

Modelling mode choice in Dutch urban areas with a focus on the (e-)bike

Master's Thesis

C.H. Huurman

June 27, 2022



MODELLING MODE CHOICE IN DUTCH URBAN AREAS WITH A FOCUS ON THE (E-)BIKE

MASTER'S THESIS

by

C.H. Huurman

Faculty of Civil Engineering and Geosciences

At Delft University of Technology
Stevinweg 1, 2628 CN Delft

and

At Witteveen+Bos
Stationsweg 5B, 4811 AX Breda

June 27, 2022

Supervisors TU Delft:
Dr.ir. W. (Winnie) Daamen
Dr.ir. A.J. (Adam) Pel
Dr. C. (Kees) Maat

Supervisor Witteveen+Bos:
Ir. S. (Sander) Veenstra

Cover photo taken by:
City of Amsterdam

PREFACE

This thesis is the final part of obtaining the master's degree in Civil Engineering, track Transport & Planning. Since I was a kid, I have loved to travel. I often travelled with my family to the South of Europe by car. Looking outside the window, I saw us going through beautiful tunnels in the hills and driving over magnificent bridges across waterways. This is where my admiration for infrastructure began. After choosing the minor Transport, Infrastructure, and Logistics during my Bachelor's, I knew that that was the field of study I wanted to continue in. Now that I am almost graduated, I love that it is possible to physically see the things that I have learned. The initial interest in infrastructure has grown to a fascination for all types of traffic networks. This master's degree has confirmed that this is the career path I want to follow to be able to contribute to the ever-changing traffic and mobility needs.

The master's degree does not have a long-term internship in their programme, so I knew I wanted to graduate at a company to gain that experience. Witteveen+Bos gave me the opportunity to graduate with a self-chosen topic, which had to be related to improving their bicycle network tool called the Fietsmonitor. Besides working on my thesis, I also had the possibility to contribute to some projects to also get an idea of how they work. It has been a pleasure and a valuable experience to work at this engineering firm.

It has been a struggle to work in a time that was influenced by a global pandemic. The start of my thesis in November began with a nationwide lockdown, wherein working from home was mandatory. Gradually the measures were reduced and I was able to work from home and at the office in a hybrid manner. Being able to physically meet with people and work at the office helped me to remain focused and motivated. I am also very grateful that it is possible to finish my master's degree by doing my final presentation in front of my friends and family at TU Delft.

I want to thank my committee members for their guidance, ideas, and feedback that helped me improve my master's thesis. Winnie, thank you for your help and knowledge in data collection and for discussing my progress every other week. Adam & Kees, thank you for your knowledge and interesting discussions about discrete choice models. Sander, thank you for your feedback and practical point of view in the weekly meetings, and for always being available to answer my questions. The support and guidance in the use of Python and QGIS have helped me a lot. I also want to thank my friends and family for the support and encouragement they have given me. Not only during the master's thesis, but during my entire study period. A special thanks goes to my dad, who has brought me where I am today by guiding me towards the right study choices for me. I am proud to finish my master's thesis and hereby present the result.

Enjoy reading!

*Chantal Huurman
Barendrecht, June 2022*

ABSTRACT

Cycling is an important part of Dutch culture. It is the second most used mode of transport in the Netherlands, closely following the use of the car. The most popular places to use the bike are within the cities, and because these urban areas are growing rapidly, the bike is also used more and more often (KiM, 2018). In addition, the use of the e-bike is growing rapidly. The e-bike is no longer a transport mode solely used by the elderly for recreational purposes, but its use by young and middle-aged people for commuting is growing. The advantage is that the e-bike could more easily be a substitution mode for the car, and thus increase the use of more sustainable modes in urban areas.

Because of the cycling culture, the bike is a very important mode in the Netherlands to consider in research about mode choice. However, the bike is in practice often modelled solely based on the travel distance or travel time. Moreover, the increasing use of the e-bike and its differences with the bike show that more research into this transport mode is needed. The most prominent differences are the higher speed and the less physical effort that is needed for the e-bike. Based on these research gaps, the objective of this research is two-fold. One, its goal is to find a mode choice model that predicts a more accurate modal split with the bike, e-bike, car, and public transport that includes multiple factors for urban areas in the Netherlands. And two, its goal is to find significant factors for the mode choice that influence the modal split, in order to support municipalities of Dutch urban areas to stimulate the use of the (e-)bike. The research question belonging to the objectives is:

What mode choice model estimates an accurate modal split for Dutch urban areas including multiple influencing factors, which is also useful for analysing changes to (e-)bicycle policies and networks?

The research steps needed to answer this question are a literature review, data analysis, modelling, and validation. The goal of the literature review is to find factors that are known to influence mode choice. Linked to the two goals of this research, factors that are modelled are chosen based on two things:

- If it has the possibility to increase the accuracy of predicting a modal split for urban areas, and/or
- If it has the possibility to give useful information for stimulating (e-)bicycle use for municipalities.

With the results of the literature review, the data is gathered and analysed to get insights into the behaviour of the factors and to form expectations of the factors for the modelling step. The data is then used to estimate a mode choice model, in which significant factors are kept in the final model. Interaction effects and quadratic components are tested and also added to the model. The final model is then internally and externally validated, and the application of the final model is presented.

Based on the literature review, Table 1 shows the factors that are shown to be significant in literature and that can be modelled with data. The data needed for these factors is from ODIN 2018 and ODIN 2019. It is combined with data from OSM, NDOV, RDW, CROW, and CBS.

Table 1: Significant factors based on literature

Spatial characteristics	Characteristics of people	Travel journey characteristics
Separate bicycle lanes	Population density	Travel distance
Access to public transport	Gender	Travel time
Access to bicycle parking	Age	Departure day
Frequency of public transport	Occupation	Departure time
Density of network	Education	
Connectivity of network	Driver's license	
Car parking zones	Wealth	
Season	No. of household members	
	No. of cars per driving person	
	E-bike availability	
	Habit	

The data is filtered on the modes chosen to model (bike, e-bike, car, train, and BTM), urban areas (address density > 1000 addresses per km²), and a maximum travel distance of 15 km. This is based on CDF-curves of the travel distance of all transport modes. It shows that below this boundary, the bike and e-bike are competitive modes with the car and public transport. In the table, a distinction is made between spatial characteristics, people's characteristics, and transport journey characteristics. The factors in the columns are in random order.

Modelling the variables individually shows that the habit of people has a large influence on mode choice. It explains the data very well, although its high explanatory power could also be related to the way the ODIN data is collected. The habit is based on the frequency that a mode is taken. Picking a random day to report the trips a respondent made, can most likely also be the day that the frequently used transport mode is chosen. Based on this limitation, and because it is generally not known what the habit is of people, the habit is not included in the final model. Other variables with a large fit with the data are the travel distance and travel time, the density of the network, the frequency of the public transport, the car parking zones, and the address density. It shows that in general the people's characteristics individually have less explanatory power. However, the models of each category of variables show that the people's characteristics explain the data the best with a rho-square-bar of 0.576. This is mostly caused by the variable habit, which individually has a rho-square-bar of 0.552. The spatial characteristics and the travel characteristics have a rho-square-bar of respectively 0.403 and 0.391. These results give a first impression of the behaviour of the factors.

The calculated travel distance and travel time are included in the modelling step to replace the travel distance and travel time as specified by respondents in ODIN. This is done to also include the travel distance and travel time of the non-chosen mode alternatives for each trip to have the possibility to model the trade-off between the modalities for these two factors. By plotting box plots that present the difference between the calculated and specified factors, it shows that the calculated value is often close to the given value. However, it also shows that the travel distance is more accurately calculated than the travel time.

Some factors are excluded from the model during the modelling step based on several reasons. The travel time, address density, and e-bike availability are excluded from the model based on correlations with other factors. Moreover, factors that seemed relevant in the literature review, but are not in the model of this research are the separate bicycle lanes, bicycle parking, and street connectivity. These factors do not show the expected results and therefore are not expected to accurately describe their intended influence. Finally, factors influencing the daily/yearly choice and factors from which the input is not generally known are not useful for the application of the final model. If the input is not known or if the input changes often, the model cannot be used by municipalities to influence the use of the (e-)bike. These factors are the departure day, departure time, season, and habit.

After excluding factors from the model, interaction effects and quadratic components are tested and the significant results are included in the model. Three types of logit models are tested, and the final model is based on the results of the nested logit model in Table 5.12. This model uses a nested structure, in which the bike and e-bike are within the same nest. The nested structure shows that if there is a higher preference or dislike for the bike, the e-bike will be more preferred or disliked as well. This is probably because of their similar mode characteristics. The included factors, interaction effects, and quadratic components of the final model can be found in Table 2. The sensitivity analysis and uncertainty analysis show that the model is robust, stable, and not very sensitive to changes in the parameter estimates. The most influential variable is as expected the travel distance. Second-order effects are not found for both the sensitivity and the uncertainty analysis.

Table 2: Final modelled factors

Spatial characteristics	Characteristics of people	Travel journey characteristics
Access to public transport	Gender	Travel distance
Frequency of public transport	Age	Travel distance ²
Density of network	Occupation	Travel speed
Car parking zones	Education	Travel distance & age
Car parking zones ²	Driver's license	Travel distance & gender
Car parking zones & residential zip code	Wealth	
	No. of household members	
	No. of cars per driving person	
	No. of cars per driving person ²	

The results of the final nested logit model show that there is a general preference for the e-bike over the car if other factors remain the same. The bike is also often preferred if there are no changes in other factors, but less than the car. The public transport modes both have a negative ASC and are thus disliked. The expectations of the variables compared to their found result can be found in Table 3. The results show that most variables had their expected outcome. The access to public transport positively influences the choices of train and BTM, just like the frequency of the BTM mode. This variable is also positive for the bike & e-bike, which is unexpected. It can be related to the fact that when the frequency of BTM increases, it is most likely an area with a higher address density where cycling is more often chosen. Anyway, the impact of this variable on the modal split outcome is not large for the bike & e-bike. The age, gender, education, occupation, household members, and wealth are all significant as expected. Their influence on the outcome is not large, but they still have an influence. Car ownership resulted in an effect that was not expected. It was expected that it would negatively influence the choice of the modes besides the car, but it is a quadratic relation wherein few cars available lead to a preferred choice for the other modes and many cars available lead to a disliked choice for the modes. The expectation is therefore partly true. The variables car parking and driver's license have the results as expected, and they can also influence the outcome a lot. The street density is as expected for the BTM and the (e-)bike, but it was not significant for the car and the train. Lastly, the travel distance is as expected estimated to be negative, with a quadratic component for the bike, e-bike, and BTM. It was expected that the higher the travel distance, the more resistance there would be for the bike and e-bike. However, these modes have a small positive quadratic component. This still leads to a larger resistance for a higher travel distance, but it was expected to have a larger effect.

Table 3: Expectation vs results: Final model

	Expectation	Outcome
Access to public transport	PT: +	PT: +
Age	All: ✓	All: ✓
Car ownership	Car: +	Car: ref
	PT & (E-)Bike: -	PT & (E-)Bike: + / -
Travel distance	All: -	All: -
Car parking	Car: -	Car: ref
	PT & (E-)Bike: +	PT & (E-)Bike: +
Education	All: ✓	All: ✓
Frequency of public transport	PT: +	PT & (E-)Bike: +
	Car & (E-)Bike: -	Car: ref
Gender	All: ✓	All: ✓
Driver's License	Car: +	Car: ref
	PT & (E-)Bike: -	PT & (E-)Bike: -
Household Members	All: ✓	All: ✓
Occupation	All: ✓	All: ✓
Street Density	All: +	BTM & (E-)Bike: +
Wealth	All: ✓	All: ✓

With the final model, the internal and external validation can be performed. The internal validation shows that 66% is estimated correctly and that the outcome of the estimated modal split of the probabilities gives the best result compared to the actual modal split of the data. The external validation shows that 62% is estimated correctly, and is thus close to the prediction power of the internal validation. However, these numbers are based on individual predictions, which is not the goal of the performance of the model. Its goal is to accurately estimate the modal split. It is assumed that a modal split is estimated accurately if the shares of the modes are within 5 percentage points of the actual share. In light of this goal, it can be concluded from the external validation that the model can be used in other cities that have a similar preference for the modes as the average in Dutch urban areas. For cities that have a large difference from this preference, it will not predict the modal split accurately enough. Figure 1 shows that the estimated modal split of Haarlem in Figure 1b is not accurate enough compared to the actual modal split in Figure 1a for municipalities to work with. Changing the alternative specific constants can be an option, but municipalities then should perform a preliminary analysis on the preference of the bike over the car in the city. In this research, the ASC of the bike is changed by trial-and-error from 0.76 to 1.40 to match the actual modal split the most. In Figure 1c, the model results can be seen with changes to the ASC of the bike.

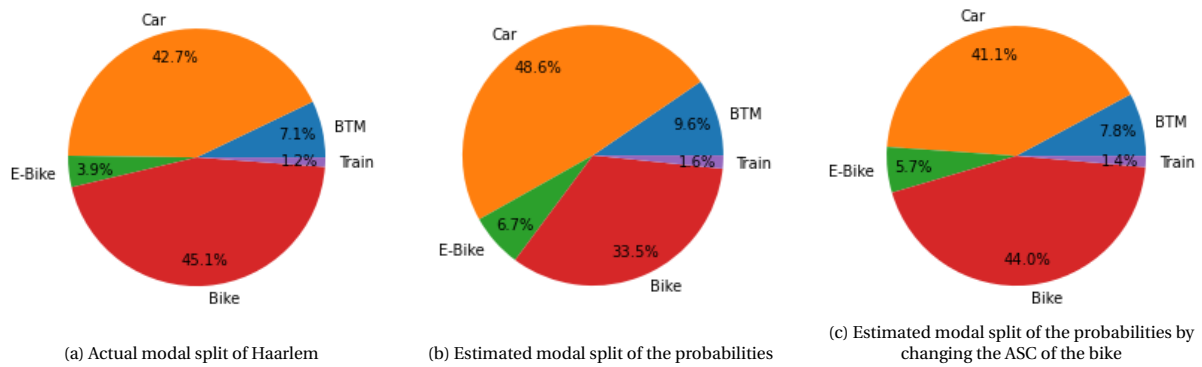


Figure 1: Modal split results of external validation

Besides validating the aggregate results, there is also looked at the accuracy of the individual predictions. Both the internal and external model performances show that the model has high accuracy in predicting for a trip if a mode is chosen or not. The modes being the least accurately predicted are the car and the bike, but if the mode is chosen or not is still predicted accurately for more than 70%. The precision and recall show less good scores and the F1-score is thus also low for some modes. The modes with an F1-score below 50% are the e-bike, BTM, and train. The performance measures thus show that for modes less often chosen (e-bike, BTM, and train), the shares are accurately predicted, but not always estimated correctly for the right individual. The modes often chosen (car and bike) have a lower accuracy of predicting for each trip if it is chosen or not, but are more correctly estimated when predicting from all trips estimated to be made by that mode and from all trips actually being made by that mode.

Conclusions of this research are that the final nested logit model is accurate enough to estimate modal splits in Dutch urban areas for travel distances shorter than 15 kilometers. This concludes the first goal of the research. Related to the second goal of the research, municipalities can use the nested logit model to estimate changes in the modal split when implementing new bicycle policies or adapting the street network. Moreover, factors that are expected to show the largest modal shift towards the bike and e-bike are decreasing travel distances for the bike and discouraging the use of the car by increasing parking restrictions and travel distances for the car. These implementations will probably lead to higher (e-)bicycle mode shares in the modal split. Also, the mode share of the e-bike is expected to increase when a larger share of the population owns an e-bike.

This research has certain limitations. Limitations to assumptions are that it is assumed that the travel behaviour stayed similar to before the covid-19 pandemic and that it is assumed that the choice for a transport mode is made separate from the decision of the destination. Limitations to the data are that it does not precisely represent the Dutch population, that the calculation for the travel distance of the public transport modes is less accurately calculated than for the other modes, and that the final model is not validated for other years. A limitation to the use of discrete choice models is that a mixed logit with random parameters could not be determined because of a lack of computational power.

Recommendations for future research based on the assumptions are to model the effect of the simultaneous choice of trip distribution and mode choice and to research the travel behaviour after the covid-19 pandemic to analyse the changes in travel behaviour and how that would affect the use of the final model. Based on the techniques and methods of this research, it is advised to model a mixed logit model with random parameters to see if it would result in a better model fit. Moreover, it is recommended to test the applicability of the model by validating the modal split outcome of changes in the input of the model based on possible changes in bicycle policies or networks. Based on the results, it is advised to gather the data in a better way and analyse the influence on mode choice for the following factors: bicycle parking, street connectivity, travel speed, and travel time.

CONTENTS

Preface	i
Abstract	ii
1 Introduction	1
1.1 Context	1
1.2 Problem definition	2
1.3 Research objective and questions	3
1.4 Scientific and practical relevance	4
1.5 Structure of the thesis	4
2 Methodological Approach	5
2.1 Use of literature	6
2.2 Gathering the data	6
2.3 Phases in modelling	7
2.4 Validation of the model	9
3 Literature Review	10
3.1 Categories of mode choice factors	10
3.2 Spatial factors explaining mode choice	10
3.2.1 Infrastructural environment factors	11
3.2.2 Natural environmental factors	13
3.3 Characteristics of people explaining mode choice	13
3.3.1 Socio-demographic factors	13
3.3.2 Psychological factors	15
3.4 Characteristics of the transport journey explaining mode choice	16
3.5 E-bike factors	18
3.6 Conclusion	19
4 Data Analysis	21
4.1 Analysis of ODiN data	21
4.1.1 Distance Boundaries	21
4.1.2 Descriptive statistics of ODiN factors	23
4.1.3 Data filtering	26
4.2 Analysis of added data	27
4.2.1 Calculation of added factors	27
4.2.2 Descriptive statistics of added factors	28
4.2.3 Comparison of specified and calculated factors	30
4.3 Correlation	31
4.4 Data for validation	33
4.5 Conclusion	34
5 Modelling	35
5.1 Model requirements	35
5.2 Discrete choice models	35
5.2.1 Influence of each variable	36
5.2.2 Influence of each category & all variables	36
5.2.3 Influence of interaction effects and quadratic components	41
5.2.4 Final model	45
5.3 Sensitivity & Uncertainty Analyses	51
5.3.1 Sensitivity Analysis	51
5.3.2 Uncertainty analysis	51
5.4 Conclusion	54

6	Validation	55
6.1	Internal validation	55
6.2	External validation	56
6.3	Conclusion	58
7	Application	59
7.1	Possible implementations.	59
7.2	Applying the final model	60
7.3	Conclusion	63
8	Conclusion	64
9	Discussion	67
9.1	Limitations to the research	67
9.2	Recommendations for future research	68
	Bibliography	70
A	Appendix A - Overview of found significance of factors	74
B	Appendix B - Overview and information of chosen variables	78
C	Appendix C - Additions to the data analysis	80
C.1	Representativeness of ODiN data	80
C.2	Bar plots from ODiN data	81
C.3	Bar plots from added data.	85
D	Appendix D - Filtering steps	90
E	Appendix E - Additions to the modelling	91
E.1	Motive models	91
E.2	Interaction effects.	95
E.3	Quadratic components	98
F	Appendix F - Base input values	101

1

INTRODUCTION

1.1. CONTEXT

Of all countries in the world, the Netherlands is known for being the largest cycle country. As of 2016, there are around 23 million bicycles in the country and a quarter of the made trips are done using a bicycle (KiM, 2018). The Netherlands is the only country in the world in which people have more than 1 bicycle per capita (Bicycle Dutch, 2018). This is because the Dutch learn to cycle at a very young age. The most common way to travel to school is by bike. Not only the commute to school or work has a large share of bicycle traffic, but recreational cycling is also very common. Throughout the country, there are a large number of cycling nodes with which you can create your own recreational cycling route. Moreover, there are also LF-routes (Langeafstand Fietsroutes) which form a network in the Netherlands and Belgium for multi-day recreational cycling trips (Nederland Fietsland, 2021). As these points show, cycling is an important part of Dutch culture.

The Netherlands is a densely populated country and the population keeps growing as well. The consequences are that the urban areas also keep expanding. Urban areas are defined by CBS as the density of surrounding addresses. These are categorised into 5 classes as shown in Figure 1.1. This figure shows the classes of urban areas at the municipal level in the Netherlands. Only the density of < 500 would be classified as 'no urban area' by CBS, which means that a large part of the Netherlands is in some level urbanised. While these urban areas are growing, research done by KiM (2018) shows the increasing use of the bike in these areas compared to rural areas. In Amsterdam, the choice for the bike has increased the most out of the analysed transport modes with 3.0 percentage points between 2005 and 2016. Moreover, between 30-50% of the trips are made by bicycle in the largest cities in the Netherlands.

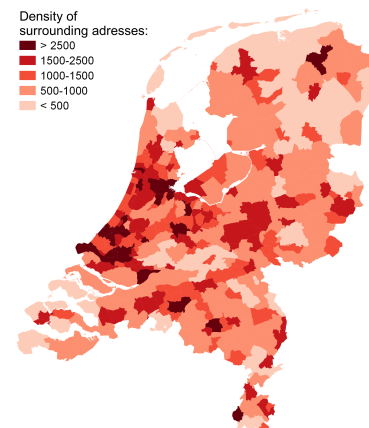


Figure 1.1: Urban areas in the Netherlands (data from CBS (2021))

Not only the bicycle is being used more and more often, the number of e-bicycles has also been growing rapidly. Out of the 23 million bicycles in the Netherlands, 2 million are e-bikes. When they were newly introduced, it was mostly seen as a mode of transport for the elderly to still enjoy cycling in their pastime without the effort. However, in the last years, the use of the e-bike among younger generations has been growing rapidly. The e-bike changed its status from 'an upgraded bicycle for the elderly' to 'a transport mode for all generations'. Between generations, the e-bike is used for different travel motives. For the elderly, these are mostly recreational, while the younger generation mostly uses the e-bike for commuting trips (KiM, 2018). The differences between the e-bike and the bike change the preferences of users on the bicycle network. For example, more space on the cycle path and safe storage facilities are possibly needed (Rijkswaterstaat, 2020). The e-bike largely competes with the bike because of their similarities. But because of the differences, the e-bike is also being chosen for longer distances where it competes with other modalities like public transport and the car.

The use of the car can be addressed as being the most comfortable option, because it can bring people from door to door, does not need any physical effort, and is resistant to all types of weather. Public transport on the other hand does also not need physical effort, is resistant to most weather (except snow and strong wind in the Netherlands), and is often cheaper than the car. However, it does not bring people from door to door and therefore another transport mode is needed for access and egress. The bike then again has different characteristics. It is the cheapest option and can bring people from door to door, but the disadvantages are that it is

not resistant to any weather and it needs physical effort. The last disadvantage can be overcome by using an e-bike as a transport mode. This small overview of characteristics shows the trade-offs people make between these modes of transport, but the real-world trade-offs are far more complicated and extensive than only these four characteristics.

Diving further into detail about the factors that possibly influence the use of the (e-)bike, the success of this transport mode in the Netherlands could be assigned to the large bicycle network that exists. It stimulates the use of the bike, while a small network can be discouraging. The greater the length of the bicycle network in a city, the more increase in bicycle shares can be observed ([Santos et al., 2013](#)). The choice for the bike is therefore often determined by the total cost of travel, in which actual costs and the cost of time or distance are included ([Holmgren and Ivehammar, 2020](#)). However, not only the size of the bicycle network matters. Other factors can play a role in stimulating the use of the bike. This could be municipality policies, but also the comparison to the resistance to other transport modes can matter ([Rietveld and Daniel, 2004](#)). Moreover, multiple network-related aspects can influence the choice for a transport mode. When a network in an area for a mode, without having chosen a route, is overall quite attractive, it can increase its preference for that transport mode. These factors can for example be the priority cyclists have at intersections, and separate lanes along heavily travelled roads ([Pucher and Buehler, 2008](#)).

Multiple researches use different kind of categories to group these influence factors ([Heinen et al., 2010](#)), ([Witte et al., 2013](#)) & ([Wang et al., 2016](#)). The categories made in this thesis research are the following:

- Spatial factors
 - Infrastructural environment factors
 - Natural environment factors
- Characteristics of people
 - Socio-demographic factors
 - Psychological factors
- Transport journey factors
 - Travel characteristics
 - Departure characteristics

Municipalities like to know where possible problems lie in their bicycle network or in their policies, to increase the use of the (e-)bike. The Fietsmonitor from [Witteveen+Bos \(2021 \[13\]\)](#) is a tool that can model the intensities of the bicycle traffic which is shown on a map of the bicycle network. Its goal is to make optimal use of the available bicycle-related data in a 4-step model to visualise the bicycle traffic. Municipalities can use this tool to better justify their choices for their bicycle policies. The tool can also be used to analyse new policies or changes to the network and how this would affect the bicycle traffic flows. Different analyses with the Fietsmonitor has already been done for the city of Haarlem, Utrecht and Province of Flevoland ([Witteveen+Bos, 2021 \[14\]](#)).

