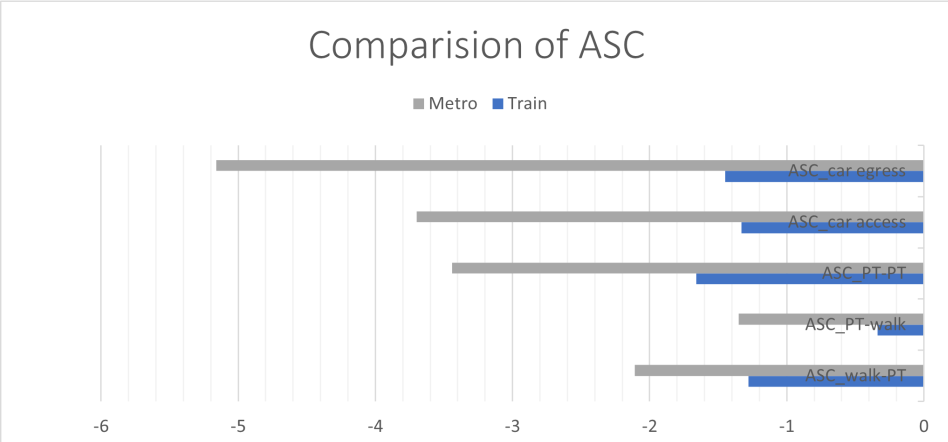


Figure 6.1: Comparison of the betas estimated in the base model

and metro. It can be observed that the individuals are more sensitive to the to the access and egress travel time to and from metro as compared to the train. The value of the travel time beta of the metro access is approximately 3 times that of the train and that of egress is 1.2 times that of the train. Contrary to the weights for access and egress the coefficient for main-mode travel time is not as significant but positive as compared to the train. In case of the train the value is negative and significant. It is expected that for the trips originating from outside Amsterdam the the in-vehicle time of the main mode travel is longer in case of the train as compared to the metro. It is likely that the time to access and egress is higher than the in-vehicle time for the metro as it is an urban rail system restricted to a smaller region.

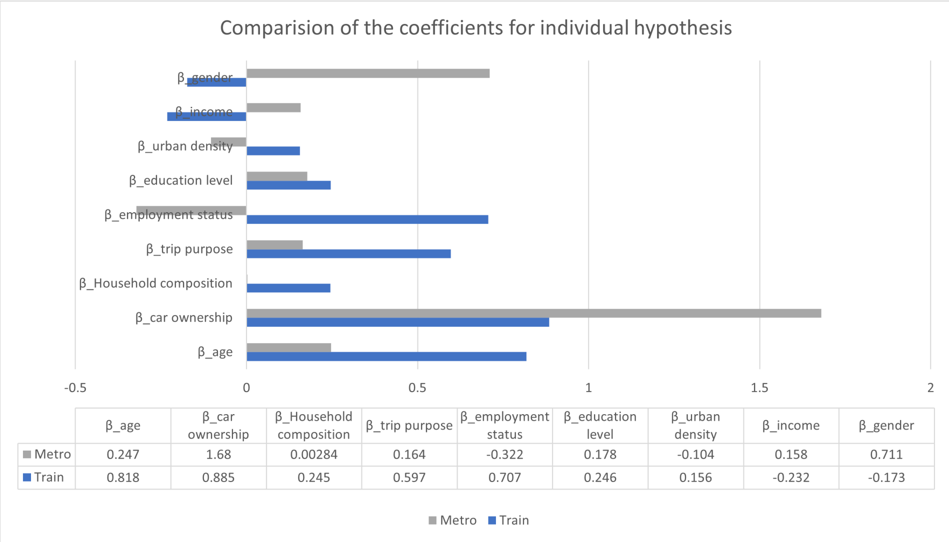


	ASC_walk-PT	ASC_PT-walk	ASC_PT-PT	ASC_car access	ASC_car egress
Metro	-2.11	-1.35	-3.44	-3.7	-5.16
Train	-1.28	-0.338	-1.66	-1.33	-1.45

Figure 6.2: Comparison of the ASC estimated in the base model

As shown in figure 6.2, in both cases all the ASC values are negative depicting that walking as access and egress is the most preferred overall. It is plausible as walking does not require any additional vehicle and is a free in terms of cost. In case of the metro the overall ASC are weighted more negatively as compared to the train. The value of the ASC suggests when the train is the main mode the Public transport access and egress is least preferred and for the alternative with mix mode and car egress is the lest preferred in case of the metro. It is plausible in case of the train it can be expected that facilities such as parking facilities for cars might not be easily available and might also have an associated cost. For the train having public transport access and egress might not be possible due to issues such as to synchronization of the different public transport time-tables and the availability of public transport in the vicinity. Moreover, with the cycling infrastructure available in the Netherlands, the cycle-train combination provides flexibility to the individual. Note that only the attributes estimated in both models are compared.

The figure 6.3 shows the betas estimated in the individual hypothesis tested with the base model. It can be observed that the in case of the betas estimated for the income, gender, employment status and urban density the value for metro and train have an opposite sign. It depicts that males would prefer to use car to access and egress metro station but not the train. For the income parameter it can be interpreted that high income earners do not prefer to use cars to access and egress train stations where as it is vice versa for metro. The beta estimated for employment status suggests that full time employed individuals favour the use of active modes such as cycling and walking to access and egress to and from the train station. However, this is not the case with metro users. The attribute is not as significant in case of the metro model. Thus, it depicts that for both the train and metro the mode choice behaviour cannot be expected to be the same. It is attributed to various other factors such as the type of network, station facilities, attitudes towards to them etc. It is to be noted that for both the models the gender, income and urban density are not significant attributes.



	β_age	β_car ownership	β_Household composition	β_trip purpose	β_employment status	β_education level	β_urban density	β_income	β_gender
Metro	0.247	1.68	0.00284	0.164	-0.322	0.178	-0.104	0.158	0.711
Train	0.818	0.885	0.245	0.597	0.707	0.246	0.156	-0.232	-0.173

Figure 6.3: Comparison of bets estimated for Hypothesis testing

Comparing the final model

Tt is observed that amongst the socio-demographic variables tested, the only significant variable is car ownership for the metro model. It is expected that the metro system is a relatively small network and is available in the city of Amsterdam and hence all respondents do not have direct access to it. Whereas, in the case of the train network, it is spread across the country and within the city. This trait is also implied via the coefficient of the travel time of the main mode in the case of the train is significant and is negative but in the case of the metro, it is positive and is insignificant. As it is a dense, high-frequency urban network it is a fast system catering to a smaller and limited region the in-vehicle time is expected to be short.

Certain aspects such as the highly educated, high-income individuals with a full-time job are more likely to adopt cycling and walking as their preferred mode of transport especially for the bike-train combination (Shelat et al., 2018). This is echoed in the train model and the metro model. The common trait observed is that the car egress travel time beta is not considered to be significant. It shall be considered that the number of observations is lower for the car alternatives. However, it is in line with the literature corresponding to the Dutch context the respondents are expected to use cars for uni-modal trips rather than multi-modal trips (Krygsman & Dijst, 2001). It can be seen that bicycle substitutes walking as the most popular access and egress mode at a lower distance for trains than for lower-level transit networks (Shelat et al., 2018). It can be attributed to the fact that the availability of bike parking facilities is higher is the possibility of using cycles for access and egress. Individuals are willing to go to the fourth closest train station for access to the station if it is said to provide facilities and better connection such as lower number of transfer (Van Kampen et al., 2020).

7

Conclusions and Recommendation

7.1. Introduction

The goal of this chapter is to provide how the findings of the study can be used as insights for future policy development goals. Furthermore, from a methodological perspective, the approach and its limitations are also discussed. The aim of the research is to provide a theoretical framework to analyze mode choice models to deduce the implications of the mode choice behaviour. The framework is applied for the case study and the context of the city of Amsterdam. The outcome of this research is a framework that contributes to modelling the mode choice simultaneously for the different urban and regional networks.

7.2. Conclusions

7.2.1. Questions

Main question: How to model access and egress mode choice simultaneously (to and from) public transport systems for regional and urban networks?
Sub-questions:

1. What are the significant factors affecting the mode choice for access and egress to and from public transport in a multi-modal trip?
As depicted in the literature review chapter umpteen studies comprising of stated and revealed preference data collection systems there are umpteen factors. They are categorized in the following manner. The following classification is made and within them, the relevant factors are defined.

Table 7.1: List of attributes from literature

Category	Probable Attributes	Study
Individual characteristics	Employment, age, gender, license, income, ownership of cars, availability of free student travel, frequency of travel using various modes.	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Ton et al. (2019), Ton and Shelat (2020)
Household characteristics	Household size, number of vehicles per household	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Ton et al.(2019)
Trip characteristics	Distance, travel Time, Purpose, frequency of travel via different modes, origin and destination locations.	Mark R. et al. (1993), Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014),Yang et al. (2015), Ton and Shelat (2020), Van Kampen et al. (2020)
Level of service variables	Frequency, number of stops, in vehicle time, waiting time	Mark R. et al. (1993), Debrezion et al.(2007), Yang et al. (2015), Ton and Shelat (2020), Brand, et al. (2017)
Built environment	Area or region, Urban density	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Van Kampen et al. (2020),Ton and Shelat (2020)
Working condition	Working hours, Travel compensation	Van Kampen et al. (2020), Ton et al. (2019)
Weather characteristics	Month of travel	Ton et al. (2019)
Station choice	Origin and destination station, area region, type of stations, parking facilities	Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Van Kampen et al. (2020), Ton et al. (2019), Ton and Shelat (2020)

Umpteen studies depict various factors that are significantly based on the context and the focus of research. However, to test the hypothesis fewer factors are selected to showcase the applicability of the model as deduced from literature and the requirements from the Municipality of Amsterdam.

2. To what extent the Discrete choice modelling approach be implemented for modelling mode choice?
As discussed in chapter 3 the existing models are compatible with the discrete choice modelling approach. Within the choice modelling domain, many methods can be implemented but considering the current research, the suitable approach is determined. Considering the current context and the findings from the literature MNL model is implemented, then optimized using the ML model. This is done so structure the estimation process so as to make it computationally less intense. The MNL model is applied to deduce the significant variables and then ML model is applied with the error-components to account for the correlations between the alternatives. It is the most commonly used approach.

The revealed preference data is used to obtain the empirical analysis. In terms of data collection, various approaches can be considered. The type of data collected is revealed or stated preference survey. In this case, to maintain the compatibility with the existing software and method of data collection is revealed preferences. For the context of the Netherlands a travel survey namely ODiN is carried out annually. The objective of doing this survey is to provide data to analyze travel behaviour trends in the Netherlands. Furthermore, it is the data source implemented in the traffic model of Amsterdam.

The alternatives are generated for specific mode-chains. The mode chains are decided based on the most frequently used modes based on the research done in the similar context. To formulate the relevant alternatives certain assumptions are made to prevent the explosion of the number of alternatives formulated. To limit the number of alternatives trips with two inter-modal transfers are considered. The main modes are the public transport networks such as train, metro bus or tram ideally. The access and egress modes considered comprise a combination of public and private modes i.e. walking, cycle, metro, bus and tram. Hence ideally the following alternatives shall be considered.

- Walk- PT-Walk
- Walk-PT- Bicycle
- Walk -PT- BTM
- Walk- PT- Car
- Bicycle - PT-Walk
- Bicycle -PT- Bicycle
- Bicycle -PT- BTM
- Bicycle -PT- Car
- BTM - PT-Walk
- BTM -PT- Bicycle
- BTM -PT- BTM
- BTM - PT- Car
- Car- PT-Walk
- Car -PT- Bicycle
- Car -PT- BTM
- Car - PT- Car

Discrete choice models are expected to be suitable for modelling the access and egress mode choice simultaneously as it is widely used in literature and in practice the existing models.

3. How can a multi-modal trip be modelled considering the context of Amsterdam?
In the case of Amsterdam as answered in the previous question the choice set generation there are ideally 16 alternatives that shall be considered. In the case of the socio-demographic variables, a selected number of variables are selected for hypothesis testing. The socio-demographic variables are extracted from the ODiN database. The only alternative dependent parameter considered is travel time. As the limitation of the reported travel times is only available for the chosen alternatives. Hence to obtain values of the chosen and non-chosen alternatives the google API algorithm is implemented to extract the travel time of access and egress. To do as the centroids of the origin and destination location are input parameters leading to a bit of discrepancy in the exact travel time values. The main mode travel time is the same as the reported time.

The framework suggests taking a structured approach to deduce the various policy implications. Two models are analysed one with the train as the main mode and the other with the metro as the main mode of transport. The modelling is done using an MNL model and once the significant variables are deduced it is implemented in the final model and optimized. As travel time is the only parameter that varies with the alternatives specified a base model is analyzed with the travel time parameter i.e. access time, egress time and main mode and the alternative specific constants. The base model is deduced with the utility function as follows;

$$V_{base-model} = \beta_{TA} * traveltime_{access} + \beta_{MT} * traveltime_{main-mode} + \beta_{TE} * traveltime_{egress} + ASC_{AM-EM}$$

Where, β_{TA}, β_{TM}, β_{TE} are the estimated variables for access time, main mode travel time and egress travel time respectively for the chosen alternatives. ASC_{AM-EM} depicts the alternative specific constant, AM is the access mode and EM is the egress mode.

The next step is to add the socio-demographic variables to the base model individually to test the formulated hypothesis is significant; The formula is as follows;

$$V_{Hypothesis-test} = V_{base-model} + \beta_{attribute} * Attribute$$

$\beta_{attribute}$ is the coefficient that is estimated for the selected attribute for testing the particular hypothesis. Once all the significant variables are deduced a combined MNL model is formulated and optimized as follows;

$$V_{Combined-model(MNL)} = V_{base-model} + \beta_{attribute_{s1}} * Attribute_{s1} + \beta_{attribute_{s2}} * Attribute_{s2} + \ldots\ldots$$

For the final round of analysis, a Mixed Logit model is implemented to account for the expected correlation amongst the alternatives. This is expected as the mode across the modes-chains overlap.

$$V_{Combined-model(ML)} = V_{Combined-model(MNL)} + \epsilon_{AM} + \epsilon_{EM}$$

Where, ϵ_{AM} is the error component for access mode and ϵ_{EM} is the error component for the egress mode for the particular alternative. This approach is considered to be suitable as it allows to deduce the impact of the different socio-demographic variables and then combine it into one model.

4. What are the impacts of the factors affecting the access and egress mode choice for train and metro (as main mode)? What are the differences between them?
The various hypothesis are tested and in both train and metro models. Though the data depicts a similar trend the outcomes are different for both. It is observed that in the case of the individual model.

In the case of the train, the travel time parameters suggest that the egress travel time is weighted more highly and negatively as compared to the access time. The main mode travel time is negatively weighted but has the lowest values. This suggests that the access and egress travel time is weighted highly and has a higher impact on the mode choice. Longer access and egress travel time deters the passengers from using a particular public transport mode.

In the case of the metro model, the access is weighted more heavily as compared to the egress time. It is a plausible finding as the availability of the metro services is only within the city of Amsterdam and is a dense network hence egressing is in an urban and dense area; making it a more walk-able environment. The access depends on the availability of the vehicles at the home end. Furthermore, for the trips made outside Amsterdam, the distance and duration for access/egress are longer.

For the metro model, there is also the limitation of the data available. As the data regarding the station choice was not available, based on the reported travel time and the mode chosen the station was deduced. Additionally, the travel time of the main mode has a positive coefficient which is insignificant. This is expected as the metro network is an urban network and the trips that originate and end outside Amsterdam are likely to higher travel times.

In case of the train model, for the policy hypothesis tested, individually the outcomes are in line with the literature. Car ownership is a significant variable. In the final MNL model, the significant variables are car ownership, age group 18-24 years, trip purpose and full-time work. Car ownership has the highest value of the coefficient. Suggesting that if the car is available, people are more likely to use it for access and egress. The age group hypothesis suggests that the people belonging to the age group of 18-24 are more likely to use public transport for access and egress. It is in line with literature wherein it is suggested that the people who work full time and are highly educated prefer to use active modes for access and egress. It is also interesting to see that given the Dutch context the use of cycles is much higher here. In addition to that having a full-time job suggests that the work timings are expected to be fixed throughout the week.

In the case of the metro model, car ownership is the only significant factor that remains. It suggests that the metro has a higher chance of being a competing mode to metro than an access and egress mode. A similar trend can be also observed for cycling trips the number of observation cycles as access and egress mode with respect to the metro is to access the regions within the city it is preferred to use that as the single mode of transport.

Furthermore, a more complex and advanced ML(Mixed Logit) model with the random error component is introduced to capture correlations across the alternatives having same access and egress modes. However, for the train and metro model the error components turned out to be insignificant suggesting that though the mode chain behaves as a synthetic combined mode itself for the user.

7.2.2. Conclusion

As suggested in the literature there is umpteen research in the field of modelling of access and egress mode choice. However, the simultaneous choice is not investigated in that depth. Furthermore, for an urban setting, the lower-level networks play an important role. Currently, the four-step model with the discrete choice modelling is adopted in western European countries. Hence it is necessary to deduce a compatible model compatible with the existing framework. Thus, discrete choice models are adopted for analysis. In the case of the Netherlands, the National and regional models are strategic models are analysed. For the larger urban regions, Amsterdam, Rotterdam and Hague area etc. have models. Literature suggests various factors are significantly based on studies carried out in different contexts and modelling methods.

The data used in this study is ODiN (on der Weg) it is a study conducted by CBS on a yearly basis. The advantage of this methodological approach is that it is designed to be flexible and applicable based on what policy implications require to be tested.

Comparing the systems of the metro and train the cycle also in a way is a competing mode on an urban level for the metro. For the visitors within the city, it is more comfortable to use bikes. It is important to note that the dominant use of cycle is a function of the cultural context and it is not always expected to be the case for other countries.

It is an adaptable framework for any city overall. The main point of difference is the contextual parameters that affect the choice set generated. Additionally, it is a more applicable system in an urban context as it is expected that the availability of public transport choices is higher in urban regions. There has been a growth in car ownership over the period time however the cycling has remained a constant mode.

The socio-demographic variables impact the access and egress mode choice for trains more prominently as compared to the metro, in the case of Amsterdam. One reason might be that the metro has a higher ability to be a feeder mode for the trains rather than the main mode. For residents within the city, it can be expected that there are other choices such as cycles that provide them with a higher convenience to travel rather than make transfers. However, they are contextual parameters, and this might not be the case in other cities and other modes might be more dominant.

Simultaneous access and egress mode choice modelling depicts how the travel time attribute is weighted among the different trip legs namely access, egress and main trip leg. Moreover, the analysis of the results suggest that the the perception of travel time access and egress to and from different modes of public transport varies. For the train model the individuals are less sensitive to access travel time as compared to egress and main mode trip legs. Where as in case of the metro it is the opposite for access and egress travel time. In addition to that the main mode travel time have completely different perceptions.

The dominance of the car ownership variable might also suggest that having a car and travelling from out of the city is a more convenient alternative than using the metro. The metro is used in conjugation with walking and other public transport modes.

The are many modelling approaches that can be adopted, studies suggest that hybrid modelling processes can also be adopted or other more complex and advanced models can be used. The trade-off is of course is the computation time. However in this case a more complex ML model was considered and the results depict that there is no correlation amongst the alternatives having the same access and egress modes.

Limitations

The pre-specified mode chains are extracted as choice alternatives. As the literature suggests there are a lot of permutations and combinations that can be generated, the feasible mode chains are determined based on the studies in the existing context and based on the data available. The number of alternatives is limited. It is a rather complex problem to address mathematically, and a particular approach is only considered. Not all the provided alternatives need to be always available to all individuals. Considering that the Netherlands has a pro cycling approach it can be assumed that cycling is an alternative that is always available. As the data suggests that the people using other modes of transport also have access to cars.

Furthermore, the integration of the route choice is also the feedback loop that is taken into consideration. The research addresses the mode choice perspective for a transport model and not the route choice component. It is assumed that the route chosen by the individual is the optimal one. Depending on the parameter that is to be optimized for the route choice such as travel time, travel costs and distance etc. there are multiple approaches that can be taken.

The data used is a travel survey which is a form of revealed preference survey. Though a lot of data is available from the data, newer ways of data collection systems such as GPS, GFTS would be able to provide more real-time information. Also, the values are reported so there might be inconsistencies in the real value and the reported value. Thus, the use of the data to make more accurate predictions would be preferred.

The travel time values are deduced using the google API algorithm which is not exactly the reported value. Additionally, the travel costs are not considered in the case only travel time is the variable considered and is expected to vary amongst the alternatives.

The validation of the model is not done due to the limitation in time and data. It would be useful to validate the results of the model using an external data source to validate it. Moreover, techniques such as the K-means clustering can be adopted to determine the predictive power of the model.

7.3. Discussion

The framework is adaptable to the different systems of public transport. Moreover the aim is to provide a framework that can be applied in different cultural contexts as well. The points to be noted for the choice set generation considered is a similar approach to other studies having maximum two intra-modal transfers. As the trip becomes more complex than that the likelihood of making a more complex trip reduces.

Comparing the travel time parameters weight of travel time access and egress is higher as compared to the main mode for both train and metro. This would suggest that more focus in the decision making process of mode selected is given to access and egress. Hence, to encourage the users to adopt public transport as their main mode the access and egress must be the emphasized upon. Average travel conditions are used in the case study, it can be expected the for off peak hours the overall value of access and egress and main mode value would be higher. Factors such as crowding, waiting etc become more prominent during peak hours.

In terms of the socio-demographic variables , it seems to have a higher impact on the access and egress mode choice in case of the train model. It is expected to be a function of the availability of such a network in the entire country, In case of the metro it is only available within the city. However, for a different city in in a different country the results might not be the same. In case of the modelling approach as suggested in the Chapter 2, there are various methods. In the current research the starting point for modelling is the starting point is the MNL model. The limitation of the MNL model is that it is assumed that the all alternatives are independent, which in reality is not always the case.

To account for the correlations an ML model is is further analysed. The results suggests that the correlations are insignificant, thus suggesting that the alternatives are independent. It depicts that the individual takes the whole trip into consideration when choosing the mode. It is an interesting insight that the whole trip is considered as a single entity without any correlations within the alternatives having the same access and egress modes. It has to be noted that it is what the data represents in this context. When applying a similar framework to another context such as a different city it is very likely that there are correlations between the alternatives.

Thus it can be deduced that modelling access and egress mode choice provides insights that the depending the context and the main mode of public transport the access and egress mode choice behavior changes. To integrate the access and egress modes more effectively the factors affecting mode choice for the different modes of public transport shall be considered. As depicted in that analysis the travel time is perceived differently for the train and metro model. There are un-observed attributes that lead to such results.

Car-ownership is the only relevant attribute for metro. It can be interpreted that as the trips to and from outside Amsterdam are considered it emphazized on the car availability to access the metro system which is only available in Amsterdam. If the case study only comprised trips within the city of Amsterdam, it can be expected that car ownership might not matter as much but factors such as frequency of cycle usage, availability of cycle parking infrastructure etc might be more important for both the train and metro model. Moreover, as the context is the Netherlands that cycling becomes a competing mode to Public transport in the city rather than a complementing mode. This is expected as it can be observed from the data that the share of trips with cycle as the access and egress mode is not high especially as compared to the train. Another significant parameter in addition to travel time is also travel costs. walking for access and egress is the popular choice amongst the individuals.

Besides focusing on the technical aspects such as synchronizing the time-tables, optimizing the frequencies it is useful to understand underlying behavioural traits. Market segmentation of the users can be carried out to understand the user profiles of the the different modes. To implement the model for another context such as another Dutch it is likely that city such as Rotterdam having a similar size and availability of infrastructure might produce similar results. However if the framework is transferred to another European city it is necessary to consider the generic modes of transport in that region to determine the alternatives.

7.4. Further research

As deduced in the first question there are many factors in literature that affect mode choice. However, it depends on the requirement and the policy insights required the model is expected to be applicable. Hence in the current analysis, not all the factors are considered. In order to analyze the impact of mobility hubs additional variables such as the availability of parking spaces, accessibility to P+R facilities, Level of service of the public transport services etc can be added to the model and it can be used to analyse and test the hypothesis. Additionally, the analysis of the impacts of the various shared micro-mobility platforms can be done using the same framework but with different access and egress choices.

Further research can be done to analyze how the metro system or an urban system competes with cycles as the main mode of transport within the urban regions. As suggested by the results of the model, in the case of the metro model the socio-demographic variables are insignificant. Additionally, micro-mobility points are becoming increasingly popular. However, it must be considered that the alternatives such as trip purpose plays an important role. Moreover, contextual variables such as the willingness to cycle for a certain distance and time matter. In the Dutch context, cycling is en-grained as a part of the lifestyle hence it is widely accepted and adopted. Thus, the analysis suggests the car ownership is a very important factor and to encourage multi-modal trips would be something that can be ventured into further. Additionally, the research how such urban transport systems and cycles at time act as more competing modes rather than complementary modes. It can be studied further.

Travel cost is also a parameter that has a significant impact on mode choice and varies with the alternative chosen. Similar analysis can be carried out considering travel costs such as fuel cost, parking costs etc. instead of travel time. The results could be useful input to design pricing schemes to encourage more public transport users. As this study focuses on the mode choice perspective the route choice can be further integrated with the system. Additionally, the results from travel time and travel cost comprising models can be compared to understand what the individuals are more sensitive to.

Points regarding the competing station choice integration are not considered. For example certain station choices would optimise the overall travel time of the individual as compared to the one chosen by them. Hence an additional step to integrate decision making for the station choice with the mode choice can be investigated. In theory the individual might have better options to chose for access and egress and may be optimum in case of time and convenience. They might not be aware of all the available alternatives.

Using the real-time data and integrating the route choice perspective would be a more advanced method such as heuristics that can be implemented to deduce the choice set generation. The ML model suggests for the current research that there are no correlations between the alternatives. Studies such as Wen et.al.,(2012) depicts the use of hybrid models to capture the un-observed latent variables. A similar approach creating a hybrid model that allows for accounting for the taste heterogeneity amongst the alternatives can be carried out to deduce if it is truly the case that the alternatives are independent though the access and egress mode is common amongst the alternatives.

In the current studies the revealed preference survey ODiN is the data source. For analysing particular factors that might be attitude or perception based can be added to the exiting framework by making a combined revealed and stated preference structure. Furthermore, the insights from the current research can also be used to determine the choice situations.

7.5. Contribution to science and practice

This section describes what the findings of the research contribute to science and practice.

7.5.1. Value to science

The methodological approach allows for modelling access and egress mode choice simultaneously. As depicted in chapter 1 most studies focus on access or egress but the simultaneous access and egress mode choice modelling is not addressed in a greater depth. The current research provides a starting point to develop it further. The other added value of this research is that it provides a transferable framework that can be applied to other networks. Furthermore, it can help compare how the access and egress mode choice behaviour differs across public transport modes.

Moreover, it is expected to be transferred to the different modes of public transport. It also depicts that all modes of public transport do not exhibit the same kind of mode choice behaviour. When designing policies for a more sustainable and car free environment and encourage public transport the different factors that can be focused on can be deduced.

7.5.2. Value to practice

As a discrete choice modelling approach is implemented it is expected to be more compatible with the existing systems. In the transport modes used in practice the 'discrete choice modelling' method is the most popular approach. The modelling approach provides sort of a tool to allowing to model access and egress mode choice simultaneously so it can be applied when framing policies to integrate the different modes. Furthermore, based on the requirement or policy goal allows to test different hypothesis so as to understand if it is significant. Hence on a strategic level is useful for practitioners to implement and use.

It is an important insight that suggests that the all modes of public transport do not exhibit a similar mode choice behaviour. Hence, having more tailor made policies could be more effective to combine modes more seamlessly. To provide a more well integrated network with public and private modes the behavioural aspects play an important role. The study depicts that the different forms of public transport cannot be treated same. It depicts which part of the trip leg impacts the access egress mode choice and more focused research can be done to find what improvements are needed to nudge individuals towards more sustainable modes. As it is suggested in the study that the trip purpose and employment status play an important role. Hence policies that enable employer's to encourage the use of public transport can be implemented e.g. designing pricing schemes that favour the the employers to provide better public transport access to their employees.

For future developments such as extension of the metro or train lines it might be useful to know if consider scenarios where in it can be checked if the adding more facilities is viable to dig deeper into infrastructure development. Moreover, the government can be aware of the factors that shall be taken into consideration during the feasibility studies of the development of such large scale infrastructure projects.

Bibliography

Arentze, T. A., & Molin, E. J. E. (2013). Travelers ' preferences in multimodal networks : Design and results of a comprehensive series of choice experiments. *TRANSPORTATION RESEARCH PART A*, *58*, 15–28. https://doi.org/10.1016/j.tra.2013.10.005

Azimi, G., Rahimi, A., Lee, M., & Jin, X. (2020). Mode choice behavior for access and egress connection to transit services. *International Journal of Transportation Science and Technology*, (December). https://doi.org/10.1016/j.ijtst.2020.11.004

Brand, J., Hoogendoorn, S., Van Oort, N., & Schalkwijk, B. (2017). Modelling multimodal transit networks integration of bus networks with walking and cycling. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings*, 750–755. https://doi.org/10.1109/MTITS.2017.8005612

Brands, T., De Romph, E., Veitch, T., & Cook, J. (2014). Modelling Public Transport Route Choice, with Multiple Access and Egress Modes. *Transportation Research Procedia*, *1*(1), 12–23. https://doi.org/10.1016/j.trpro.2014.07.003

Brands, T., Dixit, M., & van Oort, N. (2020). Impact of a new metro line in amsterdam on ridership, travel times, reliability and societal costs and benefits. *European Journal of Transport and Infrastructure Research*, *20*, 335–353.