1.2. PROBLEM DEFINITION

Reasons for municipalities to stimulate the use of the bike are often public health, and providing a more sustainable city by increasing the choice for sustainable transport modes ([CROW, 2021](#)). However, compared to the car and public transport, objective ways to substantiate bicycle policies are lacking. Modelling the mode choice including public transport and car use has been researched and applied considerably in many different cases, which also includes modelling this choice in urban areas (examples are: [Yang et al. \(2021\)](#), [Gonzalez et al. \(2021\)](#) and [Basso et al. \(2021\)](#)). Research into mode choice in urban areas for the bike, and especially for the e-bike, is much less available. Especially the increasing use of e-bikes needs more attention and research.

The bicycle is not used as much as a transport mode in other countries and that can be a reason for there being less research about it compared to car and public transport. Different studies have been done on factors influencing the mode choice for the bike for certain specific cases, but a more general model including multiple factors for estimating the mode choice in urban areas in the Netherlands is missing and is therefore a research gap that is going to be addressed in this master's thesis. Within the bicycle mode choice modelling, the differences between the bike and the e-bike create another research gap. The number of e-bikes has only been

rising to a significant number in the past couple of years, which created a different and new mode of transport that needs its separate mode alternative in mode choice modelling which has not yet been done in research.

These gaps lead to the following problem statement. Nowadays the bicycle in mode choice is not modelled and researched in the same way as the public transport and car alternative, which leads to an imbalance in comparing modal split results where different accuracies are used for the transport modes. Especially in the Netherlands, where the bike is the second most used mode of transport (KiM, 2018), a more accurate model has to be found and put into practice. 'More accurate' means that it includes more factors than the often used time or distance for determining mode choice in urban areas and that the estimated shares of the modes in the modal split are within 3 percentage points of the actual share. Furthermore, factors have to be found and analysed on their usefulness to increase the use of the (e-)bike in practice to help municipalities with their bicycle policies. Lastly, the upcoming use of the e-bike deserves the same type of modelling which has to be used (when included) as a separate transport mode in mode choice modelling instead of combining it with the ordinary bicycle. This is because of the previously mentioned differences between the e-bike and the bike.

1.3. RESEARCH OBJECTIVE AND QUESTIONS

The objective of the master's thesis is two-fold. First, its goal is to find a mode choice model that predicts a more accurate modal split with the bike, e-bike, car, and public transport that includes multiple factors for urban areas in the Netherlands, to then be able to analyse modal shifts towards the (e-)bike. Second, its goal is to find significant factors for the mode choice that influence the modal split, in order to support municipalities of Dutch urban areas to stimulate the use of the (e-)bike. The research objective leads to the following main research question:

What mode choice model estimates an accurate modal split for Dutch urban areas including multiple influencing factors, which is also useful for analysing changes to (e-)bicycle policies and networks?

The phases needed that will lead to an answer to the research question are a literature review, data collection, modelling, and validation. The sub-questions posed to support the main research question and the case study are given next and these are further explained in the methodology.

Subquestion 1: Which factors known in literature could influence the choice for the (e-)bike, car, and public transport in urban areas?

Subquestion 2: What function including which factors will describe the utility to the mode choice for the (e-)bike, car, and public transport in urban areas?

Subquestion 3: How much do the analysed factors influence the modal split in urban areas?

Subquestion 4: How accurate can the model predict the modal split of urban areas?

The main research question and the sub-questions need to be feasible to answer within the time for a master's thesis. Therefore the scope of the research will be explained by the following bullet points:

- The focus of the first goal will be on creating a generic mode choice model with e-bike, bike, car, and public transport. The decision to not include walking as a transport mode is because of the second goal of the research. This is to support municipalities to stimulate more environmentally conscious transport modes. Walking and cycling can be seen as the most environmentally friendly modes because it has no emissions. However, a change from the car to the bike, e-bike, or public transport has been found to be plausible, although a change from the car to walking is a much larger step and much less often considered. Therefore, walking is not incorporated as a transport mode in this research, because it is assumed that it will not contribute much to a modal shift.
- Travel motives will be classified as leisure and practical motives. Leisure motives will include shopping, visiting, touring, sports, and hobby purposes. Practical motives will include commute and travel-to-school purposes.
- The scale of the research during the literature review will be on a national scale, looking at data from the Netherlands and making a distinction between non-urban and urban. The scale during the modelling will be only on urban areas, focusing on the city of Haarlem in the external validation. Within the urban area, the distance range included will be determined by data analysis.

1.4. SCIENTIFIC AND PRACTICAL RELEVANCE

The scientific contribution of this research would be to improve modelling the mode choice in urban areas on distances up to 15 kilometers by using multiple factors. It gives insights into the influence of factors on the mode choice, which can further help others that want to compare the mode choices for car, public transport, and (e-)cyclists in a more accurate way. Using this model would help in the design of bicycle networks and cycling policies in urban areas. It not only focuses on one type of factor, but it models the important factors from all relevant categories as found in literature. Moreover, this thesis is one of the few papers that models the e-bike as a separate alternative in a mode choice model, and one of the first to model this choice with spatial factors. The inclusion of the e-bike in the research thus will give new insights into the different factors that can be of influence and this can be compared to the bike. This new information is useful in the design of bicycle networks and policies for urban areas but also shows an example for other studies that want to incorporate the e-bike as well. A new model is thus created to be accurate for all modes, to then be able to analyse modal shifts towards the bike and e-bike effectively.

The practical contribution of this research would be to further improve the Fietsmonitor of Witteveen+Bos. Specifically, it contributes to the increase of accuracy for the choice of the cyclist in the tool that models cyclists' traffic flows on bicycle networks. It further helps to understand which factors are of influence on the mode choice, and it helps to better advise municipalities in implementing new bicycle projects and/or policies. Furthermore, the end result of the research can be used for municipalities or other organizations to apply found factors in practice or estimate a modal split for justifying changes in bicycle policies or networks.

1.5. STRUCTURE OF THE THESIS

This thesis is structured as follows. First, the methodology of the research is given in chapter 2, which shows the method of every phase in the thesis that will answer the research questions. Chapter 3 describes the literature review and in chapter 4 the data analysis is performed. Hereafter, the modelling results are given and the application of the model is shown in chapter 5. In chapter 6, an internal validation and external validation with the city of Haarlem are done to see how the model performs in other situations. Chapter 7 shows the application of the model. Finally, the conclusion of the thesis is given in chapter 8 and the discussion & recommendations in chapter 9.

2

METHODOLOGICAL APPROACH

This chapter explains the approach to achieving the objective and answering the main research question. It shows a framework of the approach and an explanation of the methods used in each section of the approach. These are the use of literature, gathering the data, phases in modelling, and the validation.

Figure 2.1 shows the overall approach of the research. The framework distinguishes the different phases which are elaborated in this section after the figure. The numbering in the figure relates to the numbering between brackets in the text. The colour of the cells shows to which category in the approach that action belongs. The black cells are a result of actions and conclude a phase of the research. The cells are connected by arrows, and that shows for which next action the result from the cell is used.

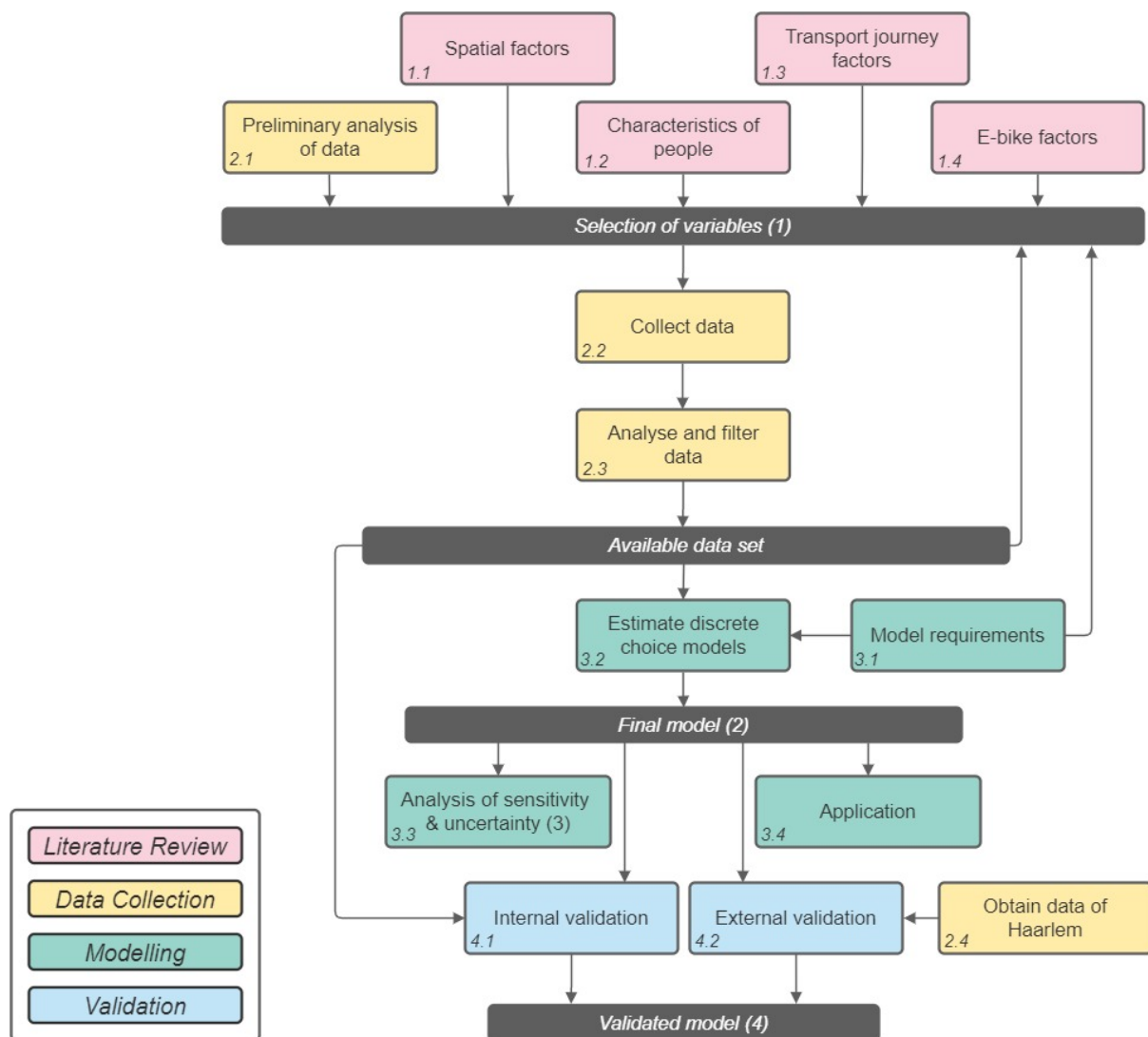


Figure 2.1: Framework of methodology

2.1. USE OF LITERATURE

A literature review is performed to get insights into the factors that are known to influence the mode choice. To give a good understanding of the structure of the literature review, the categories of mode choice factors are explained which forms the sections of the literature review.

Going more into detail, modelling the car and public transport needs input from the existing literature to model that effectively together with the bike and e-bike. Therefore, the literature review includes an assessment of factors for the bike and e-bike, as well as their influence on the car and public transport choice. The subjects in the literature review are based on several kinds of factors that can influence the mode choice. These are spatial factors (1.1), people's characteristics (1.2), and transport journey factors (1.3). Because literature on mode choice for the e-bike is lacking, all literature found on e-bike mode choice is summarized in the final section (1.4). Based on the summarized overview of literature, subquestion (1) is answered by making a selection of factors that are chosen to model.

2.2. GATHERING THE DATA

To be able to model the mode choice for Dutch urban areas, data is needed for the factors that are selected to analyse. In the literature review, factors are researched to get an insight into what can play a role in mode choice, and the method that is used with which researchers analyse and model the chosen data is read. It has been found that the most often chosen option for data collection is to use data from a national or regional survey performed by another party. Another option often seen is to gather data using a self-made survey, but surveys were often only used when the paper was focused on a specific category of factors. With this form of data, it would be difficult to achieve the first goal of the research, which is to create a more accurate modal split prediction. Achieving this requires multiple factors from different categories to analyse and one of the advantages of the national surveys is that this is possible, as opposed to conducting a survey that can only analyse a limited set of about 6-8 factors. Another disadvantage of conducting a survey is that it takes much time, starting from setting up the survey to getting enough respondents. Based on both disadvantages, it is preferred and chosen to not use a survey.

In the Netherlands, there are two large mobility surveys known as ODiN (Onderzoek Onderweg in Nederland) and MPN (MobiliteitsPanel Nederland). The main difference between these two datasets is that respondents of ODiN summarize their trips made on one day of the year, and the respondents of MPN keep a diary for multiple days. Therefore ODiN is more often used to analyse trends in mobility, and MPN is used to analyse changes in mobility behaviour, which is more focused on a personal level. Both national surveys are used in literature on mode choice in Dutch areas and are thus a good option. It depends on the type of factors which national survey is preferred. Based on the factors that are concluded from the literature review, and based on the experience of Witteveen+Bos with using ODiN datasets, ODiN is used as the primary dataset of this research. The benefit of this dataset is that ODiN is based on revealed preference instead of stated preference. It shows the real-life choices of people, which can show influences of mode choice that people possibly would not identify themselves in a stated preference. The data of ODiN is gathered via DANS for the years 2018 and 2019. It is chosen to use data of normal travel circumstances, before the Covid-19 pandemic. The format of the dataset is a CSV file that contains the respondents in rows and the information of factors in columns. ODiN contains data for people's characteristics and travel characteristics, but not for spatial characteristics. Therefore, adding data to the dataset is needed for these factors and this can be done via python using data from OpenStreetMap (OSM) and Nationale Data Openbaar Vervoer (NDOV). ODiN is therefore used as the main dataset, with additions from OSM and NDOV. An overview of the chosen variables can be found in Appendix Table B.2 with their data source, type, measurement, and expectation in modelling.

During the literature review, a preliminary analysis of the dataset from ODiN has to be performed (2.1). This is because it should be known what is available from ODiN and how it fits with the research. Moreover, the boundaries of the distance that will be included in the modelling are also analysed from ODiN data. The travel distance is assumed to be a primary influence (besides the travel time) for the choice for the (e-)bike, so CDF curves are made from the data for urban areas to gain understanding, spot differences, and measure boundaries. A boundary is formed for the travel distance that shows between which trip distances the car and public transport can be substituted by the (e-)bike.

Information that is needed for the modelling phase is data of the chosen factors for urban areas in the Netherlands for the modalities that are chosen to model. Searching for this data is, as mentioned previously, done in the ODIN data and for the infrastructural factors in OpenStreetMap (OSM) and Nationale Databank Openbaar Vervoer (NDOV) (2.2). After gathering the data, it is filtered to have a dataset useful for modelling. To be able to filter the data, an analysis is performed that includes descriptive statistics for the factors and correlation between the factors. It is used to gain insights into the behaviour of the individual factors, and the relationship between factors (2.3). For the validation phase, the filtered dataset of ODIN is split into the data for modelling and validation data of the region of Haarlem. The data for modelling is used for the internal validation and the data of Haarlem is used for the external validation (2.4).

CDF-curves

Cumulative Distribution Functions (CDF) are used to determine distance boundaries for cyclists. Based on research in the Netherlands (Schneider *et al.*, 2020), certain distance boundaries are present at which cycling becomes more attractive and at which it becomes less attractive. The paper shows that below 0.5 km walking is more attractive and that 50% of all bicycle trips are lower than 2 km. However, another paper shows that up to 2 km, the bicycle is not that attractive at all (Keijer and Rietveld, 2000). When it becomes less attractive, other transport modes like the car or public transport are more attractive. Similar boundaries are expected for the bicycle, but for the e-bike, these boundaries are expected to be higher because the travel speed is higher and the physical effort needed is lower.

The data from ODIN is used to determine these CDF curves for all modalities, in which it is expected to show different boundaries between the modalities and thus include the expected longer travel distances of the e-bike compared to the bike. These curves are then used to determine which range of distance is going to be used in the estimation of the model. From these curves, it is also determined which categories can be made for the trip distance to compare modelling results of defining trip distance as a categorical variable, a quadratic variable, or a normal linear variable. A comparable study that used different categories of distance in a logit model is from Barberan *et al.* (2017). In their use of the logit model distance is modelled as three categories, being: shorter than 0.5 km, between 0.5 and 1.0 km, and larger than 1.0 km. Another study focused on trip distances even used 13 different categories between 0.2 and 20 km (Scheiner, 2010). Papers that have modelled distance linearly include Müller *et al.* (2008), Sabir (2011), and Heinen *et al.* (2012).

2.3. PHASES IN MODELLING

The modelling is done using discrete choice models. The requirements of the model are already defined during the literature review (3.1), so that these results can be taken into account during the data collection. These depend on the wanted inputs (variables) and outputs, application, and further constraints.

In the modelling phase, the focus of the model will be on Dutch urban areas. The difference between the bike and the e-bike is part of the research objective and therefore modelled separately. The logit models are estimated to check the significance and the resistance against these factors in the mode choice (3.2).

The factors found and chosen to model in the literature review are being modelled using discrete choice models: Multinomial Logit (MNL), Nested Logit (NL), and Mixed Logit (ML). Logit and probit (less common) models are most often used in modelling probabilities of choosing a particular mode of travel (Holmgren and Ivehammar, 2020). The package used to model is Biogeme, with code language Python. All mentioned discrete choice models can be estimated with this package. The parameters following from the modelling are tested on their significance and usefulness to see which are included in the final model and which are not. The first step in this process is to estimate MNL models with the factors individually together with the alternative specific constants (ASC). The changes in log likelihood and rho-square-bar can be checked to see how much a variable adds to the model fit. Then the variables are modelled per category of factors to see how the significance of variables might change and which correlations between variables are present. After these two modelling steps, the usefulness of the variables is assessed. Some variables can have a large impact on the mode choice but are not useful to include in a model that is used to estimate a modal split between O-D pairs. Think for example of daily influence factors, which can change the behaviour of people each day and are thus hard to include for predictions. After assessing which variables are kept in the model based on usefulness and significance, interaction effects and quadratic components are added to the model and tested on their significance with an MNL model. When significant, they are included in the final model. The last step is to estimate all the variables, in-

teraction effects, and quadratic components with an MNL, NL, and ML model to further optimize results and choose the final model based on the best fit with the data. Some specifications need to be made for an NL and ML model. For both models, nests are tested for modes that possibly correlate with each other. Furthermore, the parameters that need to be estimated randomly can be determined and the presence of panel data needs to be assessed for the ML model. The final model with its parameters and variables answers subquestion (2).

Calibration is to estimate parameters in a model. In a choice model, these are already incorporated in the modelling steps and it is the same as determining the parameters for the variables. This important step in modelling is therefore not further mentioned, but it is included. Furthermore, a sensitivity and uncertainty analysis is performed (3.3), to see the influence of the variables from the model to answer subquestion (3). The final model is created to be accurate for all modes, to be able to thereafter analyse modal shifts towards the bike and e-bike effectively. The practical use of the final model to analyse this modal shift is shown by applying the model to possible bicycle projects of municipalities (3.4).

Logit Models

Discrete choice modelling is based on the utility principle. This is the principle in which it is assumed that the choice for a transport mode is based only on its utility, which can be defined with multiple parameters and variables. The choice that is made is assumed to have the highest utility. To determine the parameters of the variables, logit models are used extensively.

In a *multinomial logit model*, there are more than two dependent variables. The probability that mode i is chosen for individual n given the utilities equals the following formula:

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_j e^{V_{n,j}}} = \frac{e^{\sum_m \beta_m * x_{i,m} + \epsilon_n}}{\sum_j e^{\sum_m \beta_m * x_{j,m} + \epsilon_n}} \quad (2.1)$$

The MNL model does have certain limitations and issues:

- It assumes independence from irrelevant alternatives (IIA), in which the alternatives are assumed to be uncorrelated.
- It assumes the utility associated with common factors between alternatives to not vary across individuals.
- It does not account for possible correlation between choices made by the same individual over time.

These issues can lead to less accurate predictions of the modal split. One way to overcome the first issue is to use a nested logit model, and to overcome all issues is to use a mixed logit model.

In a *nested logit model*, correlation between alternatives is taken into account. The assumption beforehand is that the e-bike and the bike hold some correlation, so the NL model should be used to overcome that issue. The MNL is used as a comparison to the results of the NL. The NL calculation of an alternative in a nest is the probability of the nest (B) times the probability of the alternative within the nest. The formula for the probability that mode i is chosen for individual n is:

$$P_{n,i} = P[n, i | n, i \in B(n, i)] * P[n, i \in B(n, i)] \quad (2.2)$$

In a *mixed logit model*, choices have to be made about which type of distribution is used for the random parameters, which alternatives are possibly correlated, and if choice observations are correlated or not (panel data). A disadvantage is that the mixed logit requires simulation and can therefore take up a lot of time. Furthermore, finding correlation for random parameters is known to possibly lead to overfitting, which in turn leads to poor performance of the model on other data. The probability that mode i is chosen for individual n is now:

$$P_{n,i} = \int [P_{(n,i)} | v_n] * f(v_n) dv_n \quad (2.3)$$

To conclude, the choices for the ML model depend on the factors included in the modelling phase from the literature review. When these are known, the limitations of the MNL and NL can be assessed on their presence in this research and the choices mentioned for the ML can be made. The MNL is still also modelled to compare the results of the NL and ML and to see how much better the NL and ML model performs.

2.4. VALIDATION OF THE MODEL

The modelled urban utility function is validated by an internal and external validation in the final phase of the research. The goal of the validation is to see how well the model can estimate the modal split on in-sample data (4.1) and on out-of-sample data (4.2). For both validation methods, the choice and probabilities of choosing a mode are estimated for each respondent and the resulting modal splits are compared to the actual modal split which includes the actual choices of the respondents. The internal validation also includes a re-estimation of the model without a slice of the dataset to see how similar the model would be estimated without a part of the data. These results are then used to estimate the results of the removed slice. This process is done multiple times, and the log likelihood of the validation slices can be compared.

Besides the validation, the performance of the model can also be assessed. This is done by three measures, and one weighted average measure (Exsilio Solutions, 2016). These measures are calculated for each mode by using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The first is the accuracy, which shows the ratio of correctly predicted observations to the total observations. The formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.4)$$

The second measure is the precision, which shows the ratio of the correctly predicted positive observations to all estimated positive observations by the model. The formula is as follows:

$$Precision = \frac{TP}{TP + FP} \quad (2.5)$$

The third measure is the recall, which shows the ratio of the correctly predicted positive observations to all actual positive observations. The formula is as follows:

$$Recall = \frac{TP}{TP + FN} \quad (2.6)$$

The weighted average measure is the F1-score and takes both false positives and false negatives into account. When false positives and false negatives do not have the same costs, this measure is better to use than the accuracy. In mode choice, false positives and false negatives are equally wrong and thus accuracy is also a good measure to look at. The formula of the F1-score is as follows:

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.7)$$

These measures will give a better idea of the behaviour of the model and how it performs in-sample and out-of-sample. They are used as an addition to the internal and external validation and have to be interpreted lightly. These measures are calculated on the individual choices of people, which is not the main objective. The main objective is that the aggregated choice of a group of people is estimated correctly. However, these measures give a certain insight into the behaviour of the model which is interesting to add to the research. The validation phase is used to answer subquestion (4). When this is done, the research is finished and a modal split model for Dutch urban areas would have been created incorporating the e-bike as well, which answers the main research question.

3

LITERATURE REVIEW

This chapter's objective is to review the existing literature on the knowledge about factors that influence the mode choice. The mode choice originates from the four-stage transport model, in which it is the third step. It follows after the trip generation and trip distribution and comes before the route choice. However, often people's behaviour can be quite complex and thus people do not necessarily always follow this order. Some steps can be taken simultaneously (de Dios Ortúzar and Willumsen, 2011). Nevertheless, this paper only focuses on the mode choice, without the other steps involved in the process. First, this chapter explains the categorisation of factors and how these categories are linked to each other. Second, the factors associated with these categories are discussed. And third, a conclusion is given about the relevant factors for this research.

3.1. CATEGORIES OF MODE CHOICE FACTORS

As previously discussed, this paper focuses on the mode choice which is modelled by using a discrete choice model that uses determinants. These determinants can be formed into categories that are based on a literature review from Witte *et al.* (2013), which are spatial factors, people's characteristics, and transport journey factors.

Figure 3.1 shows the relationships between these different categories. The outside circle distinguishes three types of factors that relate to the modal choice. These are socio-demographic factors, spatial factors, and transport journey factors. The connections represent the interrelation and dependencies possible between the factors. The second circle represents the influence of subjective factors, being socio-psychological factors like habits and experiences. These factors determine how the possible determinants shaped in the first circle are acted upon. The modal choice is positioned in the middle of those, being a result of the interaction of the three outside categories with the influence of socio-psychological factors (Witte *et al.*, 2013). This framework shows the connections between the proposed categories, and can be kept in mind when presenting the literature on factors related to these categories in this chapter.

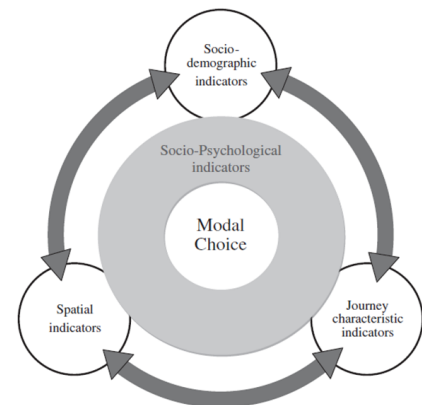


Figure 3.1: Structuring mode choice determinants (Witte *et al.*, 2013)

3.2. SPATIAL FACTORS EXPLAINING MODE CHOICE

Spatial factors encompass everything from the outside environment. This includes the changeable environment such as the network of transport modes and aspects around these networks, but also the unchangeable environment such as the landscape and weather. Assessing if spatial factors are of influence has already been done by a literature review from Wang *et al.* (2016), by summarizing the knowledge of activity for walking and cycling in a neighborhood. It formed four barriers that influence the walkability and bikeability of a neighborhood: opportunity barriers, access barriers, safety barriers, and physical setting barriers. Although this paper focused on the physical activity levels of people in a neighborhood, the mode choice is generated because of trip activities as the first step in the four-stage model. These barriers can therefore present an insight into possible spatial factors influencing the mode choice.

Spatial factors include infrastructural factors and natural environmental factors. Infrastructural factors are mostly focused on the infrastructure itself and the built environment. These are then divided into network aspects, design, and attractiveness of alternative modes. Natural environmental factors are about the weather and climate & environment, which are also its divisions. A framework of spatial factors influencing the mode choice is presented in Figure 3.2. This section discusses the factors related to the categories in the framework.

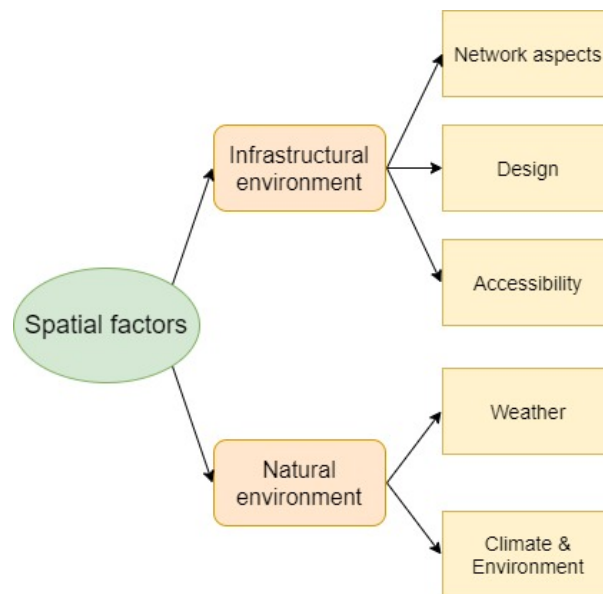


Figure 3.2: Framework for spatial factors

3.2.1. INFRASTRUCTURAL ENVIRONMENT FACTORS

Two studies focusing on decreasing car use and car ownership in urban areas have found that infrastructural factors do matter for the choice of the car. High building density, narrow street width and closely located availability of public transport were factors that decreased the use of the car in urban areas (Yang *et al.*, 2021) (Gonzalez *et al.*, 2021). This does lead to speculating if the infrastructural environment also plays a role in the choice for the bike. This subsection is split into three categories as shown in Figure 3.2 and explains the literature found for these kinds of factors. Table A.1 shows an overview of the factors in this subsection and their correlation with mode choice.

Network aspects

Network aspects focus on the network and its characteristics. Multiple studies have already shown a positive correlation between bicycle network characteristics and cycling levels. An often mentioned factor is the length or density of the bicycle network. This factor is often positively related to the choice for the bike, which means that more length of bicycle path in an area would lead to higher probabilities of people choosing to cycle. Research to this factor has been done in multiple locations, of which multiple cities in Europe (Santos *et al.*, 2013), Washington, USA (Buehler, 2012), South East Queensland, Australia (Wati and Tranter, 2015) and Hamilton, Canada (Eldeeb *et al.*, 2021). The last paper also shows a negative relation to choosing public transport, which means that the increasing choice for the bike because of increasing bicycle network length is mostly at the expense of the choice for public transport.

A factor that is often included from a route choice point of view, is the number of intersections. When choosing a route, cyclists tend to avert intersections with stop signs or traffic to be faster or ride more comfortably (Buehler and Dill, 2015). However, for the actual choice to go by bike, it is most often less relevant and not significant (Ko *et al.*, 2019) (Piatkowski and Marshall, 2015). One paper for students to school in South East Queensland, Australia shows that the same factor can be significant in some areas, although the significance of the positive correlation is not very high and only applies for trip distances lower than 3 kilometers (Wati and Tranter, 2015). Factors closely related to the number of intersections are the stop frequency, which is the number of times the cyclist has to stop per kilometer, and the hindrance frequency, which is per kilometer the frequency of hindrances such as posts of narrow infrastructure. For the Netherlands, it has been shown that both factors can be significant and correlates negatively with the choice for the bike (Rietveld and Daniel, 2004).

Connectivity is another factor that can most logically thinking be related to route choice, but it is also related and significant to mode choice. It can be measured via multiple ways, which can be measuring street connectivity by GIS (Badland *et al.*, 2013), the number of cul-de-sacs (Heesch *et al.*, 2014) or measuring by people's opinions (Titze *et al.*, 2008). Furthermore, different travel motives lead to different outcomes for this factor in terms of significance and it is generally inconclusive.

Design

Next to the network aspects is the design of and around the network. A highly used infrastructural factor related to the design is the separation of bicycle lanes. According to a literature review of [Buehler and Dill \(2015\)](#), separate bicycle tracks are most often showing a positive significance to bicycle choice. This has been proven by research from different locations around the world. This goes from Colombia ([Orozco-Fontalvo et al., \(2018\)](#)), to teenagers-to-school in Cyprus ([Kamargianni and Polydoropoulou, 2013](#)), to Trieste, Italy ([Scorano and Danielis, 2021](#)). However, including a separate bicycle lanes factor in the modelling has been done differently. The research from Colombia and Cyprus use a binary value to check the existence of a separate path on the route of the respondent, while the research from Italy checks the percentage of bicycle paths on its route. Nevertheless, both methods have been shown to be significant in their area.

Safety and comfort levels of the network or the neighborhood can also withhold people to choose a certain transport mode. Safety can be measured by the crime rates of the neighbourhood, or the number of serious accidents. However, not many results show a significance for the choice of the bike, while comfort levels can generally give a slightly positive significant correlation. Comfort level is a latent factor and it is therefore often defined by multiple underlying factors. These can include aesthetics, bicycle parking, carrying cargo, becoming sweaty, fear of theft, and sometimes safety is included as well. Research from the Netherlands shows a slight significance for both safety and comfort ([Rietveld and Daniel, 2004](#)). This result also shows for Australia ([Heesch et al., 2014](#)), in which an increase in safety shows an increase in bicycle mode choice for practical and recreational cycling, while comfort levels are measured by aesthetics and give different outcomes. For recreational cycling, it is positively correlated and for practical motives, it shows a negative correlation. This can relate to the fact that for practical motives, the aesthetics are of less importance and are even unwanted.

Accessibility

The final category within the infrastructural elements is the accessibility of the transport modes. One example is the factor of the degree of access to public transport. This factor can certainly have a large influence on the choice for public transport, but it does not have such a strong influence that it would affect the choice for the bike and the car ([Charreire et al., 2021](#)) ([Ko et al., 2019](#)). This factor is often defined in different ways, in which one method is to include the distance to the nearest public transport for either the origin or destination. With this method, it has been found that access to public transport is more important at the destination than at the origin ([Hallberg et al., 2021](#)) ([Mattisson et al., 2018](#)). Looking at travel motives, it can be found that for commuting trips access to public transport seems more important than other travel motives ([Charreire et al., 2021](#)). Other ways to define the access to public transport is to count the number of bus stops or stations in the area of origin or destination ([Ko et al., 2019](#)) and the number of buses operating in public transport ([Santos et al., 2013](#)). Which type of definition of access to public transport is the best, cannot be determined. All types show a positive correlation and the size of significance is also similar in the found literature. It mostly depends on the type of research that is conducted to determine which definition is most valuable to the research.

Instead of the availability of public transport, also the availability of the car can be assessed related to the mode choice. When a car is available for the trip, the choice for the car is often easier made. Related to infrastructural factors, the availability of the car can be defined as the possibility to go to the destination with the car. Nowadays, more and more cities want a car-free city centre to reduce emissions. These policies can highly influence the choice for the car, but also the choice for the bike as an alternative mode of transport. A literature review shows that car-free city centres do often improve physical activity and higher levels of use of active modes ([Nieuwenhuijsen and Khreis, 2016](#)). Another way of defining the availability of the car is the presence and costs of parking spaces and how these factors influence the choice for the bike. It has been found that this is significantly negatively correlated for commuting trips. One paper from [Ko et al. \(2019\)](#) analyses parking at the origin and the other paper from [Buehler \(2012\)](#) analyses parking at the destination, in which the latter shows to be more influential than at the origin.

Besides car parking, there is also bicycle parking. This is already previously mentioned as a factor sometimes incorporated into the comfort factor. However, some researchers evaluate this as a separate factor. It then has been shown to positively influence the choice for the bike as well for different places ([Nello-Deakin and Harms, 2019](#)) ([Kamargianni and Polydoropoulou, 2013](#)). It can be argued that for the Netherlands, it would not be much of an influence because bicycle parking can be found almost anywhere and thus would not be a thing to think about when choosing a transport mode. However, it is still found to be positively correlated to choosing the bike as a transport mode ([Heinen et al., 2012](#)).

3.2.2. NATURAL ENVIRONMENTAL FACTORS

The natural environment can play a role in determining mode choice. This can be divided into the daily weather, but also into the general climate & environment. This subsection explains the most relevant factors associated with both categories and Table A.2 shows an overview of the factors and their correlation with mode choice.

Weather

One of the biggest factors from this category influencing mode choice would be the rainfall. When it rains, people tend to take the car or public transport more often than the bike. The rainfall very much changes the daily choice of a trip when they normally would take the bike (Scorrano and Danielis, 2021). Sabir (2011) shows that especially the recreational trips are highly influenced by rainfall, while for commuting and educational trips it is less of influence. This may be because it can be linked with comfort, which is often more important for recreational cycling than for cycling to work or school. Other factors that are related to the weather are the temperature and wind strength. These are not often analysed in papers, but research about weather and mode choice from Sabir (2011) in the Netherlands has shown a positive correlation to an increase in temperature in general. When specifying the motive, recreational trips have shown to be the most influential, educational trips are also significant and commuting trips have shown to not be significant. Wind strength has shown to be negatively correlated to choosing the bike, which has the strongest influence on recreational trips as well. Although these significant results are helpful for the choice for the bike, another paper with data from the Netherlands from Rietveld and Daniel (2004) analyses these factors and concludes no significance to both factors.

Climate & Environment

A factor often associated with weather is the season. It is broader and covers all weather-related aspects that are generally present in a certain season. Seasons are different for locations around the world and depend on their climate. It has been found to be often significant for different papers, for example negative correlations are found with the winter season for educational trips in Germany (Müller *et al.*, 2008), for commuting trips in Washington, USA (Buehler, 2012), for trips in Copenhagen (Hallberg *et al.*, 2021) and for commuting trips in Sweden (Holmgren and Ivehammar, 2020). These papers analyse multiple transport modes and especially for choosing the bike, it is often the most influential.

Furthermore, hilliness is also an important determinant, especially for the bike and it is related to the environment. Hills are generally not found in the Netherlands and are therefore expected to not influence the choice for the bike here. However, slopes do occur in the bicycle network in the Netherlands. This is analysed as a factor by Rietveld and Daniel (2004) and it has shown a negative correlation to an increase in slopes in the analysed bicycle network.

3.3. CHARACTERISTICS OF PEOPLE EXPLAINING MODE CHOICE

This section explains the factors that can be associated with mode choice related to characteristics of people. These can be divided into socio-demographic and psychological factors. Furthermore, socio-demographic factors can be divided into categories starting from people's traits and scaling up to household traits and neighbourhood traits. The psychological characteristics of people are divided into attitudes of people and other related psychological factors. An overview can be seen in Figure 3.3.

3.3.1. SOCIO-DEMOGRAPHIC FACTORS

Socio-demographic factors are also known as socio-economic factors, and these are in this paper addressed as the same type of factors. This category of factors is highly often seen in papers analysing mode choice. Mode choice depends on one's behaviour, which can be closely linked to who they are and what their characteristics are. It provides an insight into the different types of people and thus often influences the mode choice a lot. An overview of the following categories and their significance in papers can be found in Tables A.3, A.4 and A.5.

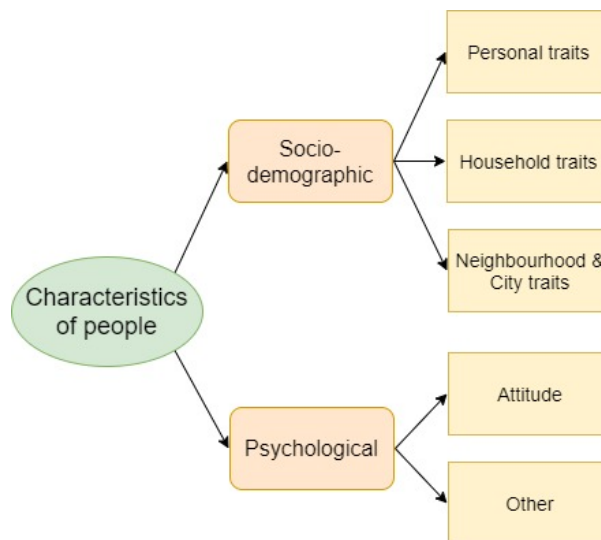


Figure 3.3: Framework for characteristics of people

Personal traits

The four most included factors from personal traits are gender, age, occupation & education. A logical relation to mode choice cannot be given or found in literature for these factors for different transport modes. Inconsistent results remain wherein positively correlated, negatively correlated, significant, and insignificant results are found in all sorts of papers. Often the region and culture play a role in these different outcomes, but even within the same countries results can differ. Although the results are not consistent, it is still used in analyses to give insights into the use of a transport mode in a study area and because it contributes to a more accurate description of the data with the model.

Two other factors that are of importance for mode choice are having a driver's license and personal health. These can logically thinking have consistent results. Having a driver's license is expected to increase the choice for the car, and sometimes has such a large influence that it also decreases the choice for the bike and public transport (Eldeeb *et al.*, 2021) (Kroesen, 2017). Furthermore, the health of people is often defined as the BMI, wherein a negative correlation with choosing the bike for a higher BMI is often found (Charreire *et al.*, 2021) (Mattisson *et al.*, 2018). A paper from Wen and Rissel (2008) specifically analysed bicycle commuting trips combined with being overweight and obesity in Australia. They concluded a strong inverse association between cycling and being obese.

One highly relevant and recent factor for the choice for public transport related to people's health is the outbreak of covid-19. Hygiene and social distancing became important, which are difficult to achieve by using shared transport modes. Therefore, it has been shown that the covid-19 pandemic influenced the mode choice drastically, where many travellers changed their choice from shared transport to private transport (Das *et al.*, 2021) and thus it has been proven that public transport is favored much less during these times (Scorrano and Danielis, 2021).

Household traits

Within households, often the availability of transport modes is a key role in choosing a mode for a trip. The availability or the number of cars in a household is a widely included factor in analyses because it often leads to the same result: significantly positively correlated with choosing the car. Apparently, the availability of a car is so influential, that even the choice for other transport modes such as the e-bike, bike, and public transport are negatively impacted by this (Kroesen, 2017). This result is true for various cases. Its result has been found in analysis of cities in Europe for commuting and non-commuting purposes (Charreire *et al.*, 2021), and also for commuting trips in the USA (Piatkowski and Marshall, 2015) (Buehler, 2012). Not only commuting trips but also travel-to-school trips in Germany show this result (Müller *et al.*, 2008). Besides the car that can be owned by a household, owning a bicycle is also an option that has been found to influence mode choice (Hallberg *et al.*, 2021) (Wati and Tranter, 2015). However, it can be questioned if this would be relevant for the Netherlands because there are more bicycles than people in the country. It would be expected that everyone almost certainly has one available. Nevertheless, it has been found that the results are indefinite. One paper

does show a positive correlation (Heinen *et al.*, 2012), while another does not find any significance in the Netherlands (Ton *et al.*, 2020).

Related to the availability of transport modes in households is the number of household members. More people in one household can mean less availability of, for example, a car in the household that may already be in use when someone wants to choose it for a trip. This then leads to having to choose between other transport modes, which increases the probability of the choice for those transport modes. An increasing number of household members then sometimes shows a positive correlation to choosing public transport and the bike (Holmgren and Ivehammar, 2020) (Ton *et al.*, 2019), although it is not always significant (von Behren *et al.*, 2020). This can be because bikes are often available one-on-one in a household, while cars are much more often shared.

Going from one place to the other can cost money. Using a car is often most costly while taking the bike is the cheapest. Therefore, the income of a household can be a defining factor in choosing a transport mode. However, the results of this factor in the literature are inconclusive. When including the income levels as categories instead of continuous variables, it can be found in one paper that low-income groups prefer active modes, middle-income groups public transport, and high-income groups take the car more often (Ko *et al.*, 2019). Nonetheless, some papers do also show no significance (Charreire *et al.*, 2021) (Rodriguez-Valencia *et al.*, 2021). Therefore, income is a similar factor as gender, age, occupation, and education. It depends on its area and can differ a lot in outcomes.

Neighbourhood & City traits

The last group of traits is again one scale up. These factors can be similar to the personal or household traits but are then analysed on neighbourhood or even city level. One that has many definitions and different outcomes is the population density. It relates to the number of household members, but then for the neighbourhood or the city. The most often found definition is the size of population density at the origin. Although the majority of papers show a positive correlation to bike and public transport choice and a negative correlation to car choice, the results still can differ a lot between papers found in literature. Another way of defining population density is to use the logarithm of population density at the origin. This has been used in a paper from Hallberg *et al.* (2021), which shows again a positive correlation to bicycle choice and negative to car choice. This can be explained by the fact that a high population density often means an urban area. This means more activities and destinations nearby compared to rural areas, which then increases the probability to choose to cycle and the car is then less preferred.

Within a neighbourhood or a city, the age distribution can also be defined. One example of how this can be incorporated into an analysis is to include the proportion of young people. A paper from the Netherlands used this as a factor within their discrete choice model and found a highly positively correlated value for choosing the bike (Rietveld and Daniel, 2004). They found that when the proportion of young people (15-19 years) in a city increases by one percentage point, the share of bicycle use would increase by more than 4%. An explanation would be that the most logical choice for this age group is the bike in the Netherlands, as it is part of the culture in the Netherlands to cycle to almost every destination within their hometown or city. The same paper from Rietveld and Daniel (2004) analysed the number of cars per capita at city level, which is similar to the car availability in a household. It found a slightly significant negative correlation to bicycle choice, which would be in line with the results from the car availability factor.

3.3.2. PSYCHOLOGICAL FACTORS

Multiple papers analyse the influence of psychological factors on people's mode choices. This subsection is divided into the different attitudes of people towards transport modes, and other psychological factors like norms and habits. Table A.6 shows a summary of papers that have results of psychological factors included.

Attitude

Attitude can be described by multiple variants and most often it is generally positively correlated with mode choice. One variant is the attitude towards a transport mode for being efficient. Efficiency can include knowing the travel time, moving quickly and having liberty (Barberan *et al.*, 2017), but also being cost-effective (Kamargianni and Polydoropoulou, 2013) and the ease of use (Simsekoglu and Klöckner, 2019). Another variant is the attitude towards the environment. This can be seen as a person who would want to pollute less with their choice of transport mode (Barberan *et al.*, 2017) (Kamargianni and Polydoropoulou, 2013) or as a person who

is environmentally conscious in general (Ko *et al.*, 2019), although the last one is not found to be significant. Heinen *et al.* (2011) concluded that awareness of the consequences of cycling in general, which includes the environment, is more positively influential on the longer commuting distances for choosing the bike.

A variant that relates to the socio-demographic factor 'health', is the attitude towards health. The difference here is that the attitude is measured by the perception of people, instead of factual data on people's health. Often this type of attitude concludes that people with an active lifestyle or people trying to improve their health more often choose cycling over the car and public transport (Barberan *et al.*, 2017) (Kamargianni and Polydoropoulou, 2013). For commuters, distance also plays a role in this attitude. It has been found that perceiving cycling as a mental and/or physical health benefit stimulate cycling longer commuting distances (Heinen *et al.*, 2011).

Whether a bike is pleasant to ride is also an attitude to consider. It can be defined as the relative convenience (Piatkowski and Marshall, 2015) or as the excitement for the bike (von Behren *et al.*, 2020). The last paper even found such a strong influence, that a significant negative correlation for car and public transport was found if people are excited about bicycle use. The last variant of attitude is comfortable & safe. It relates to the comfort and safety factors in the infrastructural factors, but here it is again measuring people's perspectives instead of using facts or data from the built environment. It can include the feeling of risk of an accident, arriving stressed or sweaty at the destination, feeling of risk of a stolen or damaged transport mode (Barberan *et al.*, 2017) (Piatkowski and Marshall, 2015).

A conclusion on the influence of attitude factors can be given by a paper from Charreire *et al.* (2021) that analyses five European urban areas. It assesses the reason why people chose a certain mode and the three most often given reasons for cycling were just that they like to travel by bike, that they want to be physically active, and that the environment was pleasant or convenient. This is in line with the significant positively correlated results found in discrete choice models in the previously mentioned papers.

Other psychological factors

Besides someone's attitude, people can have a certain opinion about their capacity to use a transport mode which can influence their choice. This is called self-efficacy and is related to bicycle use, this can include believing they can ride uphill, repair a damaged tyre, plan a route, or manoeuvre safely (Barberan *et al.*, 2017). When included in the modelling, it has been found to be positively correlated when people believe they can cycle (Heesch *et al.*, 2014). Linked to self-efficacy, is the opinion of other people about their use of a transport mode. This is the subjective norm and has inconclusive results. One paper from the Netherlands shows that only over small commuting distances (< 5 km) people can be influenced on their mode choice by other people's opinions (Heinen *et al.*, 2011).

It is generally known that habits are hard to break. This also accounts for habitual use of a transport mode. People are often used to a certain way of travelling and that habit is difficult to change, which is also often true for habitual cyclists (Heesch *et al.*, 2014). Heinen *et al.* (2011) even states that the longer the distances, the more influential a personal habit is on commuting motives.

3.4. CHARACTERISTICS OF THE TRANSPORT JOURNEY EXPLAINING MODE CHOICE

This section dives into the characteristics of the transport journey and how these can affect the mode choice. First, the characteristics of travel are discussed in which the most common factors are the travel distance, travel time, and travel costs. Second, the departure characteristics are discussed on their influence which are the departure time and the departure day. The framework in Figure 3.4 shows the factor groups that are being considered in this section, and Table A.7 shows an overview of their significance found in papers.

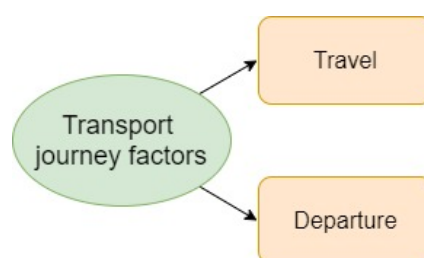


Figure 3.4: Framework for characteristics of the transport journey

Travel characteristics

Modelling mode choice is often focused on using travel distance or travel time as the most important influencing factor. Most research papers that include discrete choice modelling use travel distance and travel time as a linear effect, but also an exponential effect or boundary values have been used for distance (Heinen *et al.*, 2011). Using boundary values can be done by setting thresholds for a certain distance range. An example is to include factors for trip distances smaller than 0.5km, between 0.5-2km, and larger than 2km. These boundaries can be based on assumptions, other literature, or data. Most papers however take a guess for their research using a combination of existing literature and common sense (Kim *et al.*, 2020), and some other papers perform an analysis to determine boundary values (Schneider *et al.*, 2020). The best way to model these factors for the mode choice is still debated.

The relation between mode choice and trip distances is further analysed in a paper from Scheiner (2010). It provides hypotheses based on literature in which travel mode choice does indeed correspond closely with trip distances. Papers using odds ratios show the relation of trip distance towards the other transport modes. Buehler (2012) found that for short commute trip distances smaller than 3 miles (4.8 km), the probability of choosing the bike increases. Winters *et al.* (2010) shows the relation of all travel distances with odds ratios and then shows a decrease in bicycle choice when the distance increases, which is in line with the results from papers using parameter values. In general, an increase in travel distance decreases the utility of all possible transport modes (Müller *et al.*, 2008).