Chakour, V., & Eluru, N. (2014). Analyzing commuter train user behavior: A decision framework for access mode and station choice. *Transportation*, *41*(1), 211–228. https://doi.org/10.1007/s11116-013-9509-y

Fiorenzo-Catalano, M. S. (2007). *Choice Set Generation in Multi-Modal Transportation Networks*.

Givoni, M., & Rietveld, P. (2007). The access journey to the railway station and its role in passengers' satisfaction with rail travel. *Transport Policy*, *14*(5), 357–365. https://doi.org/10.1016/j.tranpol.2007.04.004

Harms, L., & Kansen, M. (2018). Cycling Facts Netherlands Institute for Transport Policy Analysis | KiM. *Institute for Transport Policy Analysis, KiM, Ministry of Infrastructure and Water Management,*

Hensher, D. A. (2005). Stated Preference analysis of Travel Choices: The State Of Practice. *Transport Economics, Selected Readings*, 62–82. https://doi.org/10.4324/9780203985359-12

Hensher, D. A., & Reyes, A. J. (2006). Trip chaining as a barrier to the propensity to use public transport, 341–361.

Jonkeren, O., Kager, R., Harms, L., & te Brömmelstroet, M. (2019). The bicycle-train travellers in the Netherlands: personal profiles and travel choices. *Transportation*, (0123456789). https://doi.org/10.1007/s11116-019-10061-3

Kagho, G. O., Balac, M., & Axhausen, K. W. (2020). Agent-Based Models in Transport Planning: Current State, Issues, Expectations. *Procedia Computer Science*, *170*, 726–732. https://doi.org/10.1016/j.procs.2020.03.164

Keuchel, S., & Richter, C. (2011). Applying Integrated Hierarchical Information Integration to Mode Choice Modelling in Public Transport. *20*, 875–884. https://doi.org/10.1016/j.sbspro.2011.08.096

Kim, K. (2018). Key Transport Figures 2018 for Transport.

Krygsman, S., & Dijst, M. (2001). Multimodal trips in the netherlands: Microlevel individual attributes and residential context. *Transportation Research Record*, *1753*(1), 11–19. https://doi.org/10.3141/1753-02

Li, X., Tang, J., Hu, X., & Wang, W. (2020). Assessing intercity multimodal choice behavior in a Touristy City: A factor analysis. *Journal of Transport Geography*, *86*(November 2019), 102776. https://doi.org/10.1016/j.jtrangeo.2020.102776

Mark, R., Cakmak, A. S., Hill, K., & Davidson, R. (1993). Transactions on the Built Environment vol 3, © 1993 WIT Press, www.witpress.com, ISSN 1743-3509. *3*.

Ortúzar, J. d. D., & Willumsen, L. G. (2011). Modelling Transport. https://doi.org/10.1002/9781119993308

Public transport in the Netherlands,Ministry of Transport, Public Works and Water Management. (2010). (June).

Rich, J., Nielsen, O. A., Brems, C., & Hansen, C. O. (2010). Overall design of the Danish National transport model. *Annual Transport Conference at Aalborg University*, 1–14.

Shelat, S., Huisman, R., & van Oort, N. (2018). Analysing the trip and user characteristics of the combined bicycle and transit mode. *Research in Transportation Economics*, *69*(June), 68–76. https://doi.org/10.1016/j.retrec.2018.07.017

Ton, D. (2014). *Dissertation title* (Doctoral dissertation). University of Somewhere.

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice*, *123*(August 2018), 7–23. https://doi.org/10.1016/j.tra.2018.08.023

Ton, D., Shelat, S., Nijënstein, S., Rijsman, L., van Oort, N., & Hoogendoorn, S. (2020). Understanding the Role of Cycling to Urban Transit Stations through a Simultaneous Access Mode and Station Choice Model. *Transportation Research Record*, *2674*(8), 823–835. https://doi.org/10.1177/0361198120925076

Train, K. (2002). Discrete Choice Methods with Simulation.

Transport for London. (2018). London's Strategic Transport Models. http://content.tfl.gov.uk/londons-strategic-transport-models.pdf

Van Kampen, J., Jayaraj, M. A., Pauwels, E., Van Der Mei, R., & Dugundji, E. R. (2020). Understanding the relation between distance and train station choice behavior of cyclists in the western region of the Netherlands. *Procedia Computer Science*, *170*, 98–106. https://doi.org/10.1016/j.procs.2020.03.013

van Eck, G., Brands, T., Wismans, L., & Nes, R. (2014). Model complexities and requirements for multimodal transport network design. *Transportation Research Record Journal of the Transportation Research Board*, *2429*, 178–187. https://doi.org/10.3141/2429-19

Waerden, P. V. D., & Waerden, J. V. D. (2018). The Relation between Train Access Mode Attributes and Travelers ' Transport Mode-Choice Decisions in the Context of Medium- and Long-Distance Trips in the. *2672*(8), 719–730. https://doi.org/10.1177/0361198118801346

Wen, C. H., Wang, W. C., & Fu, C. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review*, *48*(2), 545–554. https://doi.org/10.1016/j.tre.2011.09.002

Yang, H., Feng, J., Dijst, M., & Ettema, D. (2019). Mode choice in access and egress stages of high-speed railway travelers in china. *Journal of Transport and Land Use*, *12*(1), 701–721. https://doi.org/10.5198/jtlu.2019.1420

Yang, M., Zhao, J., Wang, W., Liu, Z., & Li, Z. (2015). Metro commuters' satisfaction in multi-type access and egress transferring groups. *Transportation Research Part D: Transport and Environment*, *34*, 179–194. https://doi.org/10.1016/j.trd.2014.11.004

Young, M., Blainey, S., Young, M., & Blainey, S. (2018). Railway station choice modelling : a review of methods and evidence evidence. *Transport Reviews*, *38*(2), 232–251. https://doi.org/10.1080/01441647.2017.1326537

A

Appendix - A

Modelling access and egress mode choice for multi-modal trips : Case study of Amsterdam

A Preprint

Reddhi Uday Bhatt
Department of Civil Engineering and Geo-sciences
Delft university of Technology
Mekelweg 5, 2628 CD Delft
R.U.Bhatt@student.tudelft.nl

November 3, 2021

Abstract

With the infrastructure development in the cities and urbanisation there has been a need to manage the car traffic to make the cities more livable. Moreover to improve the accessibility within urban areas providing a well integrated network of public transport is necessary. To encourage the users to use public transport, understanding the access and egress mode choice behaviour is crucial. In the current state of the art access and egress mode choice is not modelled simultaneously. Furthermore, it is often assumed that the access and egress mode choice behaviour is similar for different public transport mode, but is not always the case. Thus, to address these research gaps the current research is carried out to provide a modelling approach that allows modelling access and egress simultaneously. Additionally, compare different modes of public transport i.e. train and metro to deduce if the factors affecting mode choice are the same. A discrete choice modelling approach is implemented to estimate the models. Revealed preference data source ODiN (On der weg) is used for the analysis and estimation of the model. The modelling process comprises of a base model onto which the different hypothesis is tested to deduce the factors that impact mode choice. Significant parameters are considered and a combined model is estimated for train and metro as the main mode. The findings suggest that travel time is experienced differently in the access, egress and main leg of the trip for both train and metro. Further more, the way time is experienced for access and egress is not the same for metro and train.

Keywords access · egress · simultaneous mode choice · travel time

1 Introduction

In urbanized regions, the increased use of cars has led to traffic and congestion problems. To solve this issue, policies encouraging the users to adopt a more integrated network with public transport are necessary. Considering the case of the Netherlands, where the bike infrastructure is well-developed, cars still have the highest share i.e. 47% Harms and Kansen [2018] as compared to other modes.

However, with time there has been an increase in the use of cycles overall especially for access and egress to and from the train stations. This indicates the willingness of the users to shift towards more sustainable modes. To integrate the modes and provide a more seamless experience more insights into multi-modal transport need to be investigated. Hence, the study aims to understand the mode choices of travellers. Usually, the trips that comprise the use of public transport are multi-modal. Thus, improving the door-to-door connectivity and accessibility to public transport is the need of the hour. In order to understand the implications of certain policies to improve the current conditions and improve accessibility transport models are used. Transport models are created as tools that can be used for decision-making. The information transport models are expected to provide are impacts of policies relating to implications of alternative transport on land-use investments and policies etc.

To achieve this aim modelling approaches that take the multi-modal choice behaviour. Transport models are usually implemented to deduce policy implications. Four-step models are popular for the purpose of modelling the various choice dimensions [Ton et al., 2019]. In western European countries such as Sweden, Netherlands and Denmark, the adoption of a four-step transport model is popular. With the development in research and technology, there are different models such as activity-based models, agent-based models etc. Umpteen modelling approaches are available but discrete choice modelling approach is often implemented [Ton et al., 2019]. However, the current systems are based on the classical four-step model. Hence, the current study is focused on implementing the discrete choice modelling for mode choice models to maintain compatibility with the existing systems.

1.1 Access and egress mode choice

As discussed in the previous section the world is willing to move towards a more car free economy. To implement such policies understanding of the travel behaviour is necessary. Mode choice is a key step to deduce policy implications [Ortúzar and Willumsen, 2011]. Thus, to gain a more realistic perspective of the mode choice involving a multi-modal trip chain, it is essential to understand the mode choices for access and egress travel to and from public transport. To explain access and egress a multi-modal trip is defined. A trip is defined as the travel from an origin to a destination. A trip comprising the use of more than one mode is a multi-modal trip. Walking, cycling, public transport and car are the different modes considered [Fiorenzo-Catalano, 2007]. Thus, a multi-modal trip can be further divided into three parts i.e. access leg, main leg and egress leg [Fiorenzo-Catalano, 2007]. The access leg is the part of the trip from origin to the transfer node to the main leg. Similarly, the egress leg is the transfer from the main leg trip to a destination. In the case of the multi-modal perspective, the various transport systems are expected to be integrated in a way that there is a smooth transfer between the different modes. The following paragraphs describes what the current approach in research is carried out to model access and egress mode choice.

Studies focus on either access mode or egress mode individually or both separately. In the study conducted by Wen et al (2012) focuses on access mode choice only for high speed rail [Wen et al., 2012]. The (Nested Logit) NL model with latent classes is estimated to deduce the preferences of the users. The findings suggest that the travelers are cost sensitive. The future research shall be carried out to deduce a more generalized model allowing the integration of access and egress mode choices. By analysing access and egress mode choice separately, it is deduced that the distance plays an important role in the choice of access and egress mode [Krygsman and Dijst, 2001]. Ton et. al. (2020) investigates the factors affecting access mode choice along with station choice for cycling and walking as access modes. The study focuses on the factors affecting the access and station choice from tram as the main mode. However, for other public transit modes it is expected that the same factors might not be relevant. The mode for access and egress is cycling and walking. MNL (Multinomial Logit Model) model based on the distance of access and station. The study by Azami et. al. (2020) focuses on the region of Orlando to model access and egress to and from public transport. The factors affecting mode choice for the different modes including the micro-mobility modes are deduced [Azimi et al., 2020]. Two separate models are implemented for access and egress mode choice. MNL models are implemented for analyses of the model. The findings are limited to the context of the region. Additionally, the transfer-ability is not is not accessed. In this study it is assumed that all the forms of public transport are expected to have the same factors affecting mode choice [Azimi et al., 2020]. Hence, one of the limitations of the above-mentioned studies is that simultaneous access and egress mode choice is not considered.

A study conducted in China multi-modal choice behavior is modelled for intercity travel. The main modes comprise of airplane, train, express-bus and HSR (High Speed Railway). To model the factors affecting mode choice, to deduce the factors affecting mode choice, access and departure mode choice are considered. BMNL (Bayasian Multinomial Logit) model is implemented which is a Bayesian based discrete choice models [Yang et al., 2015]. This research considers access mode choice and departure mode choice as factors that effect the main mode choice. Thus, it provides insights how the access and egress are weighed depending on the different main modes available. Thus, it is a variable and the alternatives are the main modes, but does not address modelling the access and egress simultaneously. Moreover, it is limited to the context of the location of the case study. The research done by Waerden et. al.(2018) provides insights into the role of main modes and access modes on the decision to chose between car or train as the prime mode of travel. Hence, access and egress is not the main focus though it is considered. The study mainly focuses on access mode choice with the main travel mode. The trade off between train and car as the main mode is deduced. Mixed logit (ML) model is estimated to deduce the implications [Waerden and Waerden, 2018]. It depicts that the attributes related time and cost are influential. It provides insights into which factors shall be considered to encourage the car users to switch to train. However, various factors such as the trip purpose are not considered. The study by Yang et al (2019) focuses on access and egress mode choice to and from the high speed rail in China. The modelling of access and egress mode choice is done is separate stages. Separate models are analyzed for the business and leisure travelers [Yang et al., 2019]. Studies also determine the trade-off between uni-modal and multi-modal trips in the context of the Netherlands. Furthermore, the trade off between the service quality, travel time and travel costs are deduced [Arentze and Molin, 2013]. Insights into how the different stages of the trip such as access time, egress time and in-vehicle time are weighted is provided. However, in this case the public transport is considered as one mode and no differentiation amongst them is not carried out.

Hence, most studies focus on access or egress individually. Some consider them as factors for a generalized approach towards public transport. The study by Yang et al (2015) provide a more comprehensive approach towards mode choice modelling for access and egress mode choice for the metro in the city of Nanajing. The focus is on the satisfaction of metro commuters. In this case the focus is only on one mode of public transport. Most studies consider all the modes of public transport as one assuming the that the behavior amongst all the modes of public transport is same, or the focus is only on a particular mode of public transport. This suggests there is not much research in the comparison between the travel behavior between the modes. Studies do not necessarily compare the different public transport modes (within the same context) to deduce if similar factors affect the access and egress mode choice. Form a societal perspective to encourage the use of public transport, it would be fruitful to gain insights into what are the differences are so as to provide tailor made policies addressing the different modes of public transport. Furthermore, there is a need for a more generalized modeling approach that integrates access and egress mode choice to determine the factors affecting the access and egress mode choice[Wen et al., 2012].Hence, the simultaneous access and egress mode choice is not investigated in depth. Thus, a comprehensive method to model access and egress in a single trip and analyse the required factors in the given context is the existing gap.

The state of the art is provide a modelling framework that allows for the modelling of simultaneous access and egress mode choice. Moreover, provides insights into the factors affecting the mode choice in a single model. Secondly, provide an approach that is transferable to different public transport modes. Moreover, deduce if the assumption that mode choice behaviour for the different public transport modes is the same.
The first step is to deduce the factors affecting the access and egress mode choice as per the current state of the art. The section2, depicts the different factors that affects access and egress mode choice.