Related to travel distance is the travel time. Most papers trying to form a modal split for an area include either both or at least one of these factors. The relationship between travel distance and travel time can be shown linearly, in which the speed of the modes is assumed constant. More accurate would be to show this relation with non-constant speeds as the speed changes throughout a journey. Graphs are made to visualise this by Wee *et al.* (2006) and it is shown in Figure 3.5. Both graphs show an initial travel time before reaching a distance for cycling and the car. This is because of having to walk from one's place to the car or the bike. The D1 and D2 would define the market areas of the modes. It thus explains which transport mode would be chosen on a given distance if only distance and time were of importance.

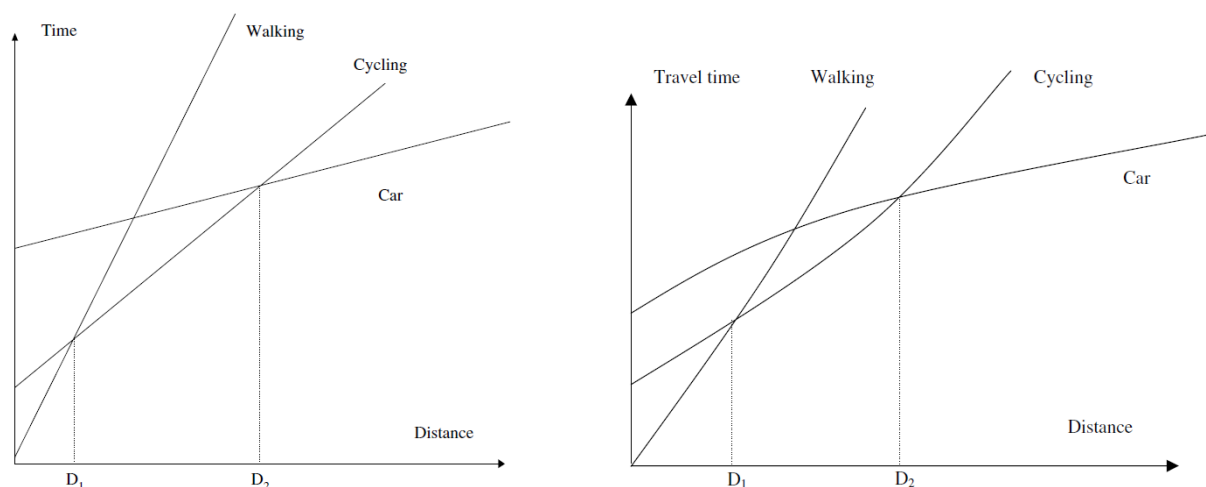


Figure 3.5: Graph representing the relation between travel time and travel distance (Wee *et al.*, 2006)

The speed of a transport mode is the combination of travel distance and travel time, and comparing the speeds between transport modes can therefore be of influence on the mode choice as well. One paper that included the bicycle speed relative to the car in the Netherlands is from Rietveld and Daniel (2004). They have found a significant positive correlation for choosing the bike if the bicycle speed compared to the car would increase. Their conclusion is that this is an essential element that can be influenced by municipalities by designing the spatial network in such a way that there are more direct routes and less number of stops for cyclists. These factors then relate to the connectivity and number of road intersections from the infrastructural factors in Section 3.2.1.

The graphs in Figure 3.5 reveal a positive correlation between travel time and travel distance. This would mean that it is expected that the travel time would also have a negative correlation, just like the travel distance. Papers including travel time in discrete choice models indeed show this relationship to mode choice. In most

cases found in literature the negative correlation to mode choice is the strongest for the bicycle compared to the car and public transport. This means that an increase in travel time is the most undesirable for cyclists. [Ko et al. \(2019\)](#) shows that for travel durations smaller than 30 minutes, the bicycle is more sensitive to travel time increases. This can be explained by the fact that for smaller travel durations, a one-minute increase in travel time is relatively larger than a one-minute increase for larger travel durations.

Literature proposes that for public transport, the travel time out of a vehicle is perceived differently than travel time within a vehicle ([Espino et al., 2021](#)). Transferring can be perceived as more stressful and more needed effort, and waiting time can be perceived as being inconvenient and taking up a longer duration of time. This means that a more accurate definition of public transport travel time would be to use them into two different variables: in-vehicle travel time and out-of-vehicle travel time.

Another common factor included to estimate modal splits are the travel costs. The actual costs for the bike are less relevant, but for using the car and public transport it is. Papers that do analyse bicycle mode choice in combination with travel costs often define it as the operational costs and parking costs, so that the influences of parking costs for the bike can be modelled. As expected, all results found in literature show that an increase in travel costs decreases the utility of the transport mode. [Rietveld and Daniel \(2004\)](#) also analysed the change in mode choice for the bike when the parking costs for the car increase. They have found that this had such a strong influence that the choice for the bike then increases.

Travel motives have been mentioned on their significance sometimes in combination with a factor. In research, it is either focused on a specific travel motive, or different models are made for the travel motives included. However, travel motives are also sometimes analysed separately as a factor within the mode choice modelling. Papers that have done so most often did find a significance to the mode choice, which included that for educational purposes the bike and public transport are often preferred over the car ([Hallberg et al., 2021](#)) and for leisure purposes, the bike is often a favorite choice ([Ton et al., 2019](#)).

A less analysed factor related to travel is the size of the travel group. It was found to be negatively correlated with cycling choice ([Ton et al., 2019](#)). This can be because it is impractical to cycle in a larger group of people. Other uncommon factors would be the number of trips per day, or the number of kilometers travelled per day ([von Behren et al., 2020](#)). These factors are mostly still uncommon because when analysed, the results are not definite. They contribute little to a more accurate mode choice model, although the results in studies when included are useful for their specific goal.

Departure characteristics

Besides the characteristics of the trip, there are also possible influences about the departure that can have an influence on mode choice. These can be the departure time or departure day. This has not been studied often, and the chance of it being significant is around 40% ([Witte et al., 2013](#)). It is related to the travel motive and the need for the trip. Leisure activities are more often chosen to do outside peak hours to avoid busy roads while commuting trips are often expected at the same time of day. For mode choice, it has been found with data from the Netherlands that a weekday has a positive relation to choosing the bike, while the time of day would not matter ([Ton et al., 2019](#)).

3.5. E-BIKE FACTORS

The focus of the review so far is mostly on the conventional bike, car, and public transport. Some could argue that the influences of the bike could also apply to the e-bike as these modes of transport are quite similar. The choice for the e-bike is not analysed that much in papers, but it is being researched more and more often. This subsection is therefore specifically about the findings of the choice for the e-bike. Furthermore, two useful papers found in literature analysed the e-bike in a discrete choice model, so they are also discussed in this subsection.

The benefits of the e-bike compared to car and public transport include lower costs, fewer emissions, and improved physical health. Moreover, the advantages of the e-bike compared to the conventional bike include travelling longer distances with less physical effort and being easier to ride. The e-bike is therefore interesting as a mode of transport on short distances (< 3 km) where it competes mostly with cycling and walking, but also on medium distances (3-50 km) where it competes mostly with the car and public transport ([Astegiano et al., 2019](#)). Findings from Ghent show that distances lower than 5 km have the largest share of trips for a regular

bike, while the e-bike reaches most of its trips under 13 km (Lopez *et al.*, 2017). These results are mostly for the commuting trips, where a car is also a compelling option.

Another way of analysing the choice for the e-bike is by looking at the travel time. One of the biggest findings in most studies is that the e-bike not only makes travel times shorter compared to the bike but also generates longer travel time trips (Hallberg *et al.*, 2021). This makes the e-bike a good substitute for both bikes and cars. The same paper concludes that dedicated infrastructure like cycle superhighways for e-bikes would lead to improved accessibility as higher travel speeds are possible. It would improve the connectivity between cities, which would increase the attractiveness of the e-bike for a larger range of trips. Different age groups could also change the attractiveness and use of the e-bike. Younger generations use the e-bike mostly for commuting with an average travel time of 20 minutes per trip. In the summer periods, it was most often used by them because of the better weather (Gorenflo *et al.*, 2017).

Often the cost of buying an e-bike is a problem for people to actually consider the e-bike as an alternative mode of transport (Ton and Duives, 2021). Multiple papers experimented with offering a trial period of using the e-bike for free to people. The conclusions were that it can qualify as a proper substitute for car use, that trials can have a long-term impact on the mode choice of participants, and that the investment costs indeed are a point of trouble for people (Cairns *et al.*, 2017) (Moser *et al.*, 2018).

Besides this research about the e-bike, two papers actually researched the choice for the e-bike in discrete choice models. Kroesen (2017) analysed the use of a mode in distance travelled in a discrete choice model with data from the Netherlands. Interesting findings are that apparently younger people tend to travel longer distances than older people on an e-bike, while for the conventional bike this would be the other way around. Moreover, owning an e-bike increases the use of the e-bike at the expense of the other transport modes, while owning a conventional bike increases the use of the bike as well as the e-bike. The conclusions of the research are that the e-bike mostly substitutes the use of the bike, but also to a lesser extent the use of the car and public transport. Also, car owners are more likely to switch to an e-bike than to public transport or a normal bike.

The other paper is from Simsekoglu and Klöckner (2019) and analysed socio-demographic & mostly psychological factors related to e-bike choice with data from Norway. Different than the results from Kroesen (2017) are the age and the income, which both show a slightly positive correlation here. Other interesting results from Simsekoglu and Klöckner (2019) are that car use increases the choice for the e-bike, while conventional bike use decreases the choice for the e-bike. Health is also included in the analysis as the level of fitness, but compared to the conventional bike where health often has an influence, it is not significant in the analysis of choosing the e-bike. This can be because not as much physical exercise is needed for an e-bike compared to a normal bike. Related to the psychological factors, Simsekoglu and Klöckner (2019) shows that mostly the image of using an e-bike and the ease of use is of importance for the attitude towards choosing the e-bike. The more positive the attitude, the higher the e-bike choice. Also, the subjective norm of people influenced the choice for the e-bike, which in turn was influenced by environmental needs and consequences. It shows that these results are similar to the findings for the conventional bike, thus it can be assumed that the results for psychological factors found for the conventional bike are possibly also applicable to the e-bike choice.

To conclude, e-bike factors often give similar results as the conventional bike factors, but definitely not in every aspect. Moreover, it is not found that spatial factors are analysed in literature. This is thus an aspect in which conventional and e-bike users could differ.

3.6. CONCLUSION

Linked to the two goals of this research mentioned in Section 1.3, factors that are going to be modelled are chosen based on two things:

- If it has the possibility to increase the accuracy of predicting a modal split for urban areas, or
- If it has the possibility to give useful information for stimulating bicycle use for municipalities.

Active mode choice has been found to be most sensitive to spatial factors and transport journey factors in the Netherlands (Ton *et al.*, 2019). These kinds of factors are also very useful to analyse in the context of both research goals. As the mode choice is most sensitive to both kinds of factors, the accuracy of the prediction of a modal split will most likely increase. Moreover, analysing spatial factors is very insightful for municipalities, as the network can be adjusted to stimulate bicycle use. Therefore, factors from both these categories are certainly included.

Moreover, personal traits would be more of an influence on mode choice than household traits (Ton *et al.*, 2019). Almost all reviewed personal traits are thus included in the modelling, and useful household factors as judged by the literature review as well. The only useful neighbourhood or city trait as found by literature is the population density, which is therefore the only one included in this category of factors.

Finally, psychological factors are often assessed in papers where the focus was set on analysing these factors only. This is because they are often dependent on other underlying attributes which have to be analysed as well. Although often influential, these factors are not chosen to analyse because of the needed underlying attributes. However, a habit can be included to model. This is then measured directly without underlying factors, being the frequency in which the transport mode is lately used.

To summarize, significant factors based on the literature review are presented in a conceptual model in Figure 3.6 about influences of mode choice (see also Appendix Table B.1). In the figure, a distinction is made between the categories as mentioned in this chapter. The relations between these categories are already shown in Figure 3.1 by Witte *et al.* (2013). The lines connecting two factors are expected to have an interaction, which is based on literature findings (solid lines) or own insight (dotted lines). The interactions between factors that are included in the research need to be tested on their significance in the modelling phase.

From the conceptual model, the factors shown in grey are not included in this research. For most factors, the reason is that there is no data available, and/or that it simply would not be expected that it would affect mode choice in the Netherlands very much. For example, the hilliness is not expected to have a major influence. Besides these reasons, the psychological factors would be too time-consuming in the modelling phase. As mentioned before, these factors all rely on underlying attributes, which need preliminary research to identify these attributes and their weights to the psychological factors. It is determined to be out of the scope of this research and therefore not included further. The conclusion is thus that the factors shown in black are analysed and modelled in this research.

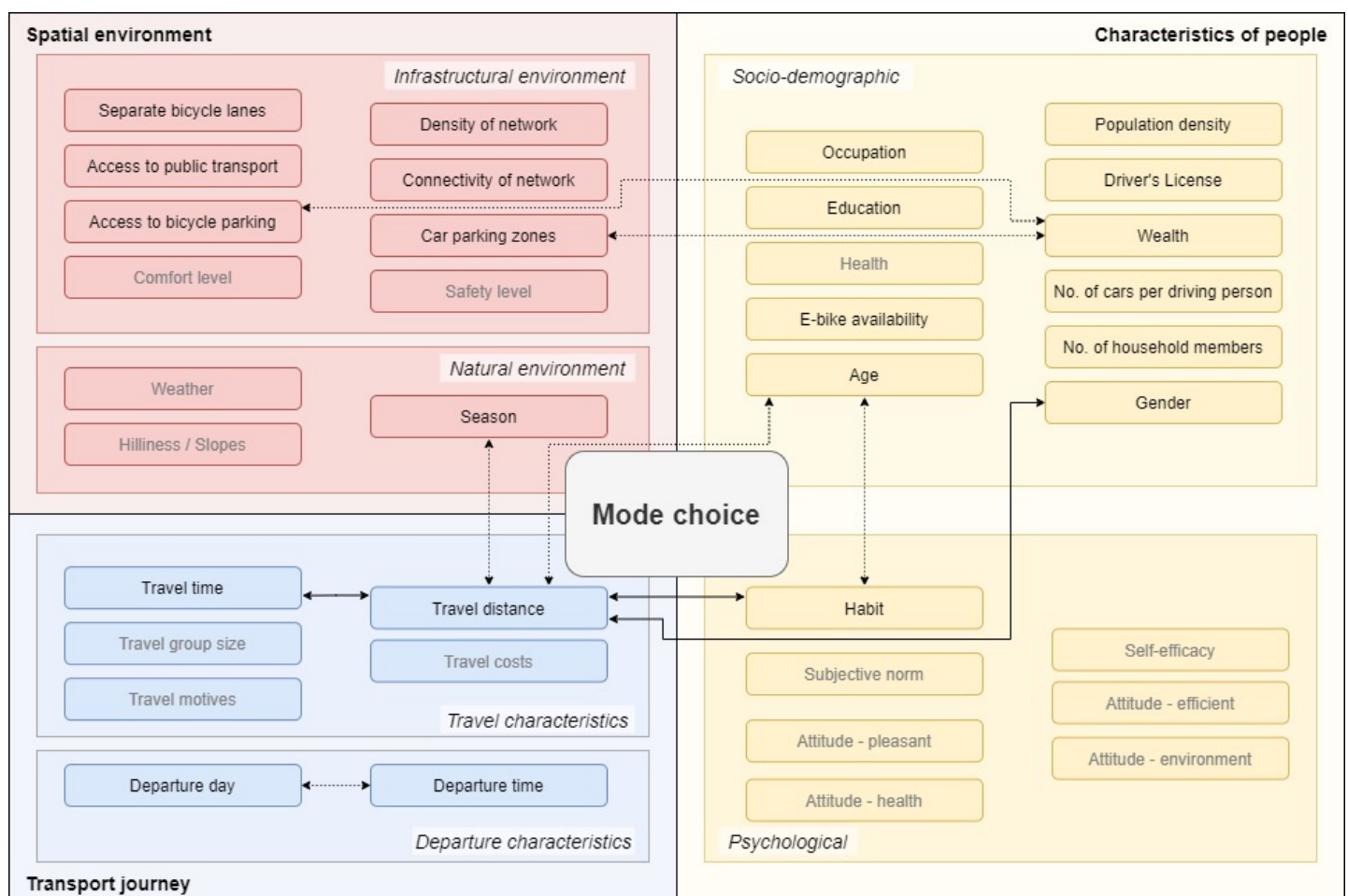


Figure 3.6: Conceptual model of mode choice factors

4

DATA ANALYSIS

This chapter discusses the data analysis of the research. First, the ODIN data is analysed and filtered, after which the added data for the remaining factors are analysed. The analysis for both data sources consist of factors being individually analysed by looking at the descriptive statistics and the behaviour of each factor related to the modes shown in bar plots. After analysing and filtering the data, correlations between all factors and between the factors and the modes are checked. Finally, it is explained what dataset is used for the validation phase and a quick analysis of how it differs from the original dataset is given. The general goal of the analyses is to analyse the behaviour of the data in itself and in relation to the transport modes. It will lead to expectations of the behaviour of the factors to improve understanding of the model results and to improve reasoning to why results may be different than expected from the literature.

4.1. ANALYSIS OF ODIN DATA

In this section, an analysis of the unfiltered data from ODIN is given. First, the number of trips of the modes for the travel distance is plotted as a CDF curve. This analysis aims to gain insights into the average travelled distance of a transport mode for changing circumstances. Moreover, the aim is to determine boundaries for the distance in which the e-bike and the bike are competitive choices compared to the car and public transport modes. There is also looked at changes per motive (practical or leisure) and changes in urbanity for the modes. An urban area is defined as the density of surrounding addresses being > 1000 addresses per km^2 , and thus a non-urban area is defined as the density of surrounding addresses being < 1000 addresses per km^2 . If a trip started or ended in an urban area, it is defined as a trip made in an urban area.

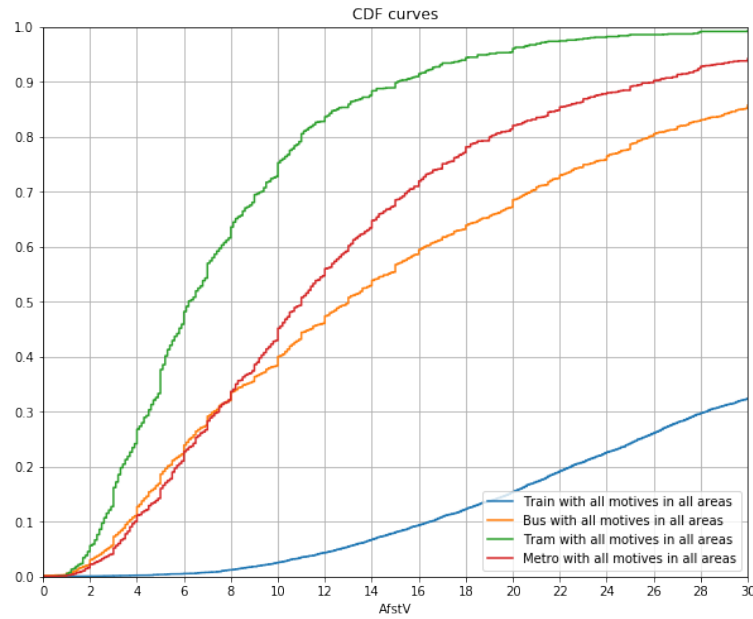
The next analysis is the individual factor analysis. Descriptive statistics and bar plots are shown for each factor. The bar plots show which categories there are and how much a category is present in a dataset. It also shows the share per mode per category of a factor. The aim is to see how the factor behaves for each mode of transport, and to form expectations of each factor for the modelling phase. This analysis is then also used to filter the data on illogical responses and useless categories.

4.1.1. DISTANCE BOUNDARIES

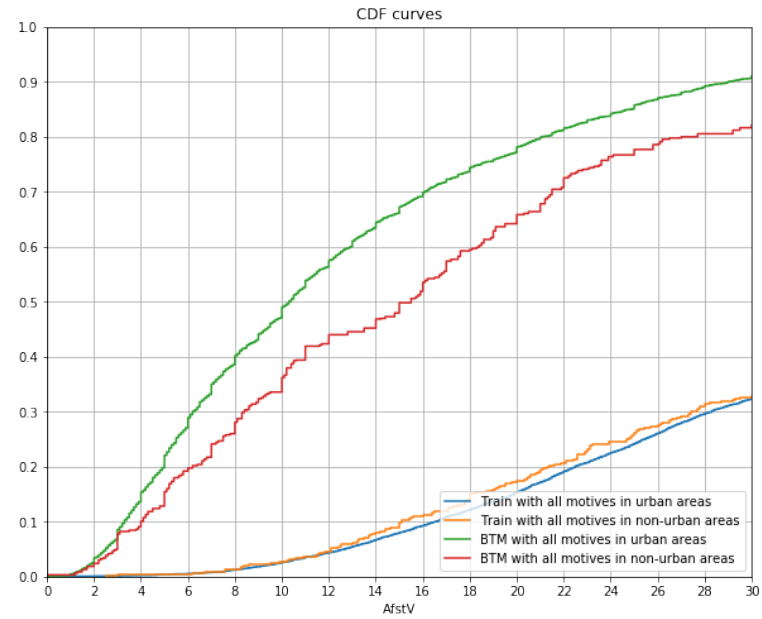
This section provides the analysis of the distance CDF curves with data from ODIN. The data used from ODIN of the travel distance is the distance which is reported by the respondent as the travel distance. Therefore, it is the total distance from the origin to the destination that is often rounded to whole kilometer distances. The curves are therefore not smooth, but more stepwise plotted. Moreover, a drawback of the CDF curves is that it shows the percentage of trips at a given distance which depends on the total amount of trips. Therefore, the interpretation should be done carefully when comparing curves with other modes.

At first, a look is taken at the differences between the public transport types. A CDF curve shows the share of trips taken for a given distance per transport mode. The distinction made in Figure 4.1a is between the Train, Bus, Tram, and Metro. It can be seen that most shorter trips are taken with the tram and the trips longer than 30 km more often with the train, because only about 35% of trips are within the 30 km range. Because the Bus, Tram, and Metro are all competitive modes on the short distance and because they do not have a large share of trips within the total dataset when being separate modes (see Figure 4.5), they are combined as a BTM mode.

Figure 4.1b shows the difference between urban and rural areas in the Netherlands. The curve of the non-urban BTM mode is less smooth than the urban curve because there is fewer data available of BTM modes in non-urban areas. Moreover, it shows that for the train trips the curves follow a similar path for a given distance. However, the BTM mode shows a large difference in the path of the curve, in which for urban areas the share of trips for a smaller distance is larger than for non-urban areas.



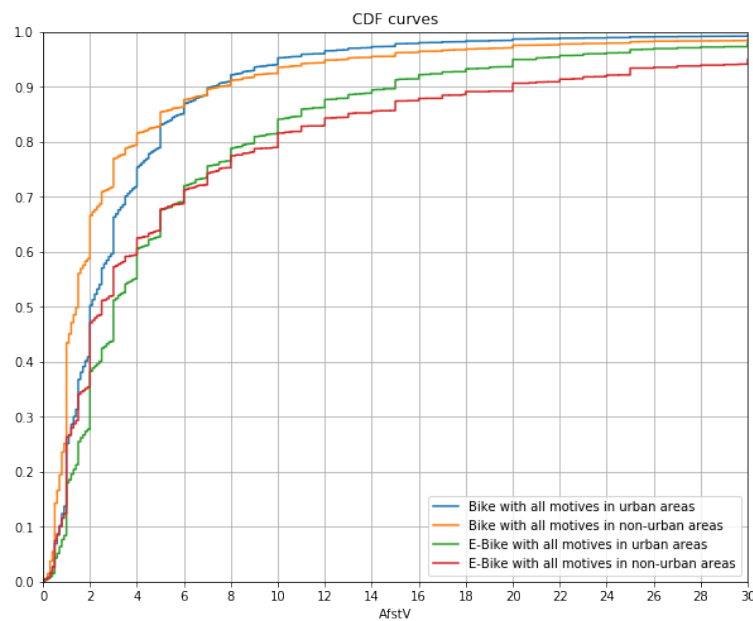
(a) Difference between Bus, Train, Metro and Train modes



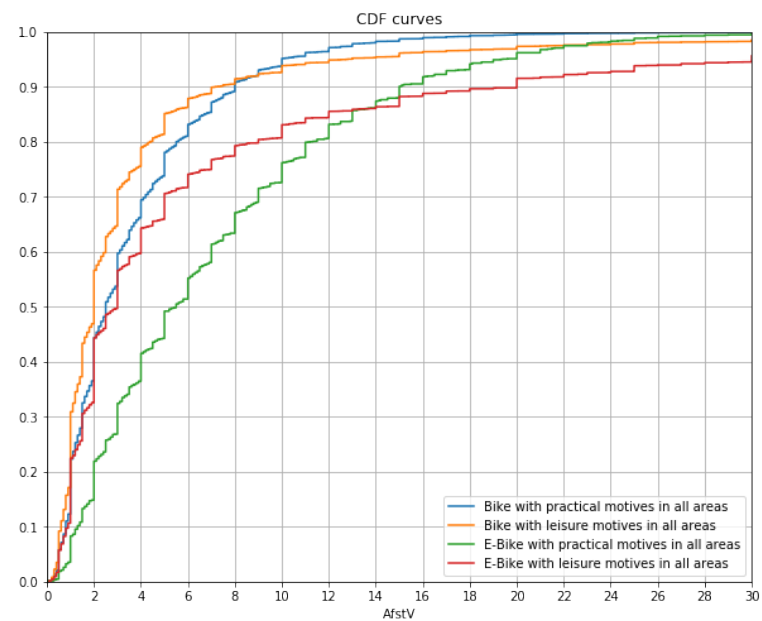
(b) Difference between urban and non-urban areas for PT

Figure 4.1: CDF-curves for PT

The differences between the e-bike and the conventional bike have been discussed, but how do they differ based on a given distance? To get better insights into these differences, the bike and e-bike are also compared to each other in CDF curves. Figure 4.2a shows the difference between urban and non-urban areas, in which it shows that the bike generally is preferred for shorter distances, and the e-bike is preferred for a larger range of distances. This can be explained by the fact that the e-bike needs less physical effort than the bike and people can thus travel further easier. The crossing of the curves shows that in non-urban areas the bike and e-bike are used on short distances (<5 km), but also on long distances (>15 km). This can be explained by the fact that travelling within the village is often for short distances in non-urban areas, but when travelling to another village or a city the travelled distance adds up rapidly. However, the differences between the curves are generally not prominent.



(a) Difference between urban and non-urban areas for bikes



(b) Difference between motives for bikes

Figure 4.2: CDF-curves for bikes

The difference between travel motives in Figure 4.2b shows that the difference between the curves of the modes is more clearly seen. Again in this figure, the two curves for each mode cross. Leisure motives lead to a larger share of short trips for the bike and e-bike. These can be for example shopping in the neighbourhood or going to the gym. Moreover, leisure activities also lead to a larger share of long-distance trips. The figure shows that especially for the e-bike about 5% of the share of trips is not captured in distances up to 30 km. These are for example the sports and hobby trips, or shopping in another city.

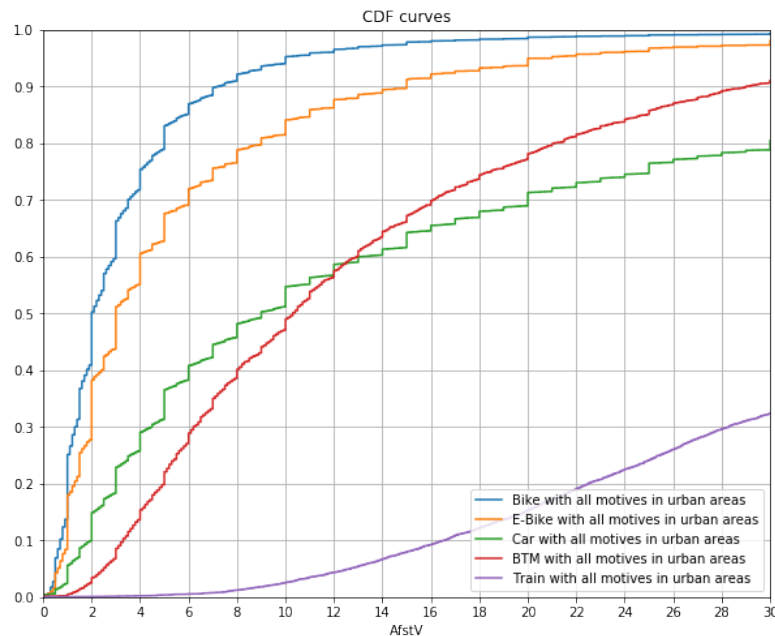


Figure 4.3: Difference between transport modes in urban areas

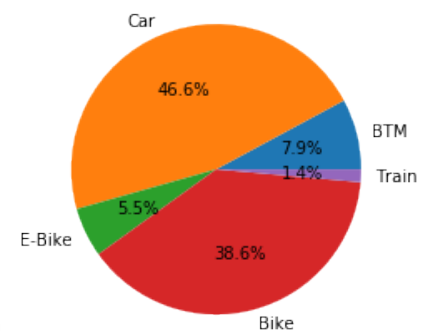


Figure 4.4: Modal Split for distances <15 km

Finally, all transport modes are combined in Figure 4.3 for urban areas. It shows that the bike is chosen for the shortest routes, while the train has the largest share of trips for longer distances. 90% of all e-bike trips and more than 95% of all bike trips are undertaken at distances of 15 km or less. The distance to cover when analysing modal shifts between the bikes and the other transport mode is thus most useful between 0-15 km. Moreover, the modal split for only the car, train, BTM, bike, and e-bike of the Dutch urban areas of 2018 and 2019 combined is shown in Figure 4.4 for distances shorter than 15 km. It shows that the largest number of trips is still taken with the car, although only about 65% of its trips lie between 0-15 km. It shows that 35% of trips taken with the car are thus on distances larger than 15 km. Almost all bicycle trips and e-bike trips are still in this distance range and for the BTM mode, it's about 68% of trips still included. For the train, only about 9% of trips are within this selection, which leads to a very low share within the modal split of 1.4%. Train trips are often taken at a larger distance as it travels between cities. The trips present below 15 km are possibly trips for which the origin and destination are close to a train station, for which it then can be feasible to take the train. Although the share is low compared to the other modalities, the train is in these cases an important alternative.

4.1.2. DESCRIPTIVE STATISTICS OF ODIN FACTORS

To get a better insight into the data of the factors, the descriptive statistics are presented for each factor for the unfiltered dataset. The analysis aims to assess the usefulness of the categories within a factor and how to possibly recategorize factors. First, an insight into the shares of trips per transport mode is given in Figure 4.5 and the shares of trips per motive are given in Figure 4.6. From the shares of trips per transport mode, it shows that the car is most often used. After that, cycling and walking are preferred options. From the public transport modes, the train is most often used. This is probably because it can reach the most destinations in the country, while the bus, tram, and metro are more regionally bound. Combining the bus, tram and metro will lead to almost a similar share of trips as for the train. From the shares of trips per motive, it shows that the commuting motive has a large share of trips for especially BTM modes, which could mean a large part of respondents work and they often go by BTM modes. The education motive has a large share of the bike and also BTM modes.

Often teenagers do not have a license and go by bike, and students have free public transport and often go by BTM modes. Moreover, the shopping motive has a large number of trips as well. Here the e-bike is most often chosen and BTM modes are then less favorable.

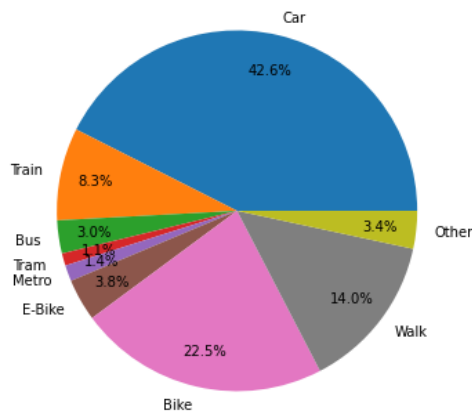


Figure 4.5: Shares of trips per transport mode

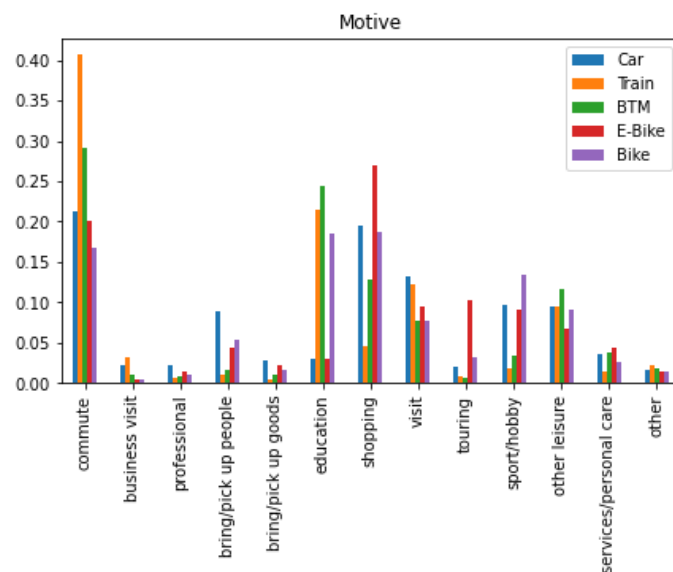


Figure 4.6: Shares of trips per motive

The motives are grouped into practical and leisure motives as these are the most common motives for travel. It provides insights into the differences between these two classes of motives in the results of this research. For practical motives, it includes commute and education. These have one of the largest shares of all motives and they determine the daily travel of most people during the week. For the leisure motive, it includes shopping, visiting, touring, sport/hobby, and other leisure. The remaining motives do not have large shares or do not belong in one of the two groups. They are thus not included in the two classes but are grouped as the 'other' motives. Analysing the shares by these groups and also specifying the transport mode gives the results shown in Table 4.1. When looking at the share of trips taken by e-bike, most is still because of leisure activities. The shares are somewhat similar to the conventional bike and the car. The shares for public transport show the largest share for the practical motives, and fewer trips for leisure motives. The car is the most widely used across travel motives as there is also 23% undefined by other motives.

Table 4.1: Shares of trips per motive

	Car	Train	BTM	E-bike	Bike
Practical motive	24%	62%	54%	23%	35%
Leisure motive	54%	29%	36%	63%	53%
Other motives	22%	9%	10%	14%	12%

To further analyse each factor, Table 4.2 shows the shares per category in percentages for categorical or binary variables, and the mean and standard deviation are shown for continuous variables. To check the representativeness of the sample, the shares of the people's characteristics are compared to the national shares per category. These can be found in Appendix C.1. The conclusion is that overall the sample is representative enough to model with the sample of the ODiN data. The only differences to consider in modelling are that the ODiN sample has more respondents being middle-aged, higher educated, and living in a wealthier household. In Appendix C.2, bar plots are also made to visualize the shares per value of a factor per transport mode. These statistics show that the split between men and women is about equal, with a preference for BTM modes and the e-bike for women and the car and bike for men. The largest share of e-bike users is still the older generation, while the normal bike is most favorite amongst younger people. The upcoming use of the e-bike amongst youth is thus not clearly present in this dataset. Car and BTM modes are more for the middle-aged group of people. Because of this difference between ages, age is defined as a categorical variable in modelling to analyse these different age groups separately.

Table 4.2: Characteristics of ODin factors

Factor	Values	Share / Mean (std)	Factor	Values	Share / Mean (std)
Gender	men	49.59%	Habit bike		3.57 (1.52)
	women	50.41%		never	18.86%
Age		41.53 (20.79)		every year	7.23%
	0-18	17.45%		every month	11.49%
	18-40	32.69%		every week	22.59%
	40-67	38.80%		(almost) every day	39.84%
	67+	11.05%	Habit e-bike		1.55 (1.21)
Occupation	part-time job	13.88%		never	80.20%
	full-time job	37.17%		every year	3.01%
	own housekeeping	10.01%		every month	3.98%
	student	16.28%		every week	6.66%
	unemployed	1.49%		(almost) every day	6.15%
	unable to work	2.24%	Habit BTM		2.15 (1.24)
	retired	14.39%		never	39.74%
	other	4.10%		every year	28.60%
	unknown	0.43%		every month	15.82%
Education	no education	1.09%		every week	8.54%
	primary education	3.75%		(almost) every day	7.29%
	vmbo/mavo	13.60%	Habit train		2.07 (1.17)
	havo/vwo	29.34%		never	38.34%
	hbo/university	38.53%		every year	36.20%
	other	2.07%		every month	12.12%
	not asked	11.62%		every week	6.59%
Driver's license	no	27.05%		(almost) every day	6.75%
	yes	72.95%	Habit car		3.66 (1.44)
Cars per driver's license		0.76 (0.50)		never	14.41%
	(0.0, 0.25]	10.38%		every year	3.73%
	(0.25, 0.5]	30.78%		every month	10.31%
	(0.5, 0.75]	5.61%		every week	24.65%
	(0.75, 1.0]	40.75%		(almost) every day	32.60%
	(1.0, 2.0]	5.77%		not relevant	14.30%
	(2.0, 10.0]	0.52%	Travel distance		15.24 (28.10)
Household members		2.88 (1.41)	Travel time		30.72 (42.32)
Wealth		6.40 (2.78)	Departure time	0:00-7:00	2.73%
	first 10% group	5.99%		7:00-9:00	13.43%
	second 10% group	5.57%		9:00-12:00	17.14%
	third 10% group	6.67%		12:00-16:00	27.42%
	fourth 10% group	9.10%		16:00-19:00	22.00%
	fifth 10% group	9.66%		19:00-24:00	12.41%
	sixth 10% group	10.45%	Departure day	Sunday	11.20%
	seventh 10% group	11.43%		Monday	14.81%
	eight 10% group	12.14%		Tuesday	14.99%
	ninth 10% group	13.34%		Wednesday	14.67%
	tenth 10% group	14.14%		Thursday	15.09%
	unknown	1.50%		Friday	15.54%
Availability of e-bike	no	77.12%		Saturday	13.70%
	yes	22.88%	Season	Spring	24.95%
Address density Origin		2018.88 (1847.13)		Summer	24.49%
Address density Destination		2019.20 (1847.20)		Autumn	26.19%
				Winter	24.37%

People with full-time jobs most often choose the car, but also BTM modes have a large share. When being a student, the BTM and the bike are by far the favorites as could also be seen from the motives. The shares per mode of unemployed and unable to work are similar and are thus grouped as one category in the model.

For the education and occupation, the categories of other, unknown, and not asked are not giving relevant information to what would influence the mode choice and are thus not used in the model. About 30% of the people do not have a driver's license of which remarkably some still chose to go by car. When choosing the car, often there is about 1 car available to every person with a driver's license in a household. When fewer cars are available, the choice for the car also lowers. The number of people in a household is most often two or four, of which the e-bike is most popular for a two-person household and the bike for a four-person household. The wealth of households is mostly in the largest 10% group, which could mean that costs for travel or availability of a transport mode would generally not be as big an issue. It shows that the largest 10% group most often chooses the car, while the lowest age group by far chooses BTM modes, probably because it is cheaper. The largest difference between categories is between the first and last 10% group. The changes between each next category are then relatively small, and thus will be categorized as five 20% groups in the model. Here also the unknown category is not used in the model for the same reason as for the education and occupation factor. The availability of the e-bike is not very large: about 20% of people own an e-bike. Remarkably, there is still a small share for the choice of the e-bike, although they would not own an e-bike. From the people who do have an e-bike, the share of the car choice is the second most often made choice. Combined with the result of the driver's license option where the e-bike is also the second most often made choice, it shows the possibility of these two modes being a substitution mode for one another. The address density at the origin or destination shows an average of 2000 addresses per kilometer, which can still vary a lot. In small address densities, the car and (e-)bikes are preferred, while public transport is more preferred in higher address density areas. These areas are more towards cities in which public transport is more present and probably thus more chosen.

Habit is defined as the frequency for which a mode is taken. These are described in five categories, being: never, every year, every month, every week or (almost) every day. The habits per transport mode show that generally, the e-bike is not taken that much. 80% never use an e-bike, which complies with the availability of the e-bike in which 80% do not even own an e-bike. If the e-bike is used, however, the most common frequency is then to use it weekly. Moreover, the conventional bike is taken almost daily, together with the car. The BTM mode and the train are rarely taken, but there are also fair shares of people using it every month. The plots show the influence of habit even further: for each mode of transport, the habit of a certain mode being every day has the largest share for that analysed certain mode. It is therefore expected that the habit has a large positive influence on mode choice in the model.

The trips taken most often are a short distance between 0 and 2.5 km, which is a logical outcome as these trips are made in urban areas. The bike is then a favorite, while BTM modes are more often taken on longer distances. Linked to this, the travel time is thus also most often of short length between 0 and 20 minutes. Looking at the departure moment, this is often in the middle of the day between 12:00 and 16:00. A difference between transport modes is not clearly seen. Throughout the week the number of trips stays the same amount, but the least amount of trips are taken on a Sunday. Moreover, BTM modes have the lowest share on the weekend, probably because fewer BTM modes are driving on the weekends compared to weekdays. Looking at the whole year, the number of trips per season is about equal, which is logical when thinking that trips normally do not depend on a season. Only a small drop in the choice for the (e-)bike in the winter can be seen, which can be explained by the worse weather in these months. It is expected that the influence of the season is therefore similar for each mode, the influence of the departure time and day will be more prominent for public transport as these shares are high in peak hours and low on weekends.

4.1.3. DATA FILTERING

After having analysed the whole dataset, the data can be filtered based on conclusions made in the analysis of the ODIN data. It is filtered to be able to select only the data that is needed and to make the results more accurate by filtering out unreliable or incorrect data to make the data useful for modelling. What data is removed from the dataset, can be found in Appendix D.

For filtering the ODIN dataset, the first step is to focus the dataset on the modes that are included in modelling. These are the car, bike, e-bike, BTM, and train. Then select only inputs from urban areas by selecting all rows that have an origin or destination in a zip code with an address density larger than 1000 addresses per square kilometer. There is also filtered on distances shorter than 15 km, which results from the analysis in Section 4.1.1. Finally, only the columns are selected that include the data for the factors, together with the origin and destination of the trip and the mode choice. Which columns these are for the factors, can be seen in Appendix Table B.2. After having filtered the data, the number of respondents is 116783, which is expected to be enough

for estimating an accurate model. If the number of respondents eventually is too few, this can be seen in large standard errors of the variables which leads to less significant factors.

4.2. ANALYSIS OF ADDED DATA

In this section, the data for the remaining factors are analysed. The data sources that are used are OSM, NDOV, RDW, and CROW. First, the explanation is given of how the factors are calculated and added to the ODiN dataset. For further explanation of the factors, see the tables of appendix B. It summarizes the description, units and type of variable, the data source, how the variable is measured, and the expectation of the significance of the variable in the model. Then, the descriptive statistics and bar plots are discussed for the added factors. The goal is to gain insight into the behaviour of the factors and form expectations for each factor of their usefulness and results in the modelling phase. Finally, a comparison of the specified and calculated travel distance and travel time is shown, which is done to check the accuracy of the calculated travel distance and travel time.

4.2.1. CALCULATION OF ADDED FACTORS

To further clarify the added factors, the calculations are explained in this section. In general, the origin and destination zip code determine the area in which these factors are calculated for each row in the filtered dataset from ODiN. The ODiN data works with a zip code level 4 as a detail level of the origin and destination locations given by respondents. The value resulting from the calculations done in Python is added as a column to the dataset from ODiN for each row, which represents a trip.

The first factors are the street density, street connectivity, and separate bicycle lanes. The data source of these factors are OSM for the car and (e-)bike, and NDOV for the train and BTM. The data is imported in Python via shapefiles that contain the network of the Netherlands. To analyse these factors on the whole route, an ellipse is drawn around the origin and destination of a row. The length is determined by the distance between the origin and destination, and the width of the ellipse is then smaller by a factor of 0,6 compared to the length. The locations are based on the center of gravity of the addresses in the zip code area, as this provides the highest probability that the trip has started/ended close to that location. Within the ellipse, the network is assessed. For the street density, the lengths of all streets for the specific mode are summed (counting two-way streets double) and then divided by the area of the drawn ellipse. This definition can be found in papers from [Eldeeb et al. \(2021\)](#) and [Ko et al. \(2019\)](#). The value is then divided by 100 to reduce the size of the betas in estimating the parameters.

$$\text{Street density (km/hm}^2\text{)} = \frac{\text{Length of streets}}{\text{Total area} * 100} \quad (4.1)$$

The street connectivity is calculated for the car and (e-)bike mode specifically by dividing the number of edges by the number of nodes in the drawn ellipse.

$$\text{Street connectivity} = \frac{\text{No. of edges}}{\text{No. of nodes}} \quad (4.2)$$

The separate bicycle lanes are calculated as a ratio of the length of separate bicycle lanes compared to the total length of all cyclable roads in the drawn ellipse. This definition is based on the factor found in literature from [Scorrano and Danielis \(2021\)](#). The length of separate bicycle paths could also be used as a definition by [Nello-Deakin and Harms \(2019\)](#) and [Santos et al. \(2013\)](#), but it is expected that it would then correlate too much with the street density of the bicycle network.

$$\text{Separate bicycle lanes} = \frac{\text{Length of separate bicycle lanes}}{\text{Length of all cyclable roads}} \quad (4.3)$$

Access to public transport as found in the literature review is split into two different factors. One is defined as the frequency of public transport and it is based on the stops in the origin or destination zip code. Data is gotten from NDOV in which the frequency of each BTM stop is given. The frequencies of the trains are not available. It is calculated as an average frequency per hour of each BTM stop. A difference is made for the weekdays, Saturdays, and Sundays. It is assumed that the weekdays have similar frequencies, while the frequencies on Saturday and Sunday can differ from each other and the weekdays. For each row in the ODiN data, the average frequency of the origin zip code and destination zip code is then added by summing all average frequencies of the stops in a zip code area. This definition is based on a similar factor found in literature from [Santos et al. \(2013\)](#), which defines access to public transport as the number of buses operating in the public

transport per 1000 population. Instead of the number of buses, the frequencies are summed in this research and also divided by 100 to reduce the size of the betas in estimating the parameters.

$$\text{Frequency of public transport} = \sum \frac{\text{Frequency of stop}}{100} \quad (4.4)$$

The next factor related to access to public transport is based on the catchment area of BTM and train stops, and it is also called access to public transport. The catchment area of BTM stops is 400m, and for train stops 800m. The factor is calculated as the ratio of area covered by catchment areas compared to the total area of a zip code. This is again calculated for both the origin and destination zip code for each row in the ODIN data. This definition is based on the factor found in literature from [Nello-Deakin and Harms \(2019\)](#).

$$\text{Access of public transport} = \frac{\text{Area covered by catchment areas}}{\text{Total area}} \quad (4.5)$$

The bicycle parking factor consists of the presence of a parking facility in each zip code for each category. It is not expected that a change from one to two bicycle parking facilities changes much to the utility of a transport mode, so it is modelled as a binary variable. The car parking zones factor is calculated as a percentage of the paid parking area in a zip code compared to the total area of a zip code. Both factors are calculated for the origin and destination location. The value is then divided by 100 to reduce the size of the betas in estimating the parameters.

$$\text{Car parking zones} = \frac{\text{Area covered by paid car parking areas}}{\text{Total area}} \quad (4.6)$$

To calculate the travel distance and travel time for the car and the (e-)bike, the package OSMnx is used in Python, which imports data from OSM. Calculating the fastest route based on distance or time takes a lot of time when having imported a large area. Therefore, the provinces are used as the area in which a route is calculated. The result of this method is that when a row in the dataset contains an origin being in another province than the destination (or the other way around), no travel distance or travel time can be calculated. The rows for which it can be calculated are based on the Dijkstra algorithm calculated by OSMnx.

The calculation of the travel distance and travel time for the public transport modes is different. The data in OSM did not have the right classification to make the difference between train and BTM modes and it did not have the roads on which the buses drive, thus data of NDOV is used which does have that information. The line network and the stops are connected to each other, and the average service speed is added to the lines to be able to calculate the travel time as well. This is transformed into the right format so that OSMnx can again calculate the shortest path with its algorithm.

4.2.2. DESCRIPTIVE STATISTICS OF ADDED FACTORS

The added data are analysed in the same way as the ODIN data. The data is added to each row in the ODIN dataset, thus it is analysed for each trip. Therefore, a table with characteristics in [Table 4.3](#) and bar plots in [Appendix C.3](#) are also made for these factors to get a better insight into their values.

For the street network characteristics of the bike, it shows that a larger street density indeed increases the use of the (e-)bike over the other modes. However, it is less prominent for street connectivity. It looks like the public transport modes benefit most from an increase in cycling connectivity and this could mean the factor does not describe the street connectivity in a representative way. Car use increases compared to the other modes for an increase in car street density, although the (e-)bikes also benefit from the increase. This can be explained by the fact that bike users can also cycle on residential roads. Moreover, the connectivity of the car network again seems to be most beneficial for public transport modes. In the model, it is thus expected that the street density will have a positive influence on the mode choice, while street connectivity may not show this. Moreover, separate bicycle lanes show to have a positive relation to choosing public transport modes instead of the expected bike and e-bike. It can therefore mean that it does not estimate the influence of separate bicycle lanes correctly. When looking at the street network of the bicycle lanes, it can be caused by the higher presence of separate bicycle lanes in the cities. Public transport modes are also more present in the cities and can therefore lead to this increased choice for public transport when there are more separate bicycle lanes.