2 Literature review

To determine the relevant variables to be considered when modelling access and egress travel behaviour the existing literature is reviewed and the factors that can be considered are classified and determined in this section. To realistically model the user's choices, the attributes impacting the individual's mode choices must be considered. The mode choice factors are classified into three parts; the first is the characteristics of the trip maker (it comprises age, income, availability of car license, etc) [Ortúzar and Willumsen, 2011]. Second is the characteristics of the journey such as trip purpose, time of day, group size, etc. The third is the characteristic of the transport facility travel time, distance, in-vehicle time, etc [Ortúzar and Willumsen, 2011].

Moreover, these sections can be sub-divided into more classifications. The division of the different categories for mode choice can be made into three parts i.e. multi-modal travel variables, socio-demographic variables and residential environmental variables [Krygsman and Dijst, 2001]. The study by Yang et al (2015) carried out to determine the satisfaction of metro commuters divides the factors into the personal attributes and the journey details [Yang et al., 2015]. Some studies consider the personal characteristics as a category and define the attributes per particular mode [Yang et al., 2019]. However, in the case of mode choice involving active modes (i.e. cycling and walking), the factors affecting access and egress mode choice can be categorised as; characteristics of individual characteristics, household characteristics, built environment, season and weather characteristics, work conditions, and trip characteristics [Ton et al., 2019, Shelat et al., 2018]. Similarly, in a multi-modal trip, the factors are allocated to categories of variables, such as socio-demographic variables, LoS (Level of Service) parameters, trip characteristics, land-use and built environment factors, and station characteristics [Chakour and Eluru, 2014]. Thus, based on the approaches considered in literature; the following classification is made for the factors affecting the access and egress mode choice.

- Individual characteristics
- Household characteristics

- Built environment
- Work conditions
- Trip characteristics
- Station characteristics
- LOS (Level of service)

The following subsections describe the significant variables for the provided classifications based on the current state of art. Moreover, the variables belonging to the various categories as depicted in the table as follows;

Table 1: List of attributes from literature

Category	Probable Attributes	Study
Individual characteristics	Employment, age, gender, license, income, ownership of cars, availability of free student travel, frequency of travel using various modes.	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Ton et al. (2019), Ton and Shelat (2020)
Household characteristics	Household size, number of vehicles per household	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Ton et al.(2019)
Trip characteristics	Distance, travel Time, Purpose, frequency of travel via different modes, origin and destination locations.	Mark R. et al. (1993), Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014),Yang et al. (2015), Ton and Shelat (2020), Van Kampen et al. (2020)
Level of service variables	Frequency, number of stops, in vehicle time, waiting time	Mark R. et al. (1993), Debrezion et al.(2007), Yang et al. (2015), Ton and Shelat (2020), Brand, et al. (2017)
Built environment	Area or region, Urban density	Mark R. et al. (1993), Chakour, V., & Eluru, N. (2014), Van Kampen et al. (2020),Ton and Shelat (2020)
Working condition	Working hours, Travel compensation	Van Kampen et al. (2020), Ton et al. (2019)
Weather characteristics	Month of travel	Ton et al. (2019)
Station choice	Origin and destination station, area region, type of stations, parking facilities	Debrezion et al.(2007), Chakour, V., & Eluru, N. (2014), Yang et al. (2015), Van Kampen et al. (2020), Ton et al. (2019), Ton and Shelat (2020)

Factors affecting mode choice selected for analysis As depicted in the table 2 literature depicts that there are umpteen factors that affect mode choice. However, there are limitations pertaining to data availability and time constraints few relevant factors are selected to be tested. The important factors based the current literature review are deduced in the following paragraph.

In case of individual characteristics, their impact are context dependent. Car ownership is a significant variable in relation to mode choice [Krygsman and Dijst, 2001, Arentze and Molin, 2013]. Moreover, age and gender in many cases are significant [Van Kampen et al., 2020]. In the study [Azimi et al., 2020] the of the GFTS (General Transit Feed Specification) and the GIS (Geological Information Systems) data is utilized to make separate models for access and egress. The findings suggest that car ownership, age, income are significant, access and egress distance [Azimi et al., 2020]. Similarly, in the study by Yang et al. the access and egress mode choice preferences were collected via a stated preference survey for the HSR (High speed Rail) corridor in Shanghai China. Two separate models were analyzed for the access and the egress stage for business and leisure travellers [Yang et al., 2019]. The findings suggest that car ownership, age, income , access and egress distance are significant. In the case of the business travellers age, income affects the choice of the users [Yang et al., 2019].

For the Household characteristics, factors such as the household composition determines the size of the household and the travel group. Households comprising of children can be expected to prefer private modes as it would more convenient for them to make complex trips [Hensher, 2005]. Many studies suggest that the mode choice behaviour is highly impacted by trip purpose [Yang et al., 2019]. In association to that the employment status and the work hours impacts the attitudes towards mode choice.
The built environment and the urbanity class in a way depicts the accessibility and attitudes towards modes [Krygsman and Dijst, 2001]. The influence of the trip characteristics is significant as well. The analysis of train commuter behaviour is carried out by latent segmentation method [Chakour and Eluru, 2014]. Wherein simultaneously two segments are considered; one is the station first and access mode second and the segment 2, with the access mode first station second. Findings suggest that the travel time is a significant indicator, a better level of service at the station is influential [Chakour and Eluru, 2014]. Additionally, the built environment plays an important role in the business and leisure travellers are less sensitive to the egress time as compared to the access time [Yang et al., 2019].

A stated preference survey is carried out to investigate travel costs and travel time trade-off that the travellers make when choosing the modes of travel available [Arentze and Molin, 2013]. Moreover, as the travel distance increases the chances of multi-modal transport increases more dominant in medium-range trips [Arentze and Molin, 2013, Krygsman and Dijst, 2001]. Modes such as walking, has a strong distance decay function, the users drop off substantially after 2.5 km (approximately the maximum walking distance) [Krygsman and Dijst, 2001]. On the egress side, the problem of asymmetric mode availability is evident, with more users having to rely on walking and public transport to reach their final destinations. Many travellers substitute the bicycle or private car for walking, bus, or taxi. Hence, depending on the mode of transport and the distance the selection of particular access and egress mode is affected. Moreover, prefer using cycles for egress distances over bicycles due to the uncertainty in cycle availability on the egress end [Shelat et al., 2018].
Thus the following attributes are considered;

- Car ownership
- Gender
- Age
- Household size
- Trip purpose
- Employment status
- Urbanity class

2.1 Modelling approach

The discrete choice modelling approach will be adopted so as to maintain the compatibility with the existing systems. The following paragraphs describe the different mathematical models used in literature for modelling access and egress mode choice

MNL (Multinomial Logit) models are adopted by a majority of the studies to analyse access and egress mode choice behaviour [Azimi et al., 2020].Hence, the MNL model is considered suitable for deducing the access and egress mode choice. In the given study by Azimi et al., the access and egress are modelled separately. This model is chosen as it is considered that the alternative mutually exclusive. Moreover, the computation time is lower due to the closed-form mathematical structure of MNL models [Young et al., 2018]. To deduce the mode choice for metro commuters, the access and egress mode choice MNL model is adopted for the four modes (i.e. subway, bus, taxi, car) considering car as the reference mode [Yang et al., 2019]. Furthermore, study by Ton et al. to deduce the determinants for the use of active mode for access and egress mode choice where the MNL model is used to determine the significant variables as it has a more efficient computation time. Once the significant variables are used to determined, an MMNL (Mixed Multinomial Logit) is applied as the model fit is better (which is tested using the likelihood ratio test).

The study by Arentze and Molin implements a ML (Mixed multinomial Logit) framework to model travellers preferences towards multi-modal networks [Arentze and Molin, 2013]. In the above mentioned study, to account for the correlated terms between the main modes a error component is added to the utilities of the alternatives having the same main mode. Thus, a 'shared error component' is added to the utility. This enables to account for the common unobserved attributes amongst the different modes [Arentze and Molin, 2013]. Similarities on the level of access and egress stages are not taken into consideration. To have a parsimonious approach, only the important sources of covariance are considered in the model [Arentze and Molin, 2013]. The added advantage of the error component model structure is that it provides more flexibility as compared to hierarchically nested model structures [Arentze and Molin, 2013].

To analyse the access mode choice behaviour to the high speed railway in Taiwan, a latent class NL (Nested Logit) model is adopted. Contrary to the popular choice of MNL models this approach is implemented as the latent class can provide further insights into the number, sizes and the characteristics of the segments [Wen et al., 2012]. The latent class MNL model exhibits the IIA (Independence of Irrelevant Alternatives) property within the segments. The modelling was done for latent class MNL models and latent class NL models. As the goodness of fit is better for the likelihood ratio test at 5% significance level [Wen et al., 2012]. In the research conducted by Waerden and Waerden, to deduce the significant attributes that affect the decision of the access and main mode choice, an ML (Mixed Logit) model is considered. The investigation of the individuals choice is carried out using a mixed logit (ML) model with panel effects. Such a model structure takes into consideration random taste variation in the population and the fact that a decision-maker can make more than one decision [Waerden and Waerden, 2018].

In the study by Li et al. the Bayesian-based model is adopted, as has higher accuracy. The advantage of this method is that it allows for complete uncertainty of the parameters through posterior distribution [Li et al., 2020]. The added advantage of this method is that it avoids over-fitting issues and numerically intensive likelihood function maximization [Li et al., 2020]. Some studies adopt the binary logistic regression model. In the study by Krygsman and Dijst the aim is to analyse the factors that impact the choice of multi-modal trips and it is analysed using the binary logistic regression model. Also, in the study by Yang et al. a similar modelling approach is implemented to determine the factors affecting the metro commuters satisfaction. In this study, the model for each predefined mode-chain, a model is estimated and the factors impacting the choice of the mode chain is analysed. Such a modelling approach estimates the variables by an iterative likelihood procedure [Yang et al., 2015]. In the study by Ton et al., station choice behaviour of cyclists is analysed wherein the MNL modelling approach is adopted to estimate the attributes that impact the station choice.

Thus, it can be observed that there are various approaches considered however MNL is the most prominent approach as it is simple and is computationally efficient. In the case of a situation that there is a correlation between the alternatives to allow more flexibility more advanced models such as the ML models are preferred. However it is computationally more intensive to implement ML models. However, the real-world validation is not possible at times so it is difficult to justify the use of more complex models [Young et al., 2018]. To sum it up, many for the various combinations of travel choice dimensions such as mode choice (in this case) the extensions or variations of the MNL and the NL (Nested Logit) model are used [Ton, 2014]. However, to account for the correlation within the trip chains for mode choice ML would be an ideal method of modelling mode choice.

3 Modelling framework

To have a more structured approach certain hypothesis are formed for the selected factors and tested individually onto a base model. the schematic diagram of the process is as shown in figure 1. The base model comprises of the travel

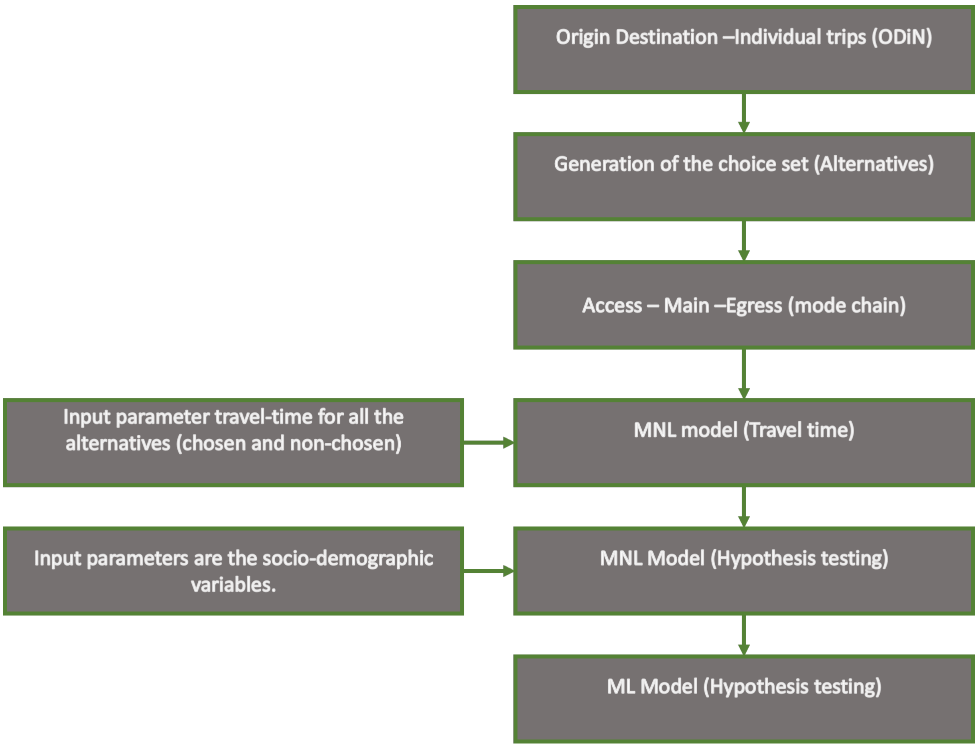


Figure 1: Flowchart of the method

time variables along with the alternative specific constants. The utility function for the base model with the travel time

parameters is as follows;
$V_{base-model} = \beta_{TA} * traveltime_{access} + \beta_{MT} * traveltime_{main-mode} + \beta_{TE} * traveltime_{egress} + ASC_{AM-EM}$
Where, β_{TA}, β_{TM}, β_{TE} are the estimated variables for access time, main mode travel time and egress travel time respectively for the chosen alternatives. ASC_{AM-EM} depicts the alternative specific constant, AM is the access mode and EM is the egress mode. The final model is a compilation of the different hypothesis. Due to the limitation of time and data, for a selected attributes hypothesis are formulated. The attributes are added to the relevant alternatives. The coefficients are estimated and checked if they are significant at 90% and 95% interval. The following equation depicts the utility function for estimating the model testing the hypothesis.

$$V_{Hypothesis-test} = V_{base-model} + \beta_{attribute} * Attribute$$

$\beta_{attribute}$ is the coefficient that is estimated for the selected attribute for testing the particular hypothesis. For the attributes that are significant a model with the combined attributes is estimated.

$$V_{Final-model(MNL)} = V_{base-model} + \beta_{attribute_{s1}} * Attribute_{s1} + \beta_{attribute_{s2}} * Attribute_{s2} + ...$$

It is optimized considering all the significant factors. The study by Wen et al. the modelling comprises of two steps where in in the initial step MNL model is estimated and then the more advanced model of Latent class model is implemented. Moreover, in the study carried out by Arentze and Molin to capture the error terms are added to correlation between the alternatives when they have same main modes and so are expected to have similar unobserved characteristics. Furthermore, once the model is optimized for the MNL model, the Final optimization is carried out by implementing the error component ML model. The ML model is implemented to as it is a more complex model which can account for the correlation between the different alternatives having the same access and egress mode.

$$V_{Final-model(ML)} = V_{Final-model(MNL)} + \epsilon_{AM} + \epsilon_{EM}$$

Where, ϵ_{AM} is the error component for access mode and ϵ_{EM} is the error component for the egress mode for the particular alternative. There are multiple approaches that can be applied using an ML model. In the current study the error-component structure is considered to be the most suitable approach. This method suggests that the error components are added to the alternatives that create the correlations among the different alternatives [Train, 2002, Arentze and Molin, 2013, Wen et al., 2012]. In case of standard modelling approach such as MNL the unobserved portion of the utility due to the IIA property and the restrictive substation patterns the correlations amongst the alternatives are not account for Arentze2013, Train2002. Hence, this approach is adopted in the study. Furthermore, the ML model is estimated using the Halton draws method. It is method that adopts a procedure and takes intelligent draws [Arentze and Molin, 2013, Train, 2002].