The characteristics of the public transport network are a little different. The density is calculated in the same way and shows indeed an increase in public transport modes when the network density increases. However, the connectivity cannot be calculated in the same way, as not all edges in an area are accessible between two points. Think of train services on a train track passing through the countryside, but no stops available

to get on the train. Then when calculating the number of services divided by the number of stops for two points in this area, the connectivity would be very high while this is misleading. A low number of stops would therefore benefit the connectivity, but this is counter-intuitive. Therefore, other characteristics are included for the public transport network. The first is the access to public transport, which is defined as the percentage catchment area of all BTM stops or train stops in the origin and destination zip code. The catchment area of a stop is hard to estimate. Research shows that it also depends on other factors, like the frequency and the speeds of the lines at a stop (Brand *et al.*, 2017). However, the focus of this research is not on analysing the influence of other factors on the catchment area of stops, so the catchment area is defined only by distance. Often the distances for access and egress vary for each location, but mostly a fixed catchment area of 400m is used for bus, tram and metro stops, and 800m for the train (Rijsman *et al.*, 2019). Based on the research, the catchment area is thus defined as a circle of 400 meters around a BTM stop and 800 meters around a train stop. This leads to an average of 80% of a zip code being accessible to BTM stops and about 20% to train stations. It indeed shows an increase in public transport modes when the access to an origin or destination increases. Moreover, access can also be explained by the frequency of the public transport services at the origin and destination. A lot of BTM stops can be in the area, but if the frequencies are low, it is still not always a preferred choice. The data show that the frequencies can differ a lot for each origin and destination zip code. It does lead to an increase in public transport modes when the frequency of the services increases for both origin and destination. It is thus expected that the access and frequency factors both have a positive influence on the public transport modes.

The car parking zones are often not present at a given origin or destination. But when it is present, it often quickly leads to a large decrease in car choice compared to the other modes and thus is expected to have a negative relation to car choice. The bicycle parking facilities are not present often in a given zip code. The bar plots show that the train benefits most from bicycle parking facilities, but an explanation can be given by the fact that often these facilities are located at train stations. It further shows only a minimal increase for (e-)bikes for both facilities compared to the car. This can mean that it probably is not a good predictor for choosing the bike, but it can be for the train. However, it probably explains another phenomenon, which is the presence of a train station in an origin or destination zip code. It is therefore better to include the influence of train and BTM stations at origin and destination, as this would probably be a better definition for this phenomenon. However, when analysing this as a factor being the number of stops at origin and destination, a correlation can be found between the number of BTM stops and the frequency of BTM stops, and a correlation between the number of train stops and the access of the train. This can be seen in Figure 4.7. Therefore, the number of stops is not included as a factor in the model.

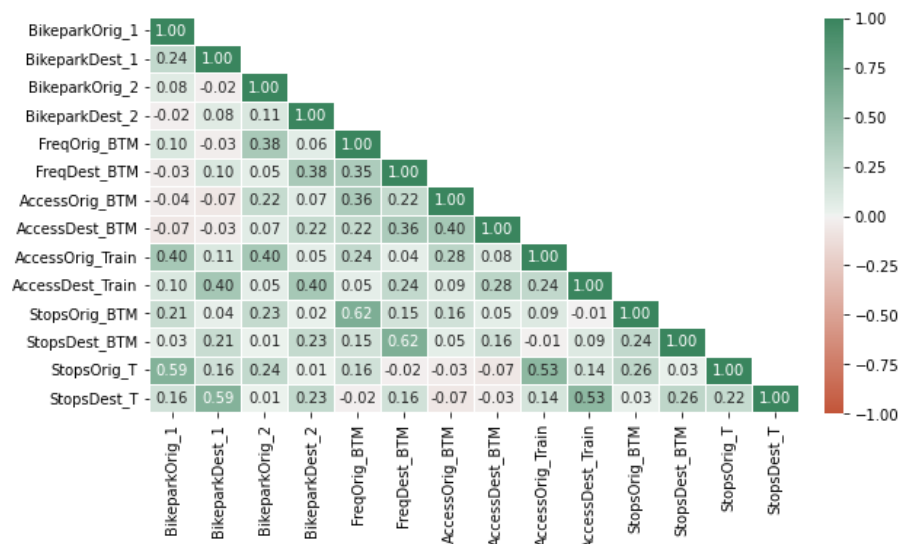


Figure 4.7: Correlation matrix train and BTM stops

The travel distance and the travel time of the modes are calculated with data from OSM and NDOV. These lead to the possibility to compare the results between modes. The bar plots show that an increase in travel distance or travel time for the bike and e-bike, does indeed lead to a lower share of (e-)bike choice compared to the other alternatives. The travel distance of the car and the bike are often quite similar, therefore the difference is

not that large between the two bar plots. It does show a lower share of car use on the long travel distances, but the short travel distances have mixed shares. The same applies to the travel time of the car. To calculate the travel distances and travel times of public transport, the walking times and transfer times are given a penalty. Shown travel distances and travel times thus include this penalty to better represent the options. The bar plots show that the train is chosen more often on longer travel distances between 10 and 20 kilometers and hardly on the short travel distances or very long travel distances. This is because on the short travel distance a train is hardly an option as the distance between two stations is often already quite large. The long travel distances in the plot are often caused by a large walking or transfer penalty, so they are also not a good option. The BTM modes have a similar plot, although these modes are most often chosen between 5 and 15 kilometers. The stations are closer together and thus the optimal distance for BTM is shorter than for the train. The average travel distances and travel times of the public transport modes in the table support this as well.

Table 4.3: Characteristics of added factors

Factor	Values	Share / Mean (std)	Factor	Share / Mean (std)
Street density bike		31.86 (8.61)	Frequency BTM Origin	55.24 (63.62)
Street connectivity bike		1.28 (0.06)	Frequency BTM Destination	55.22 (63.32)
Street density car		28.30 (7.30)	Car parking zone Origin	19.89 (35.06)
Street connectivity car		1.23 (0.05)	Car parking zone Destination	19.90 (35.04)
Street density PT		14.56 (12.09)	Calculated distance bike	3.84 (3.69)
Separate bicycle lanes		19.13 (5.63)	Calculated time bike	12.80 (12.31)
Bicycle parking Origin	Security & Paid	18.04%	Calculated time e-bike	11.64 (8.43)
	Security & Free	12.60%	Calculated distance car	4.07 (3.78)
Bicycle parking Destination	Security & Paid	18.15%	Calculated time car	4.28 (3.97)
	Security & Free	12.68%	Calculated distance train	18.76 (21.30)
Access train Origin		19.41 (25.86)	Calculated time train	87.40 (87.90)
Access train Destination		19.51 (25.92)	Calculated distance BTM	7.78 (5.60)
Access BTM Origin		78.53 (22.71)	Calculated time BTM	29.17 (15.46)
Access BTM Destination		78.53 (22.68)		

4.2.3. COMPARISON OF SPECIFIED AND CALCULATED FACTORS

To assess if the calculated travel distance and travel time are correctly calculated compared to the specified travel distance and travel time, box plots are shown in Figure 4.8 for the comparison of travel distance and Figure 4.9 for the comparison for travel time. The difference in travel distance shows that most of the time the calculated travel distance is estimated to be larger than the specified travel distance. However, the average and the quartiles are mostly around zero. Especially for the train, there are large outliers. These are caused by the calculation of the travel distance for the train, which adds two times the access and egress distance as-the-crow-flies to the travel distance taken by the train as a penalty for the distance between origin or destination and the train station. This also affects BTM in the same way, but is less extreme as can be seen in the plot. This is probably because BTM stops are more present in an area and thus have smaller walking distances.

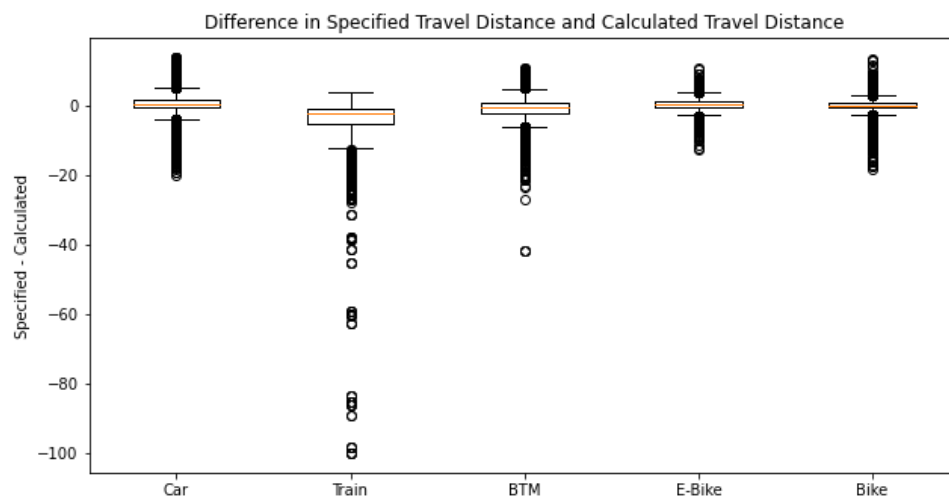


Figure 4.8: Difference in specified travel distance and calculated travel distance

The difference in travel time shows that the calculated travel time is actually smaller than the specified travel time. The outliers here are more extreme than for the distance, although the average and quartiles are again around zero. The extreme outliers are probably because the distance is filtered at a maximum of 15 kilometers, while the travel time is not filtered and can thus still be specified as anything. The outliers in the positive direction are in some cases caused by a misspecification of the travel time by respondents. The outliers in the negative direction of the train and BTM are the same reasoning as for the difference in travel distance, with the added limitation that no access or egress mode is taken into account which affects the travel time to and from a train or BTM station as well.

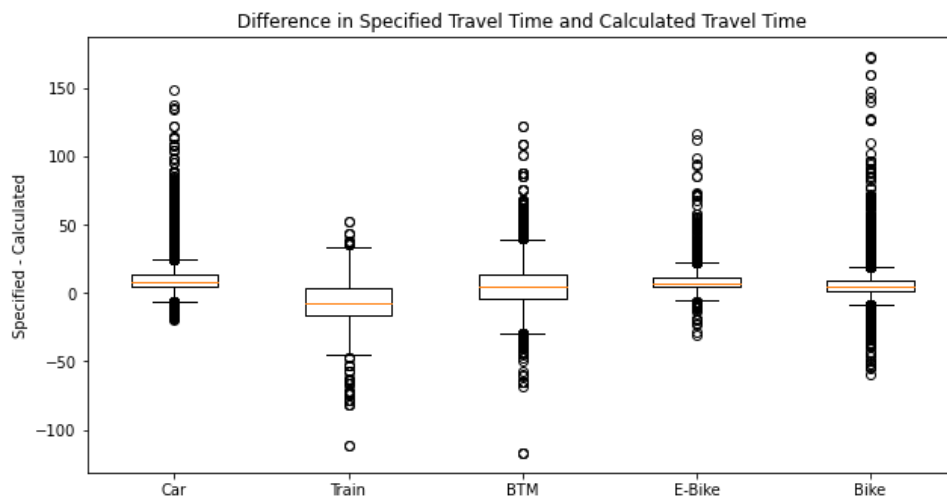


Figure 4.9: Difference in specified travel time and calculated travel time

In both box plots, the median is around zero and the quartiles are also close to the median. Only for the train and the BTM mode the quartiles are farther apart in both plots, and thus this needs to be taken into account when interpreting the results of the train and BTM in the model for the estimated parameters for travel distance and travel time. Moreover, the travel distance is more accurately calculated than the travel time as can be seen from the smaller quartile ranges for all modes. This also needs to be remembered when analysing results in the modelling phase.

4.3. CORRELATION

Besides analysing the individual factors, an analysis of the correlation between two factors and an analysis of factors with the mode alternatives is done. The first analysis aims to check whether two factors possibly explain the same phenomenon, which would then show a high correlation. It can then be argued if both factors need to be in the same model. The second analysis with the modes aims to check which variables are expected to influence the mode choice in the model. A high correlation would mean that it is expected to influence the mode choice and thus be significant. It would also show in which direction it is correlated, which would be an indication of the sign of the estimated parameters in the model.

Correlation between variables

From the correlation matrix in Figure 4.10, it can be seen that the highest positive correlations are between the calculated travel distances and travel times. The travel time of the bike and e-bike is calculated with an average speed of respectively 18 km/h and 25 km/h from the travel distance of the bike, so it is logical that these correlate to 1.0. The travel distance and travel time of the BTM modes also correlate with the travel distances and travel times of the bike and car. This is probably because the bus for example can drive on the same network as the car. The train travel distance and travel time are not correlated to the other calculated travel distances and travel times, probably because the walking penalties to stations and transfer penalties are higher and more often included. The train travel distance and travel time thus relate less to the other modes in which this is much less or not included. Furthermore, the street density of the car and the bike and separately the street connectivity of the car and the bike correlate positively. This is probably because most streets can be used for either the car or the bike, so these values can often be very similar.

The address density also correlates highly with car parking and access to BTM. This is because high address densities are mostly present in the city centres, which is also the location of most paid car parking zones and more stops for BTM modes. Moreover, the address density correlates negatively with the street density of the car and the bike. This is because an increase in address density means it is a larger city and thus has more infrastructure. Other highly positively correlated variables are the habit of the e-bike & the ownership of an e-bike and the habit of the car & if someone has a driver's license. It can mean in both cases that an increase in someone using the e-bike or the car leads to a higher chance of someone having an e-bike or a driver's license. The other way around is also possible: an increase in someone having a license or an e-bike leads to a higher chance of someone using the car or e-bike.

The most negatively correlated factors are the travel distance and calculated travel distances and travel times with the street density of the bike, car, and BTM. A higher street density then means a shorter distance, which is a logical result. The correlation is a bit stronger for the calculated travel distances. It can be because the street network is also calculated with distances. A notable result is that the public transport network does not correlate much with the travel distance of the public transport modes. The travel distance is also positively correlated with the calculated travel distances and travel times, which is a good result as these are meant to resemble each other. However, the correlation between the travel time and calculated travel times is less prominent. It could mean the calculated travel times are less accurate, but the travel times do vary more between the alternatives than the travel distance because of the travel speed of alternatives. It means the correlation is less, but it can still be accurately calculated.

Correlation between variable & modes

The correlation for each variable with each mode can be seen in Figure 4.11. It shows that for the car, the highest positive correlations are with the habit of the car, the driver's license, and the number of cars in a household. The address density, car parking zones, and the habit of the bike all show a negative correlation with choosing the car. For the train, no high correlations can be found. This can be because there are not many respondents choosing the train in the dataset. The highest correlations are with the travel distance and travel time, which means that the higher both variables, the more often the train is chosen. The BTM mode is the highest correlated with the habit of BTM, but also positively correlated with address density, car parking zones, and the frequency of BTM modes. It can therefore be seen that the BTM modes are most often chosen in the city or at places with high BTM lines frequency. The e-bike is mostly correlated with the habit of the e-bike and the ownership of the e-bike in the household. Moreover, the habit of the bike shows a negative correlation which would mean that an increased habit for the bike will lead to a lower probability of choosing the e-bike. Lastly, the bike is mostly positively correlated with the habit of the bike, and negatively correlated with the habit of the car. Often car use thus leads to a lower chance of choosing the bike. Moreover, the travel distance and travel time correlate negatively, so a longer trip will be less likely taken by bike. These findings are related and show similar results to the findings from the bar plots. A visual representation of these numbers can therefore be found in Appendix B.

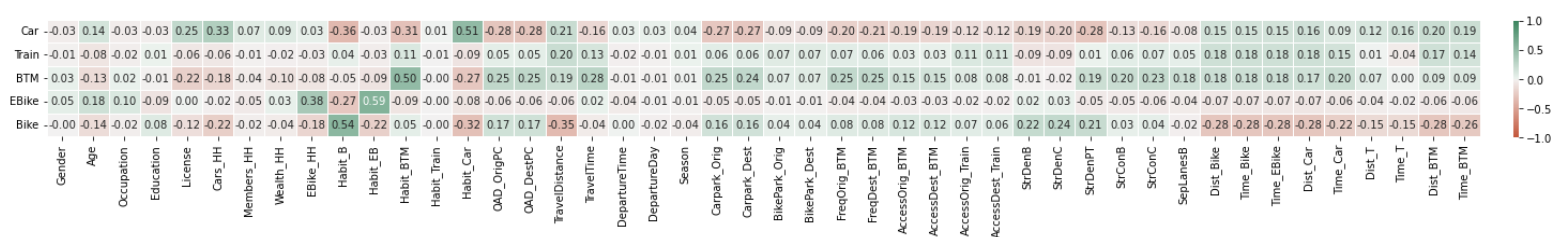


Figure 4.11: Correlation between variable & modes

4.4. DATA FOR VALIDATION

Now that the data is gathered completely, the section for validation can be split from the ODiN dataset with added factors. Respondents that had an origin or destination in Haarlem are separated, which leads to a sample size of 1783 respondents from the total of 118566 respondents. From this selection, the modal split can be seen in Figure 4.12.

In general, the modal split of the whole dataset in Figure 4.4 is similar to the modal split of Haarlem. The car and bike both have the largest amount of trips, and the BTM, e-bike, and train have much smaller shares. The difference with the modal split of the total dataset is that the bike has a 6 percentage points larger share, which is mostly at the cost of the share for the car. The shares of the e-bike, train, and BTM have remained a similar size. This means that generally in Haarlem, people tend to take the bike more than average in urban areas in the Netherlands. It is not expected that these differences result in an inaccurate validation, but it is something to remember when interpreting the validation results.

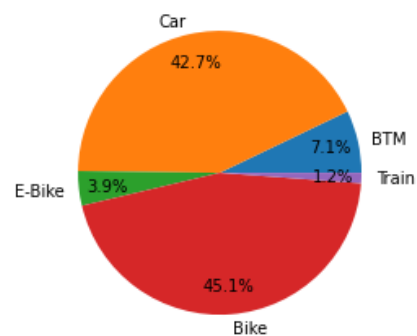


Figure 4.12: Modal split Haarlem

4.5. CONCLUSION

The data analysis aims to get an insight into the factors and how they relate to each other and to the mode alternatives. In this section, the most important expectations and findings based on the data analysis are summarized to be able to use in the modelling phase.

The calculated travel distance and travel time are included for each mode to replace the specified travel distance and travel time as specified by respondents in ODIN. The advantage of the calculated travel distance and travel time is that it has the data for all alternatives per respondent, instead of only for the chosen mode. The correlation matrix shows that indeed there is a correlation, which shows that the factors are similar. Moreover, the box plots showing the difference between the calculated and specified factors, show that most of the difference is close to zero. Therefore, the calculated travel distance and travel time can be used in the model replacing the travel distance and travel time given in ODIN.

Address density correlates highly with car parking zones, access to BTM modes, and frequency of BTM modes. Also, positive correlation can be found between the network variables street density and street connectivity. A high address density often means the area is within a city. All other mentioned factors are related to the city, as BTM modes are more present in cities (see Section 4.1.1), and street density and street connectivity are expected to be higher in cities as well. Moreover, car parking zones are mostly present in city centres to generate resistance to using the car there. It is therefore expected that address density will explain the same phenomenon as the combination of the other mentioned factors when used in a model.

Separate bicycle lanes show to be more chosen by public transport modes when increased in percentage. This is not a logical result as it was expected before the analysis to have a positive influence on the (e-)bike modes. However, in the correlation matrix can be seen that separate bicycle lanes also slightly correlate positively with the address density. It could therefore mean that a higher amount of separate bicycle lanes can be found in the cities, where also the public transport modes are present and chosen more often. It is therefore possible that the separate bicycle lanes will not show their influence the way it was intended in the model.

Bicycle parking is expected to have a positive influence on the public transport modes, and a negative influence on the (e-)bike modes. This conclusion is drawn based on the correlation in Figure 4.11 and the bar plots in Appendix B. It is therefore expected that it will not accurately describe the impact of bicycle parking facilities. An explanation of the positive impact on public transport modes can be that a lot of the secured and free bicycle parking spaces are at a train station or at a BTM stop. Access and egress will thus be done by bike, but then the main mode of the trip will be the train or BTM. Analysing if the number of train stations or BTM stops is possibly a factor that influences the mode choice, shows that it correlates with other factors like the frequency of BTM lines and the access of BTM and train stops. It is therefore not used further as a factor, as it is expected that the other factors already explain mostly the same phenomenon. These findings of the expectation of the behaviour of bicycle parking must thus be remembered.

5

MODELLING

This chapter presents the results of the modelling steps. First, the model requirements are shown, whereafter discrete choice models with their results are presented. The steps towards the final model are all performed by estimating MNL models. The first is to assess the influence of each variable, then the influence of each category of factors, and finally a complete MNL model with all significant and useful main variables. Hereafter the interaction effects and quadratic components are assessed to add to a MNL main variables model. A complete MNL model is then estimated, together with an NL and ML model. The model with the best fit with the data is then chosen as the final model. These steps towards the final model are taken to check the behaviour and significance of the variables (independently and with other variables) in the model and to build toward an accurate and useful final model. A sensitivity and uncertainty analysis are performed to see how changes in the input of the final model change the outcome of the model.

5.1. MODEL REQUIREMENTS

The output of the model is a modal split, which is directly related to the goal of the research of predicting a more accurate modal split. It should estimate a modal split in which the mode shares are estimated within 3 percentage points of the actual mode shares. By having this as an output, it presents the possibility to analyse the changes in modal split by adjusting the input of variables. This information is useful for the design of networks and policies for municipalities. It is also of importance that the input of the factors included in the model can be known by municipalities, and that the input does not change in a year for stable results.

The input of the model will be a dataset with respondents as rows and variables as columns. The distance range is already determined to be 15 kilometers maximum. To model any influences on the mode choice which are not captured in the defined variables, alternative specific constants are included. Moreover, each variable parameter is alternative specific to be able to analyse the influence of the variables for each alternative separately. In a logit model, a reference alternative needs to be specified and this is the car alternative. By using this as the reference, the values calculated of the other modes are in reference to the car. The same accounts for categorical variables in which a reference category is required. The reference is always the first category of each categorical variable. Every estimated value of a categorical variable is thus in reference to the car alternative and to the first category of that variable. An exception to having a reference alternative is when the data of the variable is different for each alternative. The travel distance is an example of this, where for each alternative the distance between origin and destination is different. In this case, no reference is required as the data is only used for one mode.

5.2. DISCRETE CHOICE MODELS

This section shows the results of the discrete choice models that are estimated. First, the variables are assessed on their influence on their model fit by only including the variable with the ASC. The aim is to understand which variables have what size of fit with the data. A higher log likelihood or a higher rho-square-bar means a better model fit. After this, MNL models are estimated for each category and for all variables. The significance of the variables can be checked by looking at the p-value. If this is below 0.05, the variable is significant. It is also assessed if the variables are useful for the goal of the research. It is important that the input of the variables is known information to municipalities, and that the input values do not change yearly. After having modelled the variables independently, interaction effects and quadratic components of variables are added based on findings in literature and own insights. When these have been tested on their significance, the final selection of variables is estimated as a Multinomial Logit, Nested Logit, and Mixed Logit. This is done to compare the results between these types of logit models and to choose the model with the best fit as the final model.

5.2.1. INFLUENCE OF EACH VARIABLE

In this subsection, the results of the models are shown in which each result is a model that contains only the variable and the alternative specific constants. The null log likelihood is shown, which is a log likelihood of the model when there are no variables or predictors that explain the data. In Table 5.1 is shown what the log likelihood and the rho-square-bar are for each variable. To assess if the change between two models is significant, the likelihood ratio test is performed. It is used when the first model can be obtained by constraining parameters from the second model. As each model is the same as the ASC model with the added variable, the likelihood ratio test can indeed be used. It shows that for each variable result, the change in log likelihood is significant. This means that in all cases the model with the added variable is a better model fit.