For a given mode the same mode is repeated for access /egress, thus, to capture the correlation across the different alternatives having the same access and egress modes, the mode across the alternatives is same it is expected that there is a correlation.

To generate the alternatives the alternatives are created using mode chains with the selected public transport mode and the for access and egress the generic modes are used such as walking, cycling, car and public transport.

4 Case Study: City of Amsterdam

The number of inhabitants in Amsterdam is about 850,000. Including the regions surrounding Amsterdam the number of inhabitants is approximately 1,350,000. The area covered $250km^2$ [Brands et al., 2020]. There are mainly there services existing in the region of Amsterdam for the public transport namely the train, bus, tram and metro. As it is a highly urbanized region there are more services available. The train network comprises 10 station within the city. The metro network has 52 station and 4 lines forming a network that is 41 km long. The bus network can be divided into the bus and night bus network consisting of 43 lines (32 regular bus / 11 night bus). The tram network has 15 lines and 500 stations forming a rail network of 80.5 km long [1].

5 Data

5.1 Data description

there predominant ways to collected data is via revealed and stated preferences. In this section the data used in literature is discussed in depth. Revealed preferences comprise the observed behaviour, such data is usually collected in the form

[1] https://amsterdammap360.com/amsterdam-tram-map: :text=Amsterdam %20Tram%20Maptext=Amsterdam%20tram%20is%20a%20transit,(80%2C5%20km

travel surveys. In the context of Netherlands a National Travel survey is carried out on a yearly basis. Revealed and stated preferences are used by various studies to deduce the mode choice behaviour. As depicted in table **??**, most studies selected are based on travel surveys such as travel diaries. NTS (National Travel Survey), OViN (Onderzoek Verplaatsingen in Nederland) and ODiN (On Der Weg in Nederland), MPN (Mobiliteitspanel Nederland) are national travel surveys carried out every year in the Netherlands. In the case of stated preference, the questionnaire is designed so as to address the perceptions of the users.
The structure of the data comprises trips made by the individual, each row depicting one trip leg made using a particular mode of transport. Each column depicts a particular variable. Hence, multiple records form a trip. Thus, there are 203 variables. They depict the record number of trips, aggregation of the different variables (Trip motive is aggregated for the overall trip as well). Urban regions such as Amsterdam, Rotterdam and The Hague area have more detailed responses. Certain questions are made addressing particular trip makers (questions associated with car ownership). The data comprises umpteen variables that address the individual characteristics, socio-demographic characteristics and travel characteristics [Shelat et al., 2018]. It can be further divided into different categories. This is depicted in the table below. Based on what is relevant to the research the variables are depicted. The variables are associated with a person, displacement, ride, and weight factor. The factors associated with a person are the socio-demographic variables, household variables, vehicles ownership etc. In case of displacement and ride, data associated with the transport mode, purpose, distance travelled in each trip etc. Weight is addressed by weighting factors. As the focus of the study is to analyze multi-modal trip to and from public transport, the data must be pre-processed to form trips that comprise public transport and the main mode and the access and egress legs. The current data-set comprises data from 2018 and 2019. However, it must be noted for the future that in 2018 the data is considered from august 2018 as it is expected to be a change in the mode choice and travel behavior due to the operation of the NZL (Noord-Zuidlijn) line. The study conducted by Brands et al., the impact of the NS line is accessed based on the analysis of the smart-card data. The findings from this research suggested that the there is a shift n the over all usage of Public transport, there is an increase of approximately 4% increase in the total ridership in terms of the journey [Brands et al., 2020]. Despite the modifications carried out on the existing bus and tram lines, there is a positive impact on the metro ridership [Brands et al., 2020].

Table 2: List of attributes based on available data

Category	**Probable Attributes (based on data available)**
Individual characteristics	Employment, age, gender, license, income, ownership of cars, availability of free student travel, frequency of travel using various modes, main mode of transport displacement
Household characteristics	Household size, number of vehicles per household
Trip characteristics	Distance, travel time, travel purpose, frequency of travel via different modes, origin and destination locations.
Level of service variables	Additional data required
Built environment	Area or region
Working condition	Working hours, car provided by the employer
Weather characteristics	Reporting month
Station choice	Origin and destination station, area, region, train stations used

5.2 Data Pre-processing

The data is extracted in such a way that the individual trips are converted into a multi-modal trip chain. Hence, the data is pre-processed to convert it from individual trips to origin–destination trips. As the data set is available in SPSS the first mode of filtration was SPSS, all the observations with trips comprising more than one trip was filtered out. Furthermore, the trips comprising of public transport as a part of the trip chain was extracted from that. To divide the trips into the different alternatives of the pre-specified mode chains. The transfer nodes are 2, so the number of trip legs are 3. Hence ,to filter out the data the multi-modal trips having 3 trip legs, observations with more than 2 trip legs are extracted. The data-set consists of a variable namely 'Ant-rit' depicting the number of trip legs carried out by the respondent. Thus, all the observations having more than 2 trip legs are filtered in SPSS using that variable.

From the filtered data the observation of each trip leg is to be formulated into a complete trip for further analysis. This process is carried out in MATLAB . The first step comprises dividing the data into socio-demographic data and trip-related data. The focus is on trip-related data to obtain the sequence of the mode-choice in the trip and the origin and destination pin code. The trip related variables comprise trip purpose, trip origin and destination pin-code, number of trips made, a mode used for the trip etc.

The data is filtered are manually to extract the different mode chains using the filter and sort function. Two models will be analyzed using the data set. One model is the train model, which comprises of train as the main mode. The main mode is fixed to train/metro and the access and egress mode are varied. Also, if walking is the part of the access or egress trip for motorized modes or cycles then walking is considered a part of single access or egress leg. Whereas, if walking is the only mode of transport to and from the station then it is considered as a separate alternative. Public transport modes if walking is part of the trip for PT access or egress trip it is considered as one mode.

5.2.1 Alternatives considered

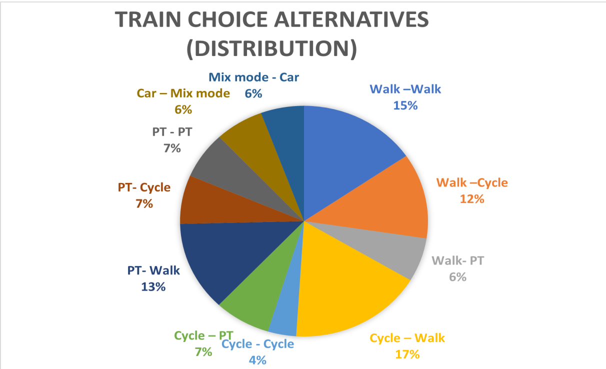
0.5 

Figure 2: Train

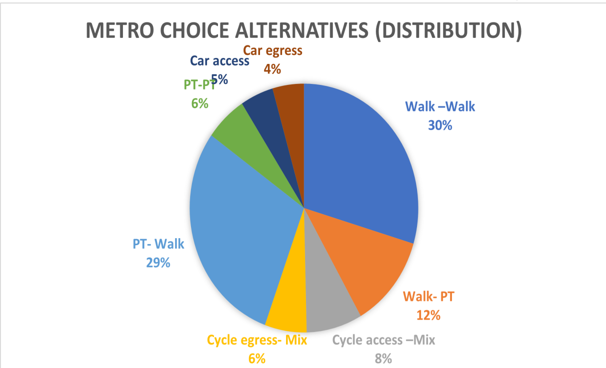
0.5 

Figure 3: Metro

Figure 4: Distribution across alternatives

As shown in the figure 2 the distribution across the alternatives suggests that the alternatives with walk as the access has the highest share. Where as the alternatives having car as the access or egress mode has the least share. The number of observations within the car alternatives is not prominent. The total number of observations for train model is 1187 and for the metro model is 405. In case of the metro as the main mode as depicted in figure 3 the alternatives with walk as access and egress has the highest share. In this case the alternatives consisting of cycle and car as the access or egress mode is lower as compared to other modes. Such a distribution is in line with expectations.

According to the study carried out by Krygsman and Dijst active modes such as walking and cycling are preferred as access modes for about 80% of all the access stages. Moreover, walking is also a connecting mode of transport. Thus, walking forms a continuous and flexible access mode that requires comparatively lesser amount of infrastructure [Fiorenzo-Catalano, 2007]. Furthermore, in the case of the egress side, asymmetric availability is expected. Walking is the more dominant mode in that case. Additionally, there are monetary costs associated with public transport services. Thus the distribution of the choices across the alternatives is plausible [Brand et al., 2017].

The distribution of the choice of the data set is also representative of the choices described in the literature as well in the Dutch context. In the study by Shelat et al. it is suggested that 82.6% of the trips are carried out 82.8% trips are carried out using the train and the rest comprise bus tram and metro as the mode of transport. As suggested in the literature the availability of cycle parking facilities is an integral part of the choice of the cycle as the preferred mode and metro stations are usually not equipped with such facilities.

5.2.2 Additional Data

The data only comprises postal pin codes of origin and destination. Thus, the centroid of the postal pin-codes extracted from GIS (Geographical Information System) is used to deduce the origin and destination nodes. In the case of the train model, the origin and destination train station are specified in the data. Hence, the access and egress travel distance and time (for the chosen and non-chosen alternatives) are extracted using the google API algorithm. The algorithm performs calculations for future events. Hence, a weekday is considered and the off-peak, hence the time fixed for deducing the travel time Ton et al. [2020]. The google API deduces the travel time (and travel distance) for the modes namely transit (Public transport), driving, walking, and cycling. The origins and destinations can be specified in the form of coordinates or addresses.

A similar approach is implemented to generate the choice sets for deducing the how further individuals are willing to cycle to access the station by Van Kampen et al.. However, in the case of the metro stations observations the stations are determined. The coordinates of the metro station locations are extracted using GIS. Based on the reported access and egress time in the data set the time the probable station location is deduced. A loop is added to the google API algorithm. The logic for all the origins and destination locations as per the chosen mode the station that has the travel time closest to the reported time is considered.

5.3 Hypothesis testing

Considering the insights from the data analysis and the re-categorized data the hypothesis is formulated. The attribute is introduced as a dummy variable except for the level of education. The level of education is added as an ordinal variable. For the car ownership the hypothesis suggesting that owning a car makes it more likely to choose the alternative that consists car for access and egress.
Hypothesis tested: The respondent is likely to use a car for access or egress if the household owns one.
For the subcategories pertaining to the age group the highest share is of the 25-34 years (refer table3). However, as suggested in multiple studies that dutch students obtain access to free public transport thus incentivising them to use public transport.
Hypothesis tested:The age-group within the range of 18-24 are highly likely to select the alternatives with public transport mode for access and egress.
This assumption is made considering that Dutch students have free public transport when pursuing education. This variable is added to the alternatives comprising public transport for access and egress.
As suggested the table 3, the higher share of individuals have the motive to travel for work or business. Furthermore, in the Dutch context professionals travelling for work purposes are expected to select active modes for access and egress Shelat2018, Ton2019. The hypothesis is as follows;]
Hypothesis tested: Respondents having their trip purpose of work are more likely to use active modes for access and egress.
The household composition has family as the highest share amongst the categories. It is expected that individuals with families would prefer to chose modes that allow for a group to travel hence alternatives with car and public transport as access and egress modes.
Hypothesis tested: Couple with others or children are more likely to opt for public transport or car as it is expected that it's easier to travel in groups.
Employment status depicts that the highest share of the individuals have a full time job (refer table 3). From table 3, it is evident that the individuals with a full time job select active modes (i.e. walking and cycling) for access and egress especially for the individuals selecting trains as the main mode as compared to the metro users. Additionally, the individuals with full time jobs are more likely to use active modes for access and egress as they have a fixed routine to follow Shelat2018. The hypothesis is formulated as follows;
Hypothesis tested: Respondents with full time jobs prefer to use active modes (walking and cycling) for access /egress.
As shown in table 3, the highest share of the individuals are highly educated. Especially, for alternatives comprising of

Variable	Categories considered	Frequency (Train)	Frequency (Metro)	Percentage (Train)	Percentage (Metro)
Car ownership	Car owned	712	238	60	59
	No car owned	475	167	40	41
Age	> 18 years	17	45	1	11
	18-24 years	262	90	22	22
	25-34 years	406	105	34	26
	35-49 years	235	80	20	20
	50-64 years	201	67	17	17
	≥ 65 years	66	18	6	4
Household composition	Single	330	91	28	22
	Couple	366	99	31	24
	Family	491	215	41	53
Trip purpose	Work/Business	674	171	57	42
	Shop/Services	58	41	5	10
	Education	149	76	13	19
	Recreation/Visit	306	117	26	29
Employment status	No pay	275	121	23	30
	Part-time (upto-12 hours)	43	30	4	7
	Part-time (12-30 hours)	124	45	10	11
	Full-time	745	209	63	52
Education level	No training	13	30	1	7
	Primary	7	11	1	3
	Lower-vocational	43	28	4	7
	Secondary-vocational	294	131	25	32
	Higher-professional	830	205	70	51
Urbanity class	Highly urban	660	241	56	60
	More urban	338	109	28	27
	Moderately urban	116	31	10	6
	Low urban	58	24	5	6
	Not urban	15	0	1	0
Gender	Male	590	183	50	45
	Female	597	222	50	55
Income	High	281	124	24	31
	Medium	434	149	37	37
	Low	472	132	40	33

Table 3: Sub categories of selected variables

cycling as the access and egress mode followed by walking. Where as, for metro users as shown in table **??** a similar trend is observed for cycling. Hence, the hypothesis posited is as follows;
Hypothesis tested: highly educated individuals are more likely to use active modes i.e., walking and cycling for access and egress.
As suggested in the data (refer table 3) higher share of the respondents live in highly urban areas. It is expected as the main mode of transport is a public transport mode. Moreover, urban areas are expected to have more access to public transport. Living in dense urban regions encourages the adoption of multi-modal trips as well [Krygsman and Dijst, 2001]. Thus, it can be expected that the individuals that lived in highly urban areas are more likely to chose public transport for access and egress. The hypothesis is as follows;
Hypothesis tested: Higher the urban density is more prone to select the alternatives with public transport access.

The gender is more or less equally distributed across the alternatives. So it can be expected that the impact of gender as high as expected. However to test it the following hypothesis is formulated;
Hypothesis tested: Men are more likely to use cars for access and egress
It can be expected that individuals earning a high income can afford more expensive modes of transport such as cars and can be expected that they are more likely to use cars for access and egress. Hence, the hypothesis is follows;
Hypothesis tested:Respondents having a higher income have more chance of using a car for access and egress.
However, it can be observed from the table 3 that the dominant classification is the middle-income individuals.