Table 5.1: Influence per variable

	<i>Log likelihood</i>	<i>Rho-square-bar</i>	<i>No. of parameters</i>	<i>Likelihood Ratio test</i>
Null log likelihood	-163603.3	0.000	0	0.00
ASC	-109827.0	0.329	4	0.00
Spatial characteristics				
Density of network	-105811.1	0.353	9	0.00
Connectivity of network	-108451.8	0.337	7	0.00
Separate bicycle lanes in network	-108020.0	0.340	8	0.00
Access to public transport	-107401.3	0.343	8	0.00
Frequency of public transport	-105773.2	0.353	12	0.00
Access to bicycle parking	-108331.8	0.338	12	0.00
Car parking zones	-105126.0	0.357	12	0.00
Season	-109563.0	0.330	16	0.00
People's characteristics				
Gender	-109580.1	0.330	8	0.00
Age	-107838.9	0.341	16	0.00
Occupation	-107209.0	0.345	24	0.00
Education	-109275.4	0.332	16	0.00
Driver's License	-107757.3	0.341	8	0.00
Wealth	-109325.3	0.332	20	0.00
No. of cars per driving person	-109013.4	0.334	8	0.00
E-Bike availability	-107829.5	0.341	8	0.00
No. of household members	-109263.6	0.332	8	0.00
Address density	-105082.7	0.358	12	0.00
Habit	-73322.2	0.552	24	0.00
Travel characteristics				
Calculated travel distance	-101864.3	0.377	9	0.00
Calculated travel time	-101482.5	0.380	9	0.00
Departure time	-109218.2	0.332	24	0.00
Departure day	-109284.3	0.332	28	0.00

Table 5.1 shows that the habit has by far the largest influence on mode choice. This thus indicates that the habit of people mostly depends on their frequent mode choices. After the habit, travel distance and travel time have the highest influence, as is also expected. The variables with the least amount of fit to the data are within the people's characteristics: gender, wealth, education, and household members. Also within the travel characteristics, the departure time and departure day are low model fits. Finally, the spatial characteristic variables show to have individually good model fits, except for the season.

5.2.2. INFLUENCE OF EACH CATEGORY & ALL VARIABLES

To know if the variables are significant or not and to check for correlation between variables, MNL models are first estimated for each category of variables. These models are checked on their model fit and the variables are checked to see if they are significant and if the results are as expected. Furthermore, there is also looked at the behaviour between variables. The correlations are checked to see if each variable explains an independent phenomenon in the model. After that, an MNL model with all remaining variables is estimated to further check correlations between variables.

MNL - Spatial characteristics

The spatial characteristics in Table 5.2 show that the bicycle parking does indeed show a positive influence on public transport modes, and no significant values for the bike and e-bike. Furthermore, the separate bicycle lanes also show this relation in which the public transport modes have a positive influence and the bike and e-bike negative. These results show that the expectations for these variables explained in the conclusion of the data analysis are found to be true in this model, and thus these two factors are excluded from further modelling. The access and frequency of public transport have the expected signs and significance for all variables, just like the influence of car parking relative to the car choice. Street density also shows to increase choice probabilities at higher densities, except for the train. It can be linked to the reason why connectivity was not included for public transport modes. When many train tracks are running through the travelled area (density of the network), but there is no possibility to get on, the train will most likely not be chosen. This problem is also present for the BTM mode but in a lesser state. This is because the stops are closer together. When tracks or roads where BTM modes are available are then within the travelled area, it is much more possible to also be able to choose that mode. Therefore, the street density is not modelled further for the train, but it is still included for the BTM mode. The street connectivity shows negative signs for each estimated mode alternative, which means that an increase in connectivity leads to a lower choice probability of the mode. The bar plots from the data analysis were also not conclusive on its effect on mode choice, which again shows to not be the expected result. The reason can be that the calculation of the connectivity is not chosen correctly for street networks, or it explains another phenomenon than expected. The calculation uses the number of edges divided by the number of nodes, which would mean that a large number of edges and a small number of nodes results in a high connectivity. Logically thinking, this would not be perceived as being a highly connected street network. To not make false conclusions about the influence of street connectivity, this factor is also excluded from further modelling. Finally, the season did not show much fit to the data individually, which can explain that not for every mode and every season the variable is significant. For the bike, it does show that there is a negative association with choosing the bike in the months with generally more bad weather.

Table 5.2: MNL Spatial Characteristics

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	2.590**	2.290**	ref: 1.0	-8.950**	-8.270**
Access to public transport (Origin)	-	-	-	1.530**	2.270**
Access to public transport (Destination)	-	-	-	1.570**	2.520**
Car parking zone (Origin)	0.850**	0.628**	ref	0.753**	0.477**
Car parking zone (Destination)	0.816**	0.592**	ref	0.656**	0.360**
Frequency of public transport (Origin)	0.138**	0.054	ref	0.348**	0.384**
Frequency of public transport (Destination)	0.172**	0.053	ref	0.364**	0.334**
Separate bicycle lanes	-4.340**	-2.930**	ref	5.940**	3.890**
Bicycle parking (Origin): Security & Paid	ref	ref	ref	ref	ref
Bicycle parking (Origin): Security & Free	-0.008	0.015	ref	0.341**	0.212**
Bicycle parking (Destination): Security & Paid	ref	ref	ref	ref	ref
Bicycle parking (Destination): Security & Free	-0.026	0.000	ref	0.395**	0.213**
Street connectivity	-6.870**	-6.200**	-3.900**	-	-
Street density	12.400**	10.900**	6.710**	0.492**	-7.490**
Season: Spring	ref	ref	ref	ref	ref
Season: Summer	0.116**	0.334**	ref	-0.150**	-0.157*
Season: Autumn	-0.066**	-0.003	ref	-0.089*	0.040
Season: Winter	-0.217**	-0.288**	ref	-0.073*	0.067
Log likelihood: -97595.9 Rho-square-bar: 0.403					
ref = reference alternative / category	red ≠ significant * = $p < 0.05$ ** = $p < 0.01$				

MNL - People's characteristics

The people's characteristics in Table 5.3 show that many variables have modes or categories that are not significant. This is also expected from the individual assessment of the variables in which some variables were not contributing much to a better model fit. Education and occupation are not shown to have significance for many modes, as well as the gender. Moreover, not all categories of the habit are shown to be significant although the model fit of the data has shown it would be very explanatory. In general, the values of the parameters increase when the habit is stronger, so the probability of choosing that mode then increases. Furthermore, not all variables are having the signs that are expected. E-bike ownership shows a negative relation to choosing the e-bike. This can be caused by the availability conditions in Biogeme, for which the e-bike ownership is used. This leads to a large correlation with the ASC of the e-bike. The same phenomenon is then explained by two variables, which leads to the model compensating the results with each other. It has thus been determined to not further include the e-bike ownership. The address density shows a positive relation, which means that a higher address density leads to an increased choice for all modes compared to the car. However, because of the correlations it has with variables of the spatial characteristics, it is not included in the full model.

Table 5.3: MNL People's Characteristics

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	-2.730**	10.800**	ref: 1.0	-4.670**	-0.743**
Address density (Origin)	0.143**	0.093**	ref	0.241**	0.170**
Address density (Destination)	0.145**	0.103**	ref	0.248**	0.169**
Age: 0-17	ref	ref	ref	ref	ref
Age: 18-40	-0.631**	-1.230**	ref	-0.787**	-0.838**
Age: 41-66	-0.596**	-1.140**	ref	-0.777**	-1.420**
Age: 67-100	-0.728**	-1.070**	ref	-0.827**	-2.640**
Cars per driver's license	0.359**	0.244**	ref	0.371**	-0.221**
E-bike ownership	-0.138**	-11.500**	ref	-0.064	-0.045
Education: primary education	ref	ref	ref	ref	ref
Education: vmbo/mavo	0.041	-0.179	ref	-0.067	-0.018
Education: havo/vwo	0.059	-0.314**	ref	-0.119	-0.026
Education: hbo/university	0.065	-0.238*	ref	-0.266**	0.110
Gender	-0.154**	0.045	ref	0.000	-0.356**
Habit: never	ref	ref	ref	ref	ref
Habit: every year	0.019	-1.040**	0.149*	1.300**	-0.21
Habit: every month	1.550**	-0.137	0.151**	3.100**	-0.550**
Habit: every week	3.000**	1.480**	0.933**	4.620**	-0.400**
Habit: (almost) every day	4.100**	2.740**	1.830**	5.590**	-0.309**
Driver's license	0.241**	0.505**	ref	0.067	-0.308**
Household members	0.058**	-0.005	ref	0.073**	0.020
Occupation: part-time	ref	ref	ref	ref	ref
Occupation: full-time	-0.047	-0.064	ref	-0.016	-0.126
Occupation: own housekeeping	-0.085	-0.041	ref	0.089	-1.490**
Occupation: student	-0.052	-0.553**	ref	-0.122*	0.210*
Occupation: unemployed / unable to work	-0.199**	-0.027	ref	0.131	-1.710**
Occupation: retired	-0.059	-0.227**	ref	0.069	-0.84**8
Wealth: first 20% group	ref	ref	ref	ref	ref
Wealth: second 20% group	0.300**	0.380**	ref	0.266**	0.363**
Wealth: third 20% group	0.445**	0.508**	ref	0.507**	0.313**
Wealth: fourth 20% group	0.461**	0.661**	ref	0.431**	0.742**
Wealth: fifth 20% group	0.520**	0.600**	ref	0.391**	0.741**
Log likelihood: -69268.0 Rho-square-bar: 0.576					
ref = reference alternative / category red ≠ significant * = $p < 0.05$ ** = $p < 0.01$					

MNL - Travel characteristics

The travel characteristics in Table 5.4 show that the calculated travel distance and travel time are negative for most modes. The results of these variables have to be interpreted lightly because they correlate highly. The unusual positive values for travel distance for BTM and the train can thus be compensation for the travel time. They both show to be significant and in the individual assessment of the fit to the data, they were both good fits compared to the other variables. However, because of the correlation, one of the two should be chosen to include in modelling further. The fit to the data of the travel time shows to be a bit better, but the data of the calculated travel distance is more reliable as was concluded in Section 4.2.3. Therefore, the calculated travel time will be excluded from further modelling. The departure day shows that every mode is more preferably chosen on weekdays than on weekends. Travelling with any mode on a Saturday has even shown to not be significant. The departure time shows that all modes are more likely to be chosen during the day compared to the middle of the night. The highest preference of all modes is to travel between 09:00-12:00.

Table 5.4: MNL Travel Characteristics

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	0.579**	0.065	ref: 1.0	-2.477**	-3.324**
Calculated travel distance	-0.009**	-0.017**	-0.172**	-0.149**	0.040**
Calculated travel time	-0.141**	-0.166**	-0.071**	0.011**	-0.047**
Departure day: Sunday	ref	ref	ref	ref	ref
Departure day: Monday	0.360**	0.443**	ref	0.482**	1.001**
Departure day: Tuesday	0.433**	0.522**	ref	0.558**	0.908**
Departure day: Wednesday	0.322**	0.467**	ref	0.446**	0.774**
Departure day: Thursday	0.338**	0.323**	ref	0.544**	0.863**
Departure day: Friday	0.123**	0.193**	ref	0.409**	0.474**
Departure day: Saturday	-0.010	-0.052	ref	-0.038	0.119
Departure time: 00:00-07:00	ref	ref	ref	ref	ref
Departure time: 07:00-09:00	0.881**	1.384**	ref	0.599**	0.437*
Departure time: 09:00-12:00	1.360**	1.891**	ref	1.530**	1.682**
Departure time: 12:00-16:00	0.645**	1.288**	ref	0.948**	0.355**
Departure time: 16:00-19:00	0.530**	1.168**	ref	0.993**	0.525**
Departure time: 19:00-24:00	0.995**	1.433**	ref	1.320**	1.160**
Log likelihood: -99606.0	Rho-square-bar: 0.391				
ref = reference alternative / category		red ≠ significant		* = $p < 0.05$	** = $p < 0.01$

MNL - All main variables

Some variables need to be excluded because they are not useful for the goal of the research. The goal is to find a model that can be used to predict modal splits between O-D pairs in urban areas. Data needs to be present for the variables that are in the model to be able to predict a modal split that is constant for a longer period of time. Variables that change throughout a year can therefore not be included. Moreover, socio-demographic variables can be included as this is often known for an area. However, the habit of people is often not generally known in an area and it is thus also excluded. Another reason to exclude the habit is the way the data is gathered. ODiN asks respondents to report the trips made on a recent day. The chance that a respondent picks a day and reports a trip made with a mode that is said to be used (almost) every day, is very large. Therefore, it can be possible that the habit is an even better fit with the model than what could be true. Another variable that is excluded, is the address density. This is based on the findings in the data analysis, which show that the variable has large correlations with multiple spatial variables: car parking, access to BTM, and street density. The correlation shows that it describes a similar phenomenon. The spatial variables actually describe which phenomenon it is, while the address density is vague. Therefore, the address density will not be included in further modelling. The variables that are being excluded thus are departure day, departure time, habit, season, and address density.

Table 5.5: MNL all main variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	1.450**	2.160**	ref: 1.0	-4.200**	-3.610**
Access to public transport (Origin)	-	-	-	1.290**	2.080**
Access to public transport (Destination)	-	-	-	1.410**	2.310**
Age: 0-17	ref	ref	ref	ref	ref
Age: 18-40	-1.180**	-0.786**	ref	-0.762**	-0.827**
Age: 41-66	-1.110**	-0.507*	ref	-0.961**	-1.290**
Age: 67-100	-1.560**	-0.739**	ref	-0.696**	-2.700**
Cars per driver's license	-0.565**	-0.611**	ref	-0.555**	-0.594**
Travel distance	-0.388**	-0.315**	-0.165**	-0.022**	0.000
Car parking (Origin)	0.841**	0.696**	ref	1.220**	0.551**
Car parking (Destination)	0.731**	0.618**	ref	1.150**	0.465**
Education: primary education	ref	ref	ref	ref	ref
Education: vmbo/mavo	0.034	-0.231**	ref	0.531**	0.140
Education: havo/vwo	0.141*	-0.418**	ref	0.456**	0.183
Education: hbo/university	0.476**	-0.303**	ref	0.227*	0.252
Frequency of BTM stops (Origin)	0.118**	0.090*	ref	0.477**	0.402**
Frequency of BTM stops (Destination)	0.176**	0.097**	ref	0.451**	0.304**
Gender	-0.096**	0.436**	ref	0.184**	-0.185**
Driver's license	-0.807**	-0.641**	ref	-1.540**	-1.310**
Household members	0.055**	-0.089**	ref	0.096**	0.031
Occupation: part-time job	ref	ref	ref	ref	ref
Occupation: full-time job	-0.250**	-0.256**	ref	-0.131**	-0.230**
Occupation: own housekeeping	-0.153**	0.041	ref	0.069	-0.945**
Occupation: student	0.753**	-0.821**	ref	1.040**	0.863**
Occupation: unemployed / Unable to work	-0.491**	-0.168*	ref	-0.033	-1.320**
Street density	2.100**	0.840**	-0.015	0.473**	-
Wealth: first 20% group	ref	ref	ref	ref	ref
Wealth: second 20% group	0.239**	0.212**	ref	0.249**	0.313**
Wealth: third 20% group	0.385**	0.292**	ref	0.564**	0.311**
Wealth: fourth 20% group	0.463**	0.423**	ref	0.434**	0.661**
Wealth: fifth 20% group	0.640**	0.318**	ref	0.321**	0.640**
Log likelihood: -89189.7 Rho-square-bar: 0.454					
ref = reference alternative / category					
red ≠ significant * = $p < 0.05$ ** = $p < 0.01$					

Table 5.5 shows the results of the complete set of variables that are left. The values are similar compared to the models per category. The access and frequency of public transport are shown to influence the public transport modes positively. It even stimulates the choice for bikes as well. The age groups show negative relations to the mode choice compared to the car and the first age group. However, these parameters correlate with the ASC's. The negative results can therefore be a compensation for the ASC of the modes. The number of cars per driver's license shows that an increase in the number of cars decreases the choice of all other modes as expected. Having a driver's license also decreases the use of all other modes compared to the car. Travel distance is also shown to negatively influence the mode choice, except for the train. This is possibly because of the little variation in the distance for the train in the dataset. Only distances below 15 kilometers are included, and short train trips are often already around this boundary. Having to pay for parking, increases the choice of other modes as expected. Education level is not significant for the train choice. Low-educated respondents more often choose BTM, while high educated respondents prefer the bike. Men more often take the bike and train, while women prefer the e-bike and BTM. Households with many members prefer BTM, and often do not choose the e-bike. People who are working prefer to take the car, while students more often take public transport or the bike. The street density does not show to be significant for the car, but it positively influences the (e-)bike and BTM. Finally, low-wealth households generally use all modes equally (as long as it is available), while high-wealth households prefer the bike and train.

5.2.3. INFLUENCE OF INTERACTION EFFECTS AND QUADRATIC COMPONENTS

This subsection is explaining the use of interaction effects and quadratic components to further increase the accuracy of the model. The interaction effects are estimated in a model together with the ASC and the main variables to check their significance and influence. The interaction effects and quadratic components that are tested are chosen based on the literature review and the goals of the research. More detailed results and figures can be found in Appendix E.

Interaction effects

Interaction effects were found in literature and summarized in Chapter 3, shown in Figure 3.6. These are going to be tested if relevant to the goal of the research. Of the variables that remain, the interaction that can be tested based on literature is gender with travel distance. It has been found by Heinen *et al.* (2012) that males have a smaller resistance against cycling longer distances. Other interaction effects which are expected are the age with the travel distance, where it is expected that the elderly have more resistance to cycling longer distances. Also car parking zones with the wealth of a household, where it is expected that lower wealth households have more resistance against taking the car to a paid parking zone. Lastly, car parking zones with the residential zip code, where it is expected that people living in a paid parking zone could have an exemption for paid parking. Another term that is going to be assessed, is the variable travel speed. It is calculated by dividing the calculated travel distance by the calculated travel time. Interaction effects normally consist of two main variables that are multiplied by each other. The travel speed also consists of two variables, but the difference is that two variables are divided instead of multiplied, and not both main effects of the variables are included in the model.

The interaction effects are tested by comparing the results from two models. One models the interaction effect and the main variables, and the other models only the main variables to check the difference in the result. Not all interaction effects are assessed by including the main effects of both variables. In the case of assessing the impact of car parking zones with residential zip code, only the main effect of the car parking zones are included. Interaction effects are included in utility functions by the following formula:

$$Utility_{interaction\ effect} = B_{variable12_mode} * variable1 * variable2 + B_{variable1_mode} * variable1 (+B_{variable2_mode} * variable2) \quad (5.1)$$

A likelihood ratio test is performed to check if the interaction leads to a better model fit. The results of the interactions can be found in Table 5.6. Distance & gender is tested for all modes and has been found to be significant for the bike and e-bike. As expected, the females have a larger resistance against cycling longer distances. The results thus match with the findings of Heinen *et al.* (2012). For the interaction of distance & age it is expected that the elderly have a larger resistance to cycling. It has indeed been found to be significant for the bike and e-bike. Speed is included as a division of the travel distance and travel time. Although it is a different formulation, it has been found to be significant for all included modes. It is not included for the bike & e-bike, because the travel times were based on a travel speed of 18 km/h and 25 km/h. Including the travel speed will thus provide no variation in the data and will lead to no useful estimation results. Car parking in combination with wealth is tested only for the car, but it was not found to be significant. Finally, the car parking with the residential zip code is tested for all modes and is also significant for all modes. The detailed results of the models can be found in Appendix Tables E.3, E.4, E.5, E.6, E.7.

Table 5.6: Interaction effects

	<i>Main effects</i>		<i>Main effects + interaction</i>		LR-test	Significance
	LL	Rho-square-bar	LL	Rho-square-bar		
Distance & Gender	-101623.7	0.379	-101484	0.380	0.00**	Bike & E-Bike
Distance & Age	-99868.4	0.389	-99717.12	0.390	0.00**	Bike & E-Bike
Distance & Speed	-101623.7	0.379	-99727.02	0.390	0.00**	Bike, Car, BTM, Train
Car parking & Wealth	-104727.4	0.360	-104717.3	0.360	0.01*	None
Car parking & Residential zip code	-105126.0	0.357	-104395.2	0.362	0.00**	All

* = $p < 0.05$ ** = $p < 0.01$

Quadratic components

Quadratic components are used in modelling to better fit the data with variables that are possibly not linearly described. The utility of the variable then decreases or increases quadratically for an increase in variable size. The quadratic components that are analysed are travel distance, car parking zones, and number of cars per driver's license. The results show a comparison of estimated values for a continuous, categorical, and quadratic relation. The categorical relation is also included to be able to check if the estimated parameter indeed shows a decreasing or increasing parameter value and to check if the quadratic component then is better than defining the variable as continuous or categorical. The categories are determined based on the bar plot analysis in Chapter 4. The quadratic components for a variable are calculated by using the following formula:

$$Utility_{quadratic\ component} = B_{variable_mode}^2 * variable^2 + B_{variable_mode} * variable \quad (5.2)$$

In the first results of the MNL models, the calculated distance does show a negative influence. However, the public transport modes have less influence than the other modes. The reason can be because the train choice is not much present in the dataset, because of the limit of 15 kilometers on travel distance. The train is mostly a choice for the longer distances as could be seen in Section 4.1.1. There could also be argued that modelling the travel distance as a continuous, linear variable is incorrect. It can be stated that the resistance to a mode is different for other distance categories. Based on the CDF curve of all modes in Figure 4.3, distance categories are made and modelled to see if the parameter values change significantly between categories of a mode. The categories are mostly based on the curves of the bike and the e-bike because it is expected that the resistance of travel distance is mostly seen for these modes. The categories made are from 0-2.5 kilometers, which includes about 40% of all e-bicycle trips and 55% of all bicycle trips. Then from 2.5-5 kilometers, which includes about 20% of e-bicycle trips and 30% of all bicycle trips. From 5-10 kilometers, which respectively includes about 20% and 10% of e-bicycle and bicycle trips, and about 50% of the BTM and car mode trips. Finally, 10+ kilometers which include the remaining (e-)bicycle trips until the made boundary of 15 kilometers for the specified travel distance. The calculated travel distances can be higher, and thus all above 15 kilometers are also within the last category.

Table 5.7: Influence of travel distance

Continuous		Categorical		Quadratic Component	
Log likelihood	-101864.3	Log likelihood (<i>LR-test</i>)	-101733.2 (0.00**)	Log likelihood (<i>LR-test</i>)	-101348.2 (0.00**)
Rho-square-bar	0.377	Rho-square-bar	0.378	Rho-square-bar	0.380
ASC_Bike	0.888**	ASC_Bike	0.738**	ASC_Bike	0.914**
ASC_BTM	-1.620**	ASC_BTM	-2.470**	ASC_BTM	-2.280**
ASC_E-Bike	0.939**	ASC_E-Bike	0.864**	ASC_E-Bike	1.020**
ASC_Train	-3.690**	ASC_Train	-5.350**	ASC_Train	-3.820**
B_CDIST_BTM	-0.093**	B_CDIST_BTM: 0-2.5 km	ref	B_CDIST_BTM	0.078**
		B_CDIST_BTM: 2.5-5 km	1.150**	B_CDIST_BTM ²	-0.009**
		B_CDIST_BTM: 5-10 km	1.200**		
		B_CDIST_BTM: 10+ km	0.752**		
B_CDIST_B	-0.429**	B_CDIST_B: 0-2.5 km	ref	B_CDIST_B	-0.515**
		B_CDIST_B: 2.5-5 km	-0.731**	B_CDIST_B ²	0.009
		B_CDIST_B: 5-10 km	-2.040**		
		B_CDIST_B: 10+ km	-3.210**		
B_CDIST_C	-0.201**	B_CDIST_C: 0-2.5 km	ref	B_CDIST_C	-0.248**
		B_CDIST_C: 2.5-5 km	-0.060**	B_CDIST_C ²	0.004**
		B_CDIST_C: 5-10 km	-0.509**		
		B_CDIST_C: 10+ km	-0.944**		
B_CDIST_EB	-0.370**	B_CDIST_EB: 0-2.5 km	ref	B_CDIST_EB	-0.492**
		B_CDIST_EB: 2.5-5 km	-0.696**	B_CDIST_EB ²	0.011**
		B_CDIST_EB: 5-10 km	-1.660**		
		B_CDIST_EB: 10+ km	-2.600**		
B_CDIST_T	-0.023**	B_CDIST_T: 0-2.5 km	ref	B_CDIST_T	-0.018**
		B_CDIST_T: 2.5-5 km	0.030	B_CDIST_T ²	0.000
		B_CDIST_T: 5-10 km	1.920**		
		B_CDIST_T: 10+ km	2.460**		

ref = reference alternative / category

red ≠ significant

* = $p < 0.05$ ** = $p < 0.01$

The results in Table 5.7 are that modelling as a categorical variable indeed increases the fit with the data in terms of the log likelihood and the rho-square-bar. Moreover, the car and (e-)bike show an increasingly negative value for longer travel distances. The public transport modes have positive values compared to the shortest travel distance category. This means that the modes are more likely chosen for the longer travel distances as was also found in Section 4.2.2. The increasingly negative values for three of the mode alternatives indicate that the resistance is not linear, but could be quadratic. Therefore, a quadratic component is tested for the calculated travel distance. It can be seen from the results that it is indeed significant for all but the train. Also, the model fit is better than categorical, thus a quadratic component is added for the travel distance. A plot of the relationship that results from these estimated parameters can be found in Appendix Figure E.1.