6 Results and Analysis

As depicted in the modelling process the base model is is estimated and the hypothesis is tested individually and the variables that are significant are added to the combined model and the values are estimated. The results depict the estimation of the parameters for the combined model. The results obtained from the individual hypothesis is discussed in analysis section.

6.1 Results

6.1.1 Combined model estimation

Combined model estimation (Train)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	β_{age}	0.841	0.146	5.76	0.00
2	ASC$_{caraccess}$	-1.39	0.396	-3.50	0.00
3	ASC$_{caregress}$	-1.46	0.536	-2.72	0.01
4	ASC$_{cycle-cycle}$	-2.33	0.428	-5.44	0.00
5	ASC$_{cycle-PT}$	-1.72	0.338	-5.11	0.00
6	ASC$_{cycle-walk}$	-0.0626	0.202	-0.31	0.76
7	ASC$_{PT-cycle}$	-1.89	0.388	-4.87	0.00
8	ASC$_{PT-PT}$	-1.36	0.336	-4.05	0.00
9	ASC$_{PT-walk}$	-0.544	0.190	-2.87	0.00
10	ASC$_{walk-cycle}$	-0.710	0.358	-1.98	0.05
11	ASC$_{walk-PT}$	-1.45	0.310	-4.66	0.00
12	$\beta_{car-ownership}$	0.912	0.208	4.39	0.00
13	β_{access}	-0.0211	0.00731	-2.88	0.00
14	β_{egress}	-0.0384	0.0197	-1.95	0.05
15	$\beta_{main-mode}$	-0.00620	0.00274	-2.27	0.02
16	$\beta_{mix-modeaccess}$	-0.0585	0.0139	-4.19	0.00
17	$\beta_{mix-modeegress}$	-0.0652	0.0251	-2.60	0.01
18	$\beta_{employment}$	0.575	0.187	3.08	0.00
19	$\beta_{trip-purpose}$	0.379	0.186	2.04	0.04

Summary statistics
Number of observations = 1187
Number of excluded observations = 0
Number of estimated parameters = 19

$$\begin{aligned}
\mathcal{L}(\beta_0) &= -2846.302 \\
\mathcal{L}(\hat{\beta}) &= -2539.503 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 613.598 \\
\rho^2 &= 0.108 \\
\bar{\rho}^2 &= 0.101
\end{aligned}$$

Alternative specific constant The alternative specific constants depict the preference of the population. The reference alternative is the walk-train-walk. The alternative specific constant is insignificant and has the highest value suggesting that the preference to use the cycle-train-walk is the same as walk-train-walk. The public transport- train- cycle has the highest negative value suggesting that is the least preferred alternative. It is expected that the passengers prefer to utilize other modes for the egress side of the trip over cycle due to asymmetry caused due to the availability of vehicles on the egress side of the trip. Furthermore, in the case of public transport, the level of service indicators such as punctuality, frequency of the service are influential factors as well.

As compared to the individual model wherein more parameters are significant; the socio-demographic variables significant in the final model are car ownership, age, trip purpose of work and full time employment. The value of the travel-time bets i.e. β_{access},β_{egress},$\beta_{main-mode}$ depict that all the trip legs are not not valued the same. Similar to the values in the base model the access and egress are weighed higher than the main mode travel time. The parameter for egress us weighed slightly higher than access. All the parameters are negatively correlated depicting that travel time causes dis-utility to the choice. Car ownership has highest positive weight indicating the if the individual's household owns a car there is a higher possibility to utilize it for access and egress. Age is also a significant factor in the combined model. It is line with expectations that for the youth i.e. the age-group 18-24 that consists of students are very likely to adopt public transport modes for access and egress. Hence, it depicts that the incentivization of free public transport for collage going students encourages that particular segment of the society to adopt public transport usage for access and egress. In terms of trip purpose of work and fully employed individuals are more likely to utilise active modes for access and egress to and from the train station. It is to be notes that the weight of the employment status i.e. full time employment is higher as compared to the purpose of travel indicating that full time employment is more influential adoption of active modes for access and egress.

Combined Model estimation (Metro)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC$_{car-access}$	-5.94	0.790	-7.52	0.00
2	ASC$_{car-egress}$	-5.37	0.784	-6.84	0.00
3	ASC$_{cycle-access}$	-2.58	0.381	-6.76	0.00
4	ASC$_{cycle-egress}$	-1.80	0.363	-4.98	0.00
5	ASC$_{PT-PT}$	-2.82	0.258	-10.93	0.00
6	ASC$_{PT-walk}$	-0.895	0.167	-5.34	0.00
7	ASC$_{walk-PT}$	-1.42	0.179	-7.91	0.00
8	$\beta_{car-ownership}$	1.66	0.637	2.61	0.01
9	β_{access}	-0.0680	0.0167	-4.07	0.00
10	$\beta_{car-mix-access}$	-0.0226	0.0113	-2.00	0.05
11	$\beta_{cycle-mix-access}$	-0.144	0.0320	-4.51	0.00
12	$\beta_{cycle-mix-egress}$	-0.0293	0.0172	-1.70	0.09
13	β_{egress}	-0.0293	0.00840	-3.49	0.00

Summary statistics

Number of observations = 405

Number of excluded observations = 0

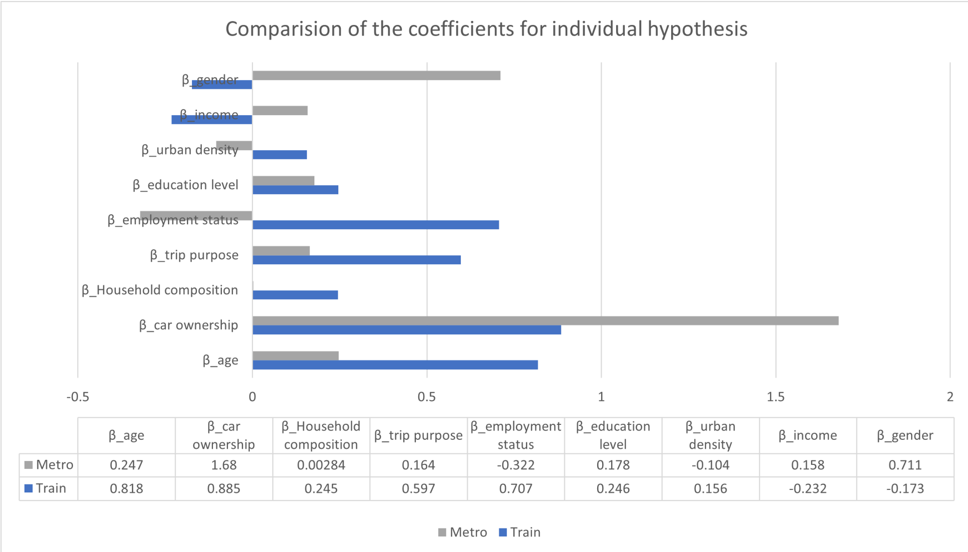
Number of estimated parameters = 13

$$\begin{aligned}
\mathcal{L}(\beta_0) &= -842.174 \\
\mathcal{L}(\hat{\beta}) &= -595.233 \\
-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] &= 493.882 \\
\rho^2 &= 0.293 \\
\bar{\rho}^2 &= 0.278
\end{aligned}$$

Alternative specific constants The reference alternative is walk-metro-walk. The ASC of the alternative with car access and mix mode has the most negative value depicting that it is the least preferred alternative as per the choices made by the population. The less negative value is of the public transport-metro–walk alternative. Hence it is the second most preferred alternative selected by the population.

The travel time parameter for the car access with mix mode egress and the main mode are not significant at 90% interval and hence the optimized version of the model does not comprise these parameters. However as the objective is to model access and egress mode choice the travel time parameter for car egress can be replaced with the generic parameter value estimated i.e. β_{egress}. In case of the metro model the car ownership is the only significant socio-demographic variable, with a very high value. The access, egress and the main mode trip legs equally. The beta estimated for the main mode is not significant (p-value greater than 0.10) and shows a positive correlation with time. Consequently on the further optimization of the model the beta for main mode is removed from the model. The egress parameter weighs lower than the access mode almost by a factor of 2. The reason behind obtaining such a value can be attributed to the fact that as the metro is only available in the city, for the individuals visiting from outside the city need to travel longer to reach the metro station location. The only significant socio-demographic variable is the car ownership. It is indicative that car is convenient to access the stations especially for the trips outside the city of Amsterdam.

6.2 Analysis



	β_age	β_car ownership	β_Household composition	β_trip purpose	β_employment status	β_education level	β_urban density	β_income	β_gender
Metro	0.247	1.68	0.00284	0.164	-0.322	0.178	-0.104	0.158	0.711
Train	0.818	0.885	0.245	0.597	0.707	0.246	0.156	-0.232	-0.173

Figure 5: Comparison of bets estimated for Hypothesis testing

The figure 5 shows the betas estimated in the individual hypothesis tested with the base model. It can be observed that the in case of the betas estimated for the income, gender, employment status and urban density the value for metro and train have an opposite sign. It depicts that males would prefer to use car to access and egress metro station but not the train. For the income parameter it can be interpreted that high income earners do not prefer to use cars to access and egress train stations where as it is vice versa for metro. The beta estimated for employment status suggests that full time employed individuals favour the use of active modes such as cycling and walking to access and egress to and from the train station. However, this is not the case with metro users. The attribute is not as significant in case of the metro model. Thus, it depicts that for both the train and metro the mode choice behaviour cannot be expected to be the same. It is attributed to various other factors such as the type of network, station facilities, attitudes towards to them etc. It is to be noted that for both the models the gender, income and urban density are not significant attributes.

Comparing the final model

Tt is observed that amongst the socio-demographic variables tested, the only significant variable is car ownership for the metro model. It is expected that the metro system is a relatively small network and is available in the city of Amsterdam and hence all respondents do not have direct access to it. Whereas, in the case of the train network, it is spread across the country and within the city. This trait is also implied via the coefficient of the travel time of the main mode in the case of the train is significant and is negative but in the case of the metro, it is positive and is insignificant. As it is a dense, high-frequency urban network it is a fast system catering to a smaller and limited region the in-vehicle time is expected to be short.

Certain aspects such as the highly educated, high-income individuals with a full-time job are more likely to adopt cycling and walking as their preferred mode of transport especially for the bike-train combination Shelat2018. This is echoed in the train model and the metro model. The common trait observed is that the car egress travel time beta is not considered to be significant. It shall be considered that the number of observations is lower for the car alternatives. However, it is in line with the literature corresponding to the Dutch context the respondents are expected to use cars for uni-modal trips rather than multi-modal trips StephanKrygsman. It can be seen that bicycle substitutes walking as the most popular access and egress mode at a lower distance for trains than for lower-level transit networks Shelat2018. It can be attributed to the fact that the availability of bike parking facilities is higher is the possibility of using cycles for access and egress. Individuals are willing to go to the fourth closest train station for access to the station if it is said to provide facilities and better connection such as lower number of transfer VanKampen2020.

6.2.1 Discussion

7 Conclusion and Recommendations

As suggested in the literature there is umpteen research in the field of modelling of access and egress mode choice. However, the simultaneous modelling of mode choice is not investigated in that depth. Furthermore, for an urban setting, the lower-level networks play an important role. Currently, the four-step model with the discrete choice modelling is adopted in western European countries. Hence it is necessary to deduce a compatible model compatible with the existing framework. Thus, discrete choice models are adopted for analysis. In the case of the Netherlands, the National and regional models are strategic models are analysed. For the larger urban regions, Amsterdam, Rotterdam and Hague area etc. have models where in the discrete choice modelling is implemented. Literature suggests various factors that affect mode choice significantly based on studies carried out in different contexts and modelling methods. The data used in this study is ODiN (on der Weg) it is a study conducted by CBS on a yearly basis. The advantage of this methodological approach is that it is designed to be flexible and applicable based on what policy implications require to be tested. Comparing the systems of the metro and train the cycle also in a way is a competing mode on an urban level for the metro stations. For the visitors within the city, it is more comfortable to use bikes. It is important to note that the dominant use of cycle is a function of the cultural context and it is not always expected to be the case for other countries. It is an adaptable framework for any city overall. The main point of difference is the contextual parameters that affect the choice set generated. Additionally, it is a more applicable system in an urban context as it is expected that the availability of public transport choices is higher in urban regions. There has been a growth in car ownership over the period time however the cycling has remained a constant mode. The socio-demographic variables impact the access and egress mode choice for trains more prominently as compared to the metro, in the case of Amsterdam. One reason might be that the metro has a higher ability to be a feeder mode for the trains rather than the main mode. For residents within the city, it can be expected that there are other choices such as cycles that provide them with a higher convenience to travel rather than make transfers. However, they are contextual parameters, and this might not be the case in other cities and other modes might be more dominant. The dominance of the car ownership variable might also suggest that having a car and travelling from out of the city is required to reach metro station. The metro is used in conjugation with walking and other public transport modes more. The are many modelling approaches that can be adopted, studies suggest that hybrid modelling processes can also be adopted or other more complex and advanced models can be used. The trade-off is of course is the computation time. However in this case a more complex ML model was considered and the results depict that there is no correlation amongst the alternatives having the same access and egress modes.

7.0.1 Limitation

The pre-specified mode chains are extracted as choice alternatives. As the literature suggests there are a lot of permutations and combinations that can be generated, the feasible mode chains are determined based on the studies in the existing context and based on the data available. The number of alternatives is limited. It is a rather complex problem to address mathematically, and a particular approach is only considered. Not all the provided alternatives need to be always available to all individuals. Considering that the Netherlands has a pro cycling approach it can be assumed that cycling is an alternative that is always available. As the data suggests that the people using other modes of transport also have access to cars.

Furthermore, the integration of the route choice is also the feedback loop that is taken into consideration. The research addresses the mode choice perspective for a transport model and not the route choice component. It is assumed that the route chosen by the individual is the optimal one. Depending on the parameter that is to be optimized for the route choice such as travel time, travel costs and distance etc. there are multiple approaches that can be taken. The data used is a travel survey which is a form of revealed preference survey. Though a lot of data is available from the data, newer ways of data collection systems such as GPS, GFTS would be able to provide more real-time information. Also, the values are reported so there might be inconsistencies in the real value and the reported value. Thus, the use of the data to make more accurate predictions would be preferred. The travel time values are deduced using the google API algorithm which is not exactly the reported value. Additionally, the travel costs are not considered in the case only travel time is the variable considered and is expected to vary amongst the alternatives. It is assumed that the optimal alternative is selected by the respondent in terms of station choice. However, there might be more optimal paths in terms of cost and time that are not known to the respondent and they end up making a particular choice.

7.1 Discussion

The framework is adaptable to the different systems of public transport. Moreover the aim is to provide a framework that can be applied in different cultural contexts as well. The points to be noted for the choice set generation considered is a similar approach to other studies having maximum two intra-modal transfers. As the trip becomes more complex than that the likelihood of making a more complex trip reduces. Comparing the travel time parameters weight of travel time access and egress is higher as compared to the main mode for both train and metro. This would suggest that more focus in the decision making process of mode selected is given to access and egress. Hence, to encourage the users to adopt public transport as their main mode the access and egress must be the emphasized upon. In terms of the socio-demographic variables , it seems to have a higher impact on the access and egress mode choice in case of the train model. It is expected to be a function of the availability of such a network in the entire country that it is more prominent. In case of the metro it is only available within the city. However, for a different city in in a different country the results might not be the same. For the current research is the starting point for modelling is the is the MNL model. The limitation of the MNL model is that it is assumed that the all alternatives are independent, which in reality is not always the case. To account for the correlations an ML model is is further analysed. The results suggests that the correlations are insignificant, thus suggesting that the alternatives are independent. It depicts that the individual takes the whole trip into consideration when choosing the mode. It is an interesting insight that the whole trip is considered as a single entity without any correlations within the alternatives having the same access and egress modes. It has to be noted that it is what the data represents in this context. When applying a similar framework to another context such as a different city it is very likely that there are correlations between the alternatives.

7.2 Further research

As deduced in the first question there are many factors in literature that affect mode choice. However, it depends on the requirement of the policy insights. Hence, in the current analysis, not all the factors are considered. In order to analyze the impact of mobility hubs additional variables such as the availability of parking spaces, accessibility to P+R facilities, Level of service of the public transport services etc can be added to the model and it can be used to analyse and test the hypothesis. Additionally, micro-mobility points are becoming increasingly popular. However, it must be considered that the alternatives such as trip purpose plays an important role. Moreover, contextual variables such as the willingness to cycle for a certain distance and time matter. In the Dutch context, cycling is en-grained as a part of the lifestyle hence it is widely accepted and adopted.
Further research can be done to analyze how the metro system or an urban system competes with cycles as the main mode of transport within the urban regions. As suggested by the results of the model, in the case of the metro model the socio-demographic variables are insignificant. Thus, the analysis suggests the car ownership is a very important factor and to encourage multi-modal trips would be something that can be ventured into further. Additionally, the research how such urban transport systems and cycles at time act as more competing modes rather than complementary modes. It can be studied further.
As this study focuses on the mode choice perspective the route choice can be further integrated with the system Points regarding the competing station choice integration. The decision making for the station choice are not integrated. In theory the individual might have better options to chose for access and egress and may be optimum in case of time and convenience. They might not be aware of all the available alternatives. Using the real-time data and integrating the route choice perspective would be a more advanced method such as heuristics that can be implemented to deduce the choice set generation.

References

Lucas Harms and Maarten Kansen. Cycling Facts Netherlands Institute for Transport Policy Analysis | KiM. *Institute for Transport Policy Analysis, KiM, Ministry of Infrastructure and Water Management,*, 2018.

Danique Ton, Dorine C. Duives, Oded Cats, Sascha Hoogendoorn-Lanser, and Serge P. Hoogendoorn. Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice*, 123 (August 2018):7–23, 2019. ISSN 09658564. doi:10.1016/j.tra.2018.08.023. URL `https://doi.org/10.1016/j.tra.2018.08.023`.

Juan de Dios Ortúzar and Luis G. Willumsen. Modelling Transport. 2011. doi:10.1002/9781119993308.

Maria Stella Fiorenzo-Catalano. *Choice Set Generation in Multi-Modal Transportation Networks*. 2007. ISBN 9789055840878.

Chieh Hua Wen, Wei Chung Wang, and Chiang Fu. Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review*, 48(2):545–554, 2012. ISSN 13665545. doi:10.1016/j.tre.2011.09.002. URL `http://dx.doi.org/10.1016/j.tre.2011.09.002`.

Stephan Krygsman and Martin Dijst. Multimodal trips in the netherlands: Microlevel individual attributes and residential context. *Transportation Research Record*, 1753(1):11–19, 2001. doi:10.3141/1753-02. URL `https://doi.org/10.3141/1753-02`.

Ghazaleh Azimi, Alireza Rahimi, Ming Lee, and Xia Jin. Mode choice behavior for access and egress connection to transit services. *International Journal of Transportation Science and Technology*, (December), 2020. ISSN 20460430. doi:10.1016/j.ijtst.2020.11.004. URL `https://doi.org/10.1016/j.ijtst.2020.11.004`.

Min Yang, Jingyao Zhao, Wei Wang, Zhiyuan Liu, and Zhibin Li. Metro commuters' satisfaction in multi-type access and egress transferring groups. *Transportation Research Part D: Transport and Environment*, 34:179–194, 2015. ISSN 13619209. doi:10.1016/j.trd.2014.11.004. URL `http://dx.doi.org/10.1016/j.trd.2014.11.004`.

Peter Van Der Waerden and Jaap Van Der Waerden. The Relation between Train Access Mode Attributes and Travelers ' Transport Mode-Choice Decisions in the Context of Medium- and Long-Distance Trips in the. 2672(8):719–730, 2018. doi:10.1177/0361198118801346.

Haoran Yang, Jianxi Feng, Martin Dijst, and Dick Ettema. Mode choice in access and egress stages of high-speed railway travelers in china. *Journal of Transport and Land Use*, 12(1):701–721, 2019. ISSN 19387849. doi:10.5198/jtlu.2019.1420.

Theo A Arentze and Eric J E Molin. Travelers ' preferences in multimodal networks : Design and results of a comprehensive series of choice experiments. *TRANSPORTATION RESEARCH PART A*, 58:15–28, 2013. ISSN 0965-8564. doi:10.1016/j.tra.2013.10.005. URL `http://dx.doi.org/10.1016/j.tra.2013.10.005`.

Sanmay Shelat, Raymond Huisman, and Niels van Oort. Analysing the trip and user characteristics of the combined bicycle and transit mode. *Research in Transportation Economics*, 69(June):68–76, 2018. ISSN 07398859. doi:10.1016/j.retrec.2018.07.017. URL `https://doi.org/10.1016/j.retrec.2018.07.017`.

Vincent Chakour and Naveen Eluru. Analyzing commuter train user behavior: A decision framework for access mode and station choice. *Transportation*, 41(1):211–228, 2014. ISSN 15729435. doi:10.1007/s11116-013-9509-y.

Jullian Van Kampen, Manoj Ashvin Jayaraj, Eric Pauwels, Rob Van Der Mei, and Elenna R. Dugundji. Understanding the relation between distance and train station choice behavior of cyclists in the western region of the Netherlands. *Procedia Computer Science*, 170:98–106, 2020. ISSN 18770509. doi:10.1016/j.procs.2020.03.013. URL `https://doi.org/10.1016/j.procs.2020.03.013`.

David A. Hensher. Stated Preference analysis of Travel Choices: The State Of Practice. *Transport Economics, Selected Readings*, pages 62–82, 2005. doi:10.4324/9780203985359-12.

Marcus Young, Simon Blainey, Marcus Young, and Simon Blainey. Railway station choice modelling : a review of methods and evidence evidence. *Transport Reviews*, 38(2):232–251, 2018. ISSN 0144-1647. doi:10.1080/01441647.2017.1326537. URL `https://doi.org/10.1080/01441647.2017.1326537`.

Xiaowei Li, Junqing Tang, Xiaojiao Hu, and Wei Wang. Assessing intercity multimodal choice behavior in a Touristy City: A factor analysis. *Journal of Transport Geography*, 86(November 2019):102776, 2020. ISSN 09666923. doi:10.1016/j.jtrangeo.2020.102776. URL `https://doi.org/10.1016/j.jtrangeo.2020.102776`.

Danique Ton, Sanmay Shelat, Sandra Nijënstein, Lotte Rijsman, Niels van Oort, and Serge Hoogendoorn. Understanding the Role of Cycling to Urban Transit Stations through a Simultaneous Access Mode and Station Choice Model. *Transportation Research Record*, 2674(8):823–835, 2020. ISSN 21694052. doi:10.1177/0361198120925076.

Danique Ton. *Dissertation Title*. PhD thesis, University of Somewhere, 2014.

Kenneth Train. Discrete Choice Methods with Simulation. 2002.

Ties Brands, Malvika Dixit, and Niels van Oort. Impact of a new metro line in amsterdam on ridership, travel times, reliability and societal costs and benefits. *European Journal of Transport and Infrastructure Research*, 20:335–353, 2020.

Judith Brand, Serge Hoogendoorn, Niels Van Oort, and Bart Schalkwijk. Modelling multimodal transit networks integration of bus networks with walking and cycling. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings*, pages 750–755, 2017. doi:10.1109/MTITS.2017.8005612.

B

Appendix - B

Train Base model iterations

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
asc_car_access	-1.70	0.294	-5.79	0.00		0.363	-4.69	0.00	
asc_car_egress	-2.19	0.333	-6.58	0.00		0.545	-4.03	0.00	
asc_ctp	-2.23	0.229	-9.74	0.00		0.489	-4.56	0.00	
asc_ctw	-0.288	0.178	-1.62	0.11	*	0.212	-1.36	0.17	*
asc_ptc	-1.86	0.234	-7.98	0.00		0.383	-4.87	0.00	
asc_ptp	-2.38	0.241	-9.87	0.00		0.479	-4.97	0.00	
asc_ptw	-0.649	0.194	-3.35	0.00		0.210	-3.08	0.00	
asc_wtc	-1.41	0.200	-7.03	0.00		0.362	-3.89	0.00	
asc_wtp	-1.71	0.221	-7.75	0.00		0.461	-3.71	0.00	
traveltime_PT_access	0.000546	0.00603	0.09	0.93	*	0.00553	0.10	0.92	*
traveltime_PT_egress	-0.00921	0.00707	-1.30	0.19	*	0.0149	-0.62	0.54	*
traveltime_car_access	0.0207	0.0149	1.39	0.16	*	0.0134	1.54	0.12	*
traveltime_car_egress	0.0232	0.0172	1.35	0.18	*	0.0280	0.83	0.41	*
traveltime_cycle_access	-0.00121	0.00286	-0.42	0.67	*	0.00293	-0.41	0.68	*
traveltime_cycle_egress	-0.0484	0.0111	-4.36	0.00		0.0286	-1.69	0.09	*
traveltime_main	-0.00651	0.00299	-2.18	0.03		0.00268	-2.43	0.02	
traveltime_mixmode_access	-0.0459	0.0148	-3.10	0.00		0.0127	-3.61	0.00	
traveltime_mixmode_egress	-0.0604	0.0140	-4.30	0.00		0.0272	-2.22	0.03	
traveltime_walk_access	-0.0224	0.00326	-6.86	0.00		0.00578	-3.87	0.00	
traveltime_walk_egress	-0.0393	0.00417	-9.42	0.00		0.0227	-1.73	0.08	*

Figure 1 Access and egress mode specific parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
asc_car_access	-1.74	0.205	-8.51	0.00		0.311	-5.61	0.00	
asc_car_egress	-1.99	0.214	-9.29	0.00		0.479	-4.15	0.00	
asc_ctc	-2.34	0.225	-10.40	0.00		0.420	-5.56	0.00	
asc_ctp	-1.55	0.195	-7.96	0.00		0.328	-4.71	0.00	
asc_ctw	-0.0812	0.168	-0.48	0.63	*	0.200	-0.41	0.68	*
asc_ptc	-1.70	0.199	-8.56	0.00		0.377	-4.52	0.00	
asc_ptp	-1.64	0.196	-8.34	0.00		0.307	-5.34	0.00	
asc_ptw	-0.381	0.172	-2.22	0.03		0.185	-2.06	0.04	
asc_wtc	-0.710	0.180	-3.95	0.00		0.352	-2.02	0.04	
asc_wtp	-1.27	0.194	-6.55	0.00		0.303	-4.19	0.00	
traveltime_access	-0.0216	0.00296	-7.32	0.00		0.00730	-2.97	0.00	
traveltime_egress	-0.0381	0.00388	-9.80	0.00		0.0194	-1.97	0.05	
traveltime_main	-0.00605	0.00295	-2.05	0.04		0.00271	-2.23	0.03	

Figure 2 Generic parameters

C

Appendix - C

Metro Base model iterations

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
asc_car_access	-4.57	0.634	-7.20	0.00		0.715	-6.39	0.00	
asc_car_egress	-6.75	0.642	-10.51	0.00		0.810	-8.33	0.00	
asc_cycle_access	-4.23	0.469	-9.02	0.00		0.695	-6.09	0.00	
asc_cycle_egress	-2.12	0.559	-3.79	0.00		0.522	-4.06	0.00	
asc_pmp	-4.68	0.434	-10.78	0.00		0.498	-9.40	0.00	
asc_pmw	-2.26	0.361	-6.24	0.00		0.345	-6.53	0.00	
asc_wmp	-2.10	0.336	-6.23	0.00		0.436	-4.80	0.00	
traveltime_PT_access	-0.0209	0.0129	-1.63	0.10	*	0.0174	-1.20	0.23	*
traveltime_PT_egress	-0.0140	0.0125	-1.12	0.26	*	0.0129	-1.08	0.28	*
traveltime_car_access	0.0147	0.0310	0.47	0.64	*	0.0470	0.31	0.76	*
traveltime_car_egress	0.0640	0.0313	2.05	0.04		0.0319	2.01	0.04	
traveltime_car_mix_access	0.000955	0.00681	0.14	0.89	*	0.0114	0.08	0.93	*
traveltime_car_mix_egress	-0.0997	0.0463	-2.15	0.03		0.0588	-1.69	0.09	*
traveltime_cycle_access	-0.0148	0.0104	-1.41	0.16	*	0.0178	-0.83	0.41	*
traveltime_cycle_egress	-0.0724	0.0253	-2.86	0.00		0.0212	-3.41	0.00	
traveltime_cycle_mix_access	-0.169	0.0468	-3.61	0.00		0.0382	-4.42	0.00	
traveltime_cycle_mix_egress	-0.0263	0.0130	-2.02	0.04		0.0239	-1.10	0.27	*
traveltime_mainmode	0.00207	0.0129	0.16	0.87	*	0.0122	0.17	0.87	*
traveltime_walk_access	-0.118	0.0166	-7.08	0.00		0.0166	-7.07	0.00	
traveltime_walk_egress	-0.0446	0.00835	-5.35	0.00		0.0226	-1.98	0.05	

Figure 1 Access and egress mode specific parameters

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
asc_car_access	-3.65	0.369	-9.88	0.00		0.488	-7.48	0.00	
asc_car_egress	-3.68	0.383	-9.59	0.00		0.480	-7.66	0.00	
asc_cycle_access	-2.53	0.288	-8.77	0.00		0.331	-7.62	0.00	
asc_cycle_egress	-2.77	0.305	-9.08	0.00		0.322	-8.61	0.00	
asc_pmp	-2.90	0.306	-9.49	0.00		0.314	-9.25	0.00	
asc_pmw	-0.852	0.242	-3.52	0.00		0.243	-3.51	0.00	
asc_wmp	-1.63	0.269	-6.08	0.00		0.255	-6.39	0.00	
traveltime_access	-0.0488	0.00728	-6.71	0.00		0.0176	-2.77	0.01	
traveltime_egress	-0.0423	0.00684	-6.19	0.00		0.0151	-2.81	0.00	
traveltime_mainmode	0.00481	0.0121	0.40	0.69	*	0.0109	0.44	0.66	*

Figure 2 Generic parameters

D

Appendix - D

Mixed logit model iterations

TRAIN

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_egress_std	-0.0241	0.429	-0.06	0.96	*	0.0506	-0.48	0.63	*
age_18_24	0.845	0.145	5.84	0.00		0.145	5.83	0.00	
asc_car_access	-1.27	0.321	-3.94	0.00		0.364	-3.48	0.00	
asc_car_egress	-1.47	0.341	-4.31	0.00		0.530	-2.77	0.01	
asc_ctp	-1.59	0.195	-8.14	0.00		0.317	-5.02	0.00	
asc_ctw	0.0755	0.165	0.46	0.65	*	0.170	0.44	0.66	*
asc_ptc	-1.77	0.206	-8.62	0.00		0.382	-4.64	0.00	
asc_ptp	-1.31	0.214	-6.15	0.00		0.324	-4.05	0.00	
asc_ptw	-0.427	0.173	-2.47	0.01		0.173	-2.47	0.01	
asc_wtc	-1.31	0.181	-7.22	0.00		0.365	-3.58	0.00	
asc_wtp	-1.46	0.197	-7.44	0.00		0.311	-4.70	0.00	
car_ownership_1	0.889	0.210	4.24	0.00		0.209	4.26	0.00	
cycle_egress_std	0.111	0.692	0.16	0.87	*	0.172	0.64	0.52	*
traveltime_access	-0.0130	0.00225	-5.79	0.00		0.00380	-3.43	0.00	
traveltime_egress	-0.0383	0.00392	-9.79	0.00		0.0196	-1.95	0.05	*
traveltime_main	-0.00589	0.00294	-2.00	0.05		0.00273	-2.16	0.03	
traveltime_mixmode_access	-0.0485	0.0141	-3.44	0.00		0.0118	-4.10	0.00	
traveltime_mixmode_egress	-0.0636	0.0133	-4.77	0.00		0.0247	-2.58	0.01	
trip_employment_fulltime	0.774	0.155	4.99	0.00		0.154	5.03	0.00	
walk_egress_std	-0.0112	0.211	-0.05	0.96	*	0.0150	-0.75	0.45	*

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_std	1.14	0.940	1.21	0.23	*	1.47	0.77	0.44	*
age_18_24	1.07	0.359	2.97	0.00		0.522	2.04	0.04	
asc_car_access	-1.38	0.456	-3.03	0.00		0.512	-2.70	0.01	
asc_car_egress	-1.58	0.446	-3.54	0.00		0.642	-2.46	0.01	
asc_ctp	-1.79	0.324	-5.53	0.00		0.493	-3.64	0.00	
asc_ctw	0.0418	0.178	0.23	0.81	*	0.196	0.21	0.83	*
asc_ptc	-1.98	0.326	-6.07	0.00		0.538	-3.68	0.00	
asc_ptp	-1.50	0.320	-4.70	0.00		0.477	-3.15	0.00	
asc_ptw	-0.610	0.296	-2.06	0.04		0.390	-1.56	0.12	*
asc_wtc	-1.33	0.187	-7.12	0.00		0.376	-3.54	0.00	
asc_wtp	-1.65	0.310	-5.32	0.00		0.482	-3.41	0.00	
car_ownership_1	0.941	0.237	3.97	0.00		0.253	3.72	0.00	
car_std	-0.366	1.09	-0.34	0.74	*	1.18	-0.31	0.76	*
cycle_std	-0.561	0.743	-0.75	0.45	*	0.814	-0.69	0.49	*
traveltime_access	-0.0136	0.00238	-5.72	0.00		0.00381	-3.58	0.00	
traveltime_egress	-0.0386	0.00400	-9.64	0.00		0.0204	-1.89	0.06	*
traveltime_main	-0.00548	0.00315	-1.74	0.08	*	0.00308	-1.78	0.08	*
traveltime_mixmode_access	-0.0487	0.0149	-3.28	0.00		0.0134	-3.64	0.00	
traveltime_mixmode_egress	-0.0643	0.0138	-4.66	0.00		0.0259	-2.48	0.01	
trip_employment_fulltime	0.801	0.172	4.65	0.00		0.196	4.09	0.00	
walk_std	0.0346	0.181	0.19	0.85	*	0.0483	0.72	0.47	*

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_access_std	0.110	0.488	0.23	0.82	*	0.285	0.39	0.70	*
PT_egress_std	-0.0775	0.322	-0.24	0.81	*	0.198	-0.39	0.70	*
age_18_24	0.895	0.167	5.37	0.00		0.165	5.42	0.00	
asc_car_access	1.40	0.462	3.03	0.00		0.481	2.91	0.00	
asc_car_egress	2.02	0.493	4.09	0.00		0.613	3.29	0.00	
asc_ctp	-0.0892	0.280	-0.32	0.75	*	0.353	-0.25	0.80	*
asc_ctw	1.53	0.262	5.85	0.00		0.255	6.00	0.00	
asc_ptc	1.02	0.362	2.82	0.00		0.475	2.15	0.03	
asc_ptp	1.55	0.373	4.15	0.00		0.428	3.61	0.00	
asc_ptw	2.40	0.348	6.89	0.00		0.338	7.10	0.00	
asc_wtc	-0.608	0.214	-2.85	0.00		0.369	-1.65	0.10	*
asc_wtp	-1.35	0.232	-5.81	0.00		0.333	-4.05	0.00	
car_ownership_1	0.960	0.237	4.05	0.00		0.231	4.15	0.00	
cycle_access_std	-6.14	0.851	-7.22	0.00		1.01	-6.09	0.00	
cycle_egress_std	0.164	0.357	0.46	0.65	*	0.229	0.71	0.48	*
traveltime_access	-0.0396	0.00622	-6.37	0.00		0.00823	-4.82	0.00	
traveltime_egress	-0.0376	0.00403	-9.34	0.00		0.0207	-1.81	0.07	*
traveltime_main	-0.00903	0.00422	-2.14	0.03		0.00419	-2.16	0.03	
traveltime_mixmode_access	-0.109	0.0181	-5.99	0.00		0.0180	-6.05	0.00	
traveltime_mixmode_egress	-0.0701	0.0143	-4.91	0.00		0.0249	-2.82	0.00	
trip_employment_fulltime	0.633	0.230	2.75	0.01		0.233	2.72	0.01	
trip_purpose_work	0.350	0.227	1.54	0.12	*	0.229	1.53	0.13	*
walk_access_std	10.0	3.37e-008	296691287.14	0.00		1.80e+308	0.00	1.00	*
walk_egress_std	-0.0419	0.206	-0.20	0.84	*	0.0825	-0.51	0.61	*

METRO

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_access_std	0.0474	0.546	0.09	0.93	*	0.177	0.27	0.79	*
asc_car_access	-13.1	2.70	-4.84	0.00		2.70	-4.84	0.00	
asc_car_egress	-15.7	2.88	-5.44	0.00		2.86	-5.47	0.00	
asc_cycle_access	-12.7	1.12	-11.33	0.00		1.20	-10.55	0.00	
asc_cycle_egress	-11.6	1.23	-9.38	0.00		1.22	-9.45	0.00	
asc_pmp	-3.74	0.737	-5.07	0.00		0.707	-5.29	0.00	
asc_pmw	-1.26	0.291	-4.32	0.00		0.277	-4.54	0.00	
asc_wmp	-1.65	0.315	-5.24	0.00		0.338	-4.88	0.00	
car_access_std	5.97	1.49	4.01	0.00		1.39	4.28	0.00	
cycle_access_std	-10.0	4.85e-008	-206031775.40	0.00		1.80e+308	0.00	1.00	*
traveltime_access	-0.150	0.0240	-6.26	0.00		0.0247	-6.08	0.00	
traveltime_car_access	0.0376	0.0589	0.64	0.52	*	0.0749	0.50	0.62	*
traveltime_car_egress	0.167	0.0754	2.22	0.03		0.0814	2.06	0.04	
traveltime_car_mix_access	0.00790	0.0109	0.72	0.47	*	0.0137	0.58	0.56	*
traveltime_car_mix_egress	-0.0596	0.0358	-1.67	0.10	*	0.0237	-2.52	0.01	
traveltime_cycle_access	-0.0866	0.0215	-4.02	0.00		0.0212	-4.09	0.00	
traveltime_cycle_egress	-0.108	0.0545	-1.98	0.05		0.0529	-2.04	0.04	
traveltime_cycle_mix_access	-0.176	0.0588	-2.99	0.00		0.0452	-3.89	0.00	
traveltime_cycle_mix_egress	-0.0700	0.0389	-1.80	0.07	*	0.0443	-1.58	0.11	*
traveltime_egress	-0.0509	0.0108	-4.73	0.00		0.0325	-1.57	0.12	*
traveltime_mainmode	-0.00974	0.0147	-0.66	0.51	*	0.0143	-0.68	0.50	*
walk_access_std	1.08	1.01	1.07	0.28	*	1.00	1.08	0.28	*

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_egress_std	-0.00810	0.503	-0.02	0.99	*	0.0317	-0.26	0.80	*
asc_car_access	-4.91	0.831	-5.91	0.00		0.910	-5.39	0.00	
asc_car_egress	-6.39	0.758	-8.44	0.00		0.911	-7.02	0.00	
asc_cycle_access	-3.13	0.437	-7.17	0.00		0.507	-6.18	0.00	
asc_cycle_egress	-2.43	0.535	-4.54	0.00		0.458	-5.31	0.00	
asc_pmp	-3.41	0.410	-8.32	0.00		0.390	-8.74	0.00	
asc_pmw	-1.33	0.363	-3.65	0.00		0.339	-3.91	0.00	
asc_wmp	-2.10	0.383	-5.47	0.00		0.356	-5.89	0.00	
car_ownership_metro	1.68	0.629	2.67	0.01		0.641	2.62	0.01	
traveltime_access	-0.0627	0.00911	-6.89	0.00		0.0151	-4.14	0.00	
traveltime_car_mix_access	-0.0182	0.00616	-2.96	0.00		0.0114	-1.59	0.11	*
traveltime_car_mix_egress	-0.120	0.0492	-2.44	0.01		0.0698	-1.72	0.09	*
traveltime_cycle_mix_access	-0.142	0.0401	-3.55	0.00		0.0324	-4.38	0.00	
traveltime_cycle_mix_egress	-0.0438	0.0122	-3.59	0.00		0.0231	-1.90	0.06	*
traveltime_egress	-0.0459	0.00746	-6.16	0.00		0.0174	-2.64	0.01	
traveltime_mainmode	0.135	0.0932	1.45	0.15	*	0.0857	1.57	0.12	*
walk_egress_std	-0.0672	0.529	-0.13	0.90	*	0.0630	-1.07	0.29	*

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_std	-0.0838	0.473	-0.18	0.86	*	0.189	-0.44	0.66	*
asc_car_access	-15.0	4.74	-3.16	0.00		6.50	-2.31	0.02	
asc_car_egress	-17.1	4.84	-3.52	0.00		6.66	-2.56	0.01	
asc_cycle_access	-12.6	0.868	-14.53	0.00		0.883	-14.30	0.00	
asc_cycle_egress	-11.5	1.01	-11.40	0.00		0.970	-11.90	0.00	
asc_pmp	-3.92	0.498	-7.87	0.00		0.461	-8.51	0.00	
asc_pmw	-1.75	0.419	-4.17	0.00		0.388	-4.51	0.00	
asc_wmp	-2.28	0.438	-5.20	0.00		0.452	-5.04	0.00	
car_ownership_metro	5.93	2.76	2.15	0.03		3.81	1.55	0.12	*
car_std	6.04	2.16	2.80	0.01		3.04	1.99	0.05	
cycle_std	-10.0	1.80e+308	0.00	1.00	*	1.80e+308	0.00	1.00	*
traveltime_access	-0.124	0.0155	-7.97	0.00		0.0159	-7.77	0.00	
traveltime_car_mix_access	-0.0205	0.00656	-3.13	0.00		0.00962	-2.14	0.03	
traveltime_car_mix_egress	-0.0953	0.0420	-2.27	0.02		0.0345	-2.76	0.01	
traveltime_cycle_mix_access	-0.196	0.0568	-3.44	0.00		0.0428	-4.57	0.00	
traveltime_cycle_mix_egress	-0.0349	0.0136	-2.56	0.01		0.0234	-1.49	0.14	*
traveltime_egress	-0.0536	0.0107	-5.01	0.00		0.0323	-1.66	0.10	*
traveltime_mainmode	0.130	0.105	1.23	0.22	*	0.0961	1.35	0.18	*
walk_std	0.213	1.49	0.14	0.89	*	1.68	0.13	0.90	*

Name	Value	Std err	t-test	p-value		Robust Std err	Robust t-test	p-value	
PT_access_std	0.00320	0.314	0.01	0.99	*	0.0139	0.23	0.82	*
asc_car_access	-4.91	0.831	-5.91	0.00		0.910	-5.39	0.00	
asc_car_egress	-6.39	0.757	-8.44	0.00		0.911	-7.02	0.00	
asc_cycle_access	-3.13	0.437	-7.17	0.00		0.507	-6.18	0.00	
asc_cycle_egress	-2.43	0.535	-4.54	0.00		0.457	-5.31	0.00	
asc_pmp	-3.41	0.410	-8.32	0.00		0.390	-8.74	0.00	
asc_pmw	-1.33	0.363	-3.65	0.00		0.339	-3.91	0.00	
asc_wmp	-2.10	0.383	-5.47	0.00		0.356	-5.89	0.00	
car_ownership_metro	1.68	0.629	2.67	0.01		0.641	2.62	0.01	
traveltime_access	-0.0627	0.00910	-6.89	0.00		0.0151	-4.14	0.00	
traveltime_car_mix_access	-0.0182	0.00616	-2.96	0.00		0.0114	-1.59	0.11	*
traveltime_car_mix_egress	-0.120	0.0492	-2.44	0.01		0.0698	-1.72	0.09	*
traveltime_cycle_mix_access	-0.142	0.0401	-3.55	0.00		0.0324	-4.38	0.00	
traveltime_cycle_mix_egress	-0.0438	0.0122	-3.59	0.00		0.0231	-1.90	0.06	*
traveltime_egress	-0.0459	0.00746	-6.16	0.00		0.0174	-2.64	0.01	
traveltime_mainmode	0.135	0.0932	1.45	0.15	*	0.0857	1.57	0.12	*
walk_access_std	-0.0374	0.398	-0.09	0.93	*	0.0281	-1.33	0.18	*