Table 5.8: Influence of no. of cars per driver's license

Continuous		Categorical		Quadratic Component	
Log likelihood	-109013.4	Log likelihood (<i>LR-test</i>)	-106479 (0.00**)	Log likelihood (<i>LR-test</i>)	-108752.6 (0.00**)
Rho-square-bar	0.334	Rho-square-bar	0.349	Rho-square-bar	0.335
ASC_Bike	0.604**	ASC_Bike	-7.900**	ASC_Bike	0.231**
ASC_EBike	0.864**	ASC_EBike	-7.260**	ASC_EBike	0.681**
ASC_BTM	-0.494**	ASC_BTM	-8.840**	ASC_BTM	-0.773**
ASC_Train	-2.210**	ASC_Train	-10.700**	ASC_Train	-2.600**
B_CAR_B	-0.464**	B_CAR_B: no cars per license	<i>ref</i>	B_CAR_B	0.683**
		B_CAR_B: < 0.5 cars per license	8.640**	B_CAR_B ²	-0.709**
		B_CAR_B: 0.5 - 1 cars per license	8.050**		
		B_CAR_B: > 1 cars per license	7.880**		
B_CAR_EB	-0.631**	B_CAR_EB: no cars per license	<i>ref</i>	B_CAR_EB	-0.287**
		B_CAR_EB: < 0.5 cars per license	8.010**	B_CAR_EB ²	-0.131*
		B_CAR_EB: 0.5 - 1 cars per license	7.100**		
		B_CAR_EB: > 1 cars per license	7.420**		
B_CAR_C	<i>ref</i>	B_CAR_C	<i>ref</i>	B_CAR_C	<i>ref</i>
				B_CAR_C ²	<i>ref</i>
B_CAR_BTM	-1.100**	B_CAR_BTM: no cars per license	<i>ref</i>	B_CAR_BTM	-0.788**
		B_CAR_BTM: < 0.5 cars per license	7.900**	B_CAR_BTM ²	0.018
		B_CAR_BTM: 0.5 - 1 cars per license	7.570**		
		B_CAR_BTM: > 1 cars per license	7.170**		
B_CAR_T	-0.856**	B_CAR_T: no cars per license	<i>ref</i>	B_CAR_T	0.625**
		B_CAR_T: < 0.5 cars per license	8.350**	B_CAR_T ²	-1.050**
		B_CAR_T: 0.5 - 1 cars per license	8.020**		
		B_CAR_T: > 1 cars per license	7.440**		

ref = reference alternative / category

red ≠ significant

* = $p < 0.05$ ** = $p < 0.01$

The categorical and quadratic effect of the number of cars per driver's license in a household is shown in Table 5.8. It is expected that the other modes are being chosen less when there are more cars available. The categories for modelling as a categorical variable range from having no cars, having to share cars to always having a car available. These show to be a better fit with the data than a continuous variable, and the parameter estimates show a value that becomes less positive when more cars are available for most mode alternatives. The positive main effect and negative quadratic component for the bike and train show a more negative utility when more cars are available. The negative main effect and negative quadratic component for the e-bike show that the utility is always negative and the change in utility for larger values is larger than for smaller values of the number of cars. Only for the BTM mode, the quadratic component is not significant. A plot of the relationship that results from these estimated parameters can be found in Appendix Figure E.3. The model fit of the quadratic component is lower than for the categorical variable, thus the categorical variable would be preferred. However, this variable is also used in the availability conditions of Biogeme in which the presence of a car means it is available. Modelling it as a categorical variable then leads to the same correlations as could be seen for the e-bike. Therefore, it is not modelled as a categorical variable but as a quadratic component.

Table 5.9: Influence of car parking zones

Continuous		Categorical		Quadratic Component	
Log likelihood	-105126.0	Log likelihood (<i>LR-test</i>)	-105072.6 (0.00**)	Log likelihood (<i>LR-test</i>)	-105038.9 (0.00**)
Rho-square-bar	0.357	Rho-square-bar	0.358	Rho-square-bar	0.358
ASC_Bike	-0.051**	ASC_Bike	-0.135**	ASC_Bike	-0.072**
ASC_EBike	0.214**	ASC_EBike	0.131**	ASC_EBike	-2.040**
ASC_BTM	-2.030**	ASC_BTM	-2.150**	ASC_BTM	0.182**
ASC_Train	-3.290**	ASC_Train	-3.500**	ASC_Train	-3.460**
<i>Origin</i>					
B_CPARK_B	0.890**	B_CPARK_B: No paid parking	<i>ref</i>	B_CPARK_B	1.360**
		B_CPARK_B: < 50% paid parking	0.231**	B_CPARK_B ²	-0.515**
		B_CPARK_B: > 50% paid parking	0.728**		
		B_CPARK_B: 100% paid parking	1.100**		
B_CPARK_EB	0.656**	B_CPARK_EB: No paid parking	<i>ref</i>	B_CPARK_EB	1.440**
		B_CPARK_EB: < 50% paid parking	0.206**	B_CPARK_EB ²	-0.866**
		B_CPARK_EB: > 50% paid parking	0.562**		
		B_CPARK_EB: 100% paid parking	0.718**		
B_CPARK_C	<i>ref</i>	B_CPARK_C	<i>ref</i>	B_CPARK_C	<i>ref</i>
				B_CPARK_C ²	<i>ref</i>
B_CPARK_BTM	1.670**	B_CPARK_BTM: No paid parking	<i>ref</i>	B_CPARK_BTM	2.070**
		B_CPARK_BTM: < 50% paid parking	0.418**	B_CPARK_BTM ²	-0.445*
		B_CPARK_BTM: > 50% paid parking	1.350**		
		B_CPARK_BTM: 100% paid parking	1.980**		
B_CPARK_T	1.260**	B_CPARK_T: No paid parking	<i>ref</i>	B_CPARK_T	3.420**
		B_CPARK_T: < 50% paid parking	0.508**	B_CPARK_T ²	-2.250**
		B_CPARK_T: > 50% paid parking	1.220**		
		B_CPARK_T: 100% paid parking	1.320**		
<i>Destination</i>					
B_CPARK_B	0.874**	B_CPARK_B: No paid parking	<i>ref</i>	B_CPARK_B	1.340**
		B_CPARK_B: < 50% paid parking	0.234**	B_CPARK_B ²	-0.512**
		B_CPARK_B: > 50% paid parking	0.711**		
		B_CPARK_B: 100% paid parking	1.110**		
B_CPARK_EB	0.620**	B_CPARK_EB: No paid parking	<i>ref</i>	B_CPARK_EB	1.370**
		B_CPARK_EB: < 50% paid parking	0.216**	B_CPARK_EB ²	-0.837**
		B_CPARK_EB: > 50% paid parking	0.522**		
		B_CPARK_EB: 100% paid parking	0.758**		
B_CPARK_C	<i>ref</i>	B_CPARK_C	<i>ref</i>	B_CPARK_C	<i>ref</i>
				B_CPARK_C ²	<i>ref</i>
B_CPARK_BTM	1.620**	B_CPARK_BTM: No paid parking	<i>ref</i>	B_CPARK_BTM	1.960**
		B_CPARK_BTM: < 50% paid parking	0.408**	B_CPARK_BTM ²	-0.380*
		B_CPARK_BTM: > 50% paid parking	1.290**		
		B_CPARK_BTM: 100% paid parking	1.980**		
B_CPARK_T	1.200**	B_CPARK_T: No paid parking	<i>ref</i>	B_CPARK_T	3.510**
		B_CPARK_T: < 50% paid parking	0.397**	B_CPARK_T ²	-2.410**
		B_CPARK_T: > 50% paid parking	1.110**		
		B_CPARK_T: 100% paid parking	1.200**		

ref = reference alternative / category*red* ≠ significant* = $p < 0.05$ ** = $p < 0.01$

The expectation of the influence of car parking zones is that a higher percentage of car parking zones in a zip code leads to a higher resistance against using the car. Besides the quadratic component, also a categorical definition of the variable is tested. The results in Table 5.9 show that the categorical variable significantly fits indeed better than the continuous variable. The parameter estimates also show that there is an increase in utility for each category. Therefore, a quadratic component can be tested. This is again a better fit with the data, and everything is also significant.

The positive main effect and the negative quadratic component of all modes mean a smaller change in utility towards a higher percentage of car parking. The highest change in utility is therefore present at the moment paid car parking can be found in a zip code. Because the best fit with the data can be seen for the quadratic component, and because all are significant, the quadratic component for car parking zones is added to the model. A plot of the relationship that results from these estimated parameters can be found in Appendix Figure E.2.

From the results of this subsection, it can be concluded that there are 4 interaction effects: distance & gender, distance & age, distance & speed, and car parking & residential zip code. Moreover, there are 3 quadratic components which are the travel distance, the car parking zones, and the number of cars in a household.

5.2.4. FINAL MODEL

This section shows the results of three types of models which includes all remaining variables, interaction effects, and quadratic components. These results are shown and interpreted for a MultiNomial Logit. After that, nests are determined for the alternatives for the Nested Logit and a panel structure is added for the Mixed Logit. These three types of models are estimated, to further optimize the model fit and to find the most accurate model for the data. Each type of model includes other assumptions and methods of calculating the influence of the variables, which can lead to a difference in model fit. The results following from these models are shown and compared to each other to see how these different logit models influence the model fit. Finally, the best fitting model is chosen and determined to be the final model of the research at the end of this section.

MultiNomial Logit

A MultiNomial logit model is estimated for all motives. The results of the multinomial logit can be found in Table 5.10. Compared to the results of the MNL with all main variables in Table 5.5, the significance and most values of the main variables did not change. However, there are differences in the interaction effects and quadratic components. The quadratic component of the travel distance for the car is not significant, just like most quadratic components of car parking. The interaction of car parking with residential zip code is not significant for the e-bike. Because the interaction effect of the travel distance and age categories of 18-40 and 41-66 were not significant for the e-bike, they are not included in this model to ensure that the main age category variables are significant for the e-bike. The interaction of distance & gender has the same behaviour as before, but including this interaction effect probably led to the change in the estimated main gender parameter for the bike from a negative to a positive value. It means that the interaction leads to a change from men preferring to choose the bike, to women preferring to choose the bike.

Besides estimating an MNL model with trips from all travel motives, Appendix Section E.1 shows two MNL models that are estimated with trips from only practical motives or leisure motives. These models are estimated to gain insights into the difference in variable parameters when specifying a travel motive, but their results are not further used in this research. In the appendix, the parameter estimates of the two travel motive MNL models are explained and they are compared to the parameter estimates of the MNL model with all travel motives in Table 5.10.

Table 5.10: MNL all variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	0.623**	1.650**	ref: 1.0	-4.740**	-4.300**
Access to public transport (Origin)	-	-	-	1.190**	0.577**
Access to public transport (Destination)	-	-	-	1.280**	0.734**
Age: 0-17	ref	ref	ref	ref	ref
Age: 18-40	-0.838**	-0.757**	ref	-0.728**	-0.835**
Age: 41-66	-0.787**	-0.471*	ref	-0.932**	-1.310**
Age: 67-100	-0.998**	-0.481*	ref	-0.703**	-2.840**
Cars per driver's license	0.805**	-0.013	ref	0.513**	1.610**
Quadratic: cars per driver's license	-0.793**	-0.286**	ref	-0.600**	-1.580**
Travel distance	-0.323**	-0.304**	-0.138**	0.159**	-0.099**
Quadratic: travel distance	0.009**	0.008**	0.001	-0.007**	0.000*
Interaction: distance & age 0-17	ref	ref	-	-	-
Interaction: distance & age 18-40	-0.068**	-	-	-	-
Interaction: distance & age 41-66	-0.063**	-	-	-	-
Interaction: distance & age 67-100	-0.165**	-0.070**	-	-	-
Interaction: distance & gender	-0.068**	-0.076**	-	-	-
Travel speed	-	-	0.008*	-0.556*	10.500**
Car parking (Origin)	0.670**	1.000**	ref	0.808**	1.390**
Quadratic: car parking (Origin)	0.097	-0.352	ref	-0.028	-1.140**
Interaction: car parking (Origin) & residential zip code	0.144**	0.028	ref	0.810**	0.478**
Car parking (Destination)	0.712**	1.010**	ref	0.560**	1.830**
Quadratic: car parking (Destination)	-0.084	-0.497	ref	0.087	-1.730**
Interaction: car parking (Destination) & residential zip code	0.180**	0.092	ref	0.870**	0.656**
Education: primary education	ref	ref	ref	ref	ref
Education: vmbo/mavo	0.013	-0.235*	ref	0.546**	0.123
Education: havo/vwo	0.129*	-0.417**	ref	0.423**	0.197
Education: hbo/university	0.473**	-0.301**	ref	0.231*	0.279
Frequency of BTM stops (Origin)	0.109**	0.078*	ref	0.449**	0.389**
Frequency of BTM stops (Destination)	0.161**	0.080*	ref	0.431**	0.289**
Gender	0.098**	0.688**	ref	0.122**	-0.264**
Driver's license	-0.888**	-0.609**	ref	-1.610**	-1.490**
Household members	0.047**	-0.083**	ref	0.083**	-0.024
Occupation: part-time job	ref	ref	ref	ref	ref
Occupation: full-time job	-0.240**	-0.272**	ref	-0.139**	-0.287**
Occupation: own housekeeping	-0.164**	0.042	ref	0.112	-0.873**
Occupation: student	0.794**	-0.794**	ref	1.040**	0.885**
Occupation: unemployed / unable to work	-0.470**	-0.157*	ref	-0.009	-1.290**
Street density	2.290**	0.839**	-0.143	0.461*	-
Wealth: first 20% group	ref	ref	ref	ref	ref
Wealth: second 20% group	0.201**	0.214**	ref	0.195**	0.277**
Wealth: third 20% group	0.338**	0.314**	ref	0.488**	0.317**
Wealth: fourth 20% group	0.413**	0.444**	ref	0.350**	0.662**
Wealth: fifth 20% group	0.595**	0.340**	ref	0.259**	0.664**
Log likelihood: -87296.97	Rho-square-bar: 0.466				
ref= reference alternative / category	red ≠ significant		* = p < 0.05 ** = p < 0.01		

Nested Logit

A nested logit consists of nests within the mode alternatives. To test which nests are present in the data, NL models are estimated with only the ASC and the calculated travel distance. The nests are changed for each test, and the results can be found in Table 5.11. The first shown combination of modes is the reference for each test. The log likelihood and the rho-square-bar show that the best nested structure would be the first test. This contains the bike & e-bike in one nest and the train & BTM in another. However, the nest of the train & BTM in tests 1 and 3 show to have a very large nest parameter, which is not a logical result. Moreover, including a nest for the train & BTM lead to a correlation between the nest coefficient and the alternative specific constants. The nest parameter of the bike & e-bike is much smaller in tests 1 and 2. Because the change in log likelihood is not that different, and because of the illogical results and correlations of the train & BTM nest, the second test is chosen as the nested structure to use for the data.

Table 5.11: Testing the nests

	Log likelihood	Rho-square-bar	Reference	Mu	Nest	Mu	Nest	Mu
1	-100660.7	0.385	Car	1	Bike & E-Bike	1.91	Train & BTM	44
2	-101546.9	0.379	Car & Train & BTM	1	Bike & E-Bike	1.93		
3	-100971.7	0.383	Car & Bike & E-Bike	1			Train & BTM	39.6

The car, train, and BTM are not nested and are thus the reference. The reason that the train & BTM are not a nest can be because the train is not often chosen at distances shorter than 15 kilometers, and the behaviour of people choosing the train under these circumstances can be different than what is normally seen. A visualisation of the nested structure can be seen in Figure 5.1.

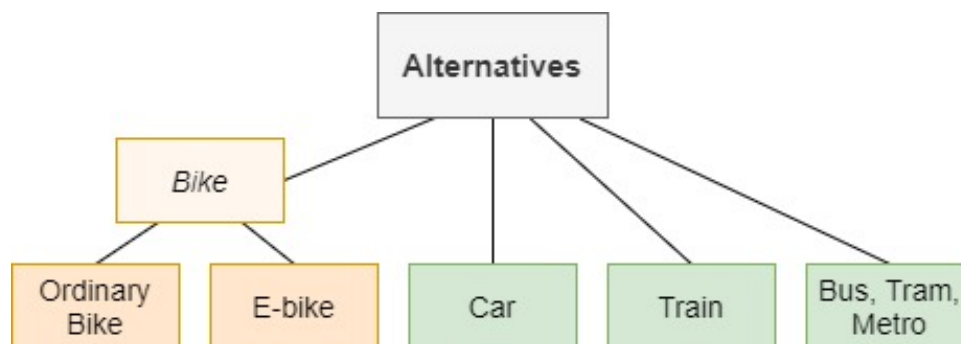


Figure 5.1: Structure of nests

The results of the nested logit can be found in Table 5.12. Differences with the MNL are that the number of cars per driver's license and the travel speed of the e-bike is significant in this model and that the education havo/vwo and the education hbo/university are not significant respectively for the bike and the e-bike. Also, the education level estimates for the e-bike are less negative. All other variables have the same significant result and the values are also similar to the MNL model. The largest changes are thus of the bike and e-bike, because of the nest implemented for these modes. The nest parameter of the bike & e-bike is significant and positive compared to the other modes. It shows that there is a positive difference in attitude towards the nest of the bike modes compared to the other modes.

Table 5.12: NL all variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	0.760**	1.630**	<i>ref</i>	-4.700**	-4.250**
Access to public transport (Origin)	-	-	-	1.200**	0.578**
Access to public transport (Destination)	-	-	-	1.290**	0.735**
Age: 0-17	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Age: 18-40	-0.834**	-0.855**	<i>ref</i>	-0.731**	-0.837**
Age: 41-66	-0.750**	-0.615**	<i>ref</i>	-0.925**	-1.310**
Age: 67-100	-0.903**	-0.702**	<i>ref</i>	-0.692**	-2.850**
Cars per driver's license	0.731**	0.387**	<i>ref</i>	0.430**	1.520**
Quadratic: cars per driver's license	-0.756**	-0.496**	<i>ref</i>	-0.561**	-1.530**
Travel distance	-0.324**	-0.327**	-0.140**	0.160**	-0.099**
Quadratic: travel distance	0.009**	0.009**	0.001	-0.007**	0.000*
Interaction: distance & age 0-17	<i>ref</i>	<i>ref</i>	-	-	-
Interaction: distance & age 18-40	-0.065**	-	-	-	-
Interaction: distance & age 41-66	-0.059**	-	-	-	-
Interaction: distance & age 67-100	-0.150**	-0.061**	-	-	-
Interaction: distance & gender	-0.069**	-0.057**	-	-	-
Travel speed	-	-	0.008**	-0.551*	10.500**
Car parking (Origin)	0.659**	0.845**	<i>ref</i>	0.811**	1.390**
Quadratic: car parking (Origin)	0.082	-0.208	<i>ref</i>	-0.037	-1.150**
Interaction: car parking (Origin) & residential zip code	0.149**	0.101	<i>ref</i>	0.806**	0.475**
Car parking (Destination)	0.704**	0.840**	<i>ref</i>	0.560**	1.820**
Quadratic: car parking (Destination)	-0.102	-0.333	<i>ref</i>	0.081	-1.730**
Interaction: car parking (Destination) & residential zip code	0.185**	0.150	<i>ref</i>	0.866**	0.654**
Education: primary education	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Education: vmbo/mavo	0.035	-0.152*	<i>ref</i>	0.543**	0.122
Education: havo/vwo	0.094	-0.281**	<i>ref</i>	0.405**	0.180
Education: hbo/university	0.406**	-0.118	<i>ref</i>	0.202*	0.253
Frequency of BTM stops (Origin)	0.102**	0.090**	<i>ref</i>	0.447**	0.387**
Frequency of BTM stops (Destination)	0.152**	0.115**	<i>ref</i>	0.428**	0.287**
Gender	0.120**	0.499**	<i>ref</i>	0.128**	-0.259**
Driver's license	-0.857**	-0.668**	<i>ref</i>	-1.600**	-1.480**
Household members	0.042**	-0.054**	<i>ref</i>	0.080**	-0.027
Occupation: part-time job	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Occupation: full-time job	-0.254**	-0.273**	<i>ref</i>	-0.141**	-0.288**
Occupation: own housekeeping	-0.149**	-0.024	<i>ref</i>	0.116	-0.870**
Occupation: student	0.751**	-0.247**	<i>ref</i>	1.020**	0.872**
Occupation: unemployed / unable to work	-0.449**	-0.227**	<i>ref</i>	0.006	-1.270**
Street density	2.140**	1.020**	-0.122	0.459*	-
Wealth: first 20% group	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Wealth: second 20% group	0.212**	0.231**	<i>ref</i>	0.200**	0.285**
Wealth: third 20% group	0.364**	0.326**	<i>ref</i>	0.496**	0.328**
Wealth: fourth 20% group	0.447**	0.449**	<i>ref</i>	0.363**	0.676**
Wealth: fifth 20% group	0.615**	0.429**	<i>ref</i>	0.268**	0.674**
Nest parameter: Bike & E-Bike	1.790**	-	-	-	-
Log likelihood: -86918.4	Rho-square-bar: 0.468				
<i>ref</i> = reference alternative / category	<i>red</i> ≠ significant				
	* = $p < 0.05$ ** = $p < 0.01$				

Mixed Logit

A mixed logit contains three extra components compared to the MNL model. The first one is that it also includes nests. These are the same as shown in the nested logit structure in Figure 5.1. Secondly, a panel structure of the data can be added to the model. It is used when the data consists of people making multiple entries. In this case, respondents reported multiple trips which are used as one observation and thus this panel structure of the data is also needed in the mixed logit. The distribution of the number of trips given per respondent can be found in Figure 5.2. It shows that most often two trips are taken in a day, with an average of 2.87 trips per day.

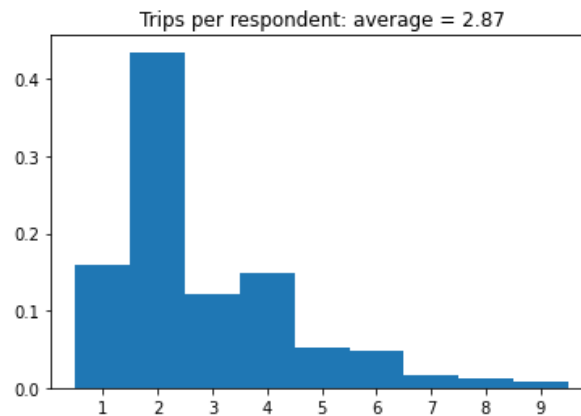


Figure 5.2: Trips per respondents

Lastly, respondents can have different tastes which could indicate that a fixed value for each estimated parameter is less accurate. To further specify and model these different tastes, a parameter can be estimated randomly by being normally distributed. Besides the estimation of the parameter, an estimation is also given for the standard deviation of the parameter. However, computational times increase rapidly when including random parameters because the computation then needs Monte-Carlo simulation techniques. It is therefore determined that there is not enough computational power to test any random parameters in the model.

The results of the mixed logit can be found in Table 5.13. Because only a panel structure is added to the model compared to the nested logit, there is again not much difference in the estimation of the variables compared to the multinomial logit and nested logit. However, the log likelihood has increased highly. This is caused by the change in respondents because of the panel structure. First, the number of respondents was based on the number of rows, while in a panel structure the rows are seen as observations. The number of respondents is determined by adding a column to the dataset that indicates which trips are taken by which respondent. The rho-square-bar does not depend on the number of respondents and can therefore be used to compare the models. For the mixed logit model, the rho-square-bar is 0.467. The nested logit has a rho-square-bar of 0.468. Therefore, the panel structure does not add much difference to the model fit. This can be caused by many respondents reporting two trips on one day as seen in Figure 5.2. Often these two trips are to and from the same locations, which means there is little variation and most often the same transport mode is chosen. This does not add much extra information and can explain the fact that there is hardly a difference in model fit.

Although the estimated parameter values are not changed much, the significance of the parameters did change. By including the panel structure and thus indicating there are fewer respondents, there is less information available to estimate the parameters. This leads to changes in the significance of variables. Compared to the nested logit, the education level of havo/vwo has become significant and positive for the bike. This was also the case in the MNL model. For the e-bike, all education categories, the frequency of BTM at the origin, and being a student have become insignificant. Because the car is the reference category, nothing changed for this mode. The number of cars per driver's license, paid car parking at the destination, and having a full-time job are insignificant for BTM. Finally, the variables changed for the train are the quadratic component of travel distance (although this was already estimated to be zero), the quadratic component of car parking at the origin, and the occupation of full-time working and own housekeeping.

Table 5.13: ML all variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	0.529*	1.280**	<i>ref</i>	-4.750**	-4.470**
Access to public transport (Origin)	-	-	-	1.090**	0.672**
Access to public transport (Destination)	-	-	-	1.410**	0.780**
Age: 0-17	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Age: 18-40	-0.772**	-0.890**	<i>ref</i>	-0.770**	-0.730**
Age: 41-66	-0.709**	-0.623**	<i>ref</i>	-0.924**	-1.180**
Age: 67-100	-0.818**	-0.713**	<i>ref</i>	-0.579**	-2.640**
Cars per driver's license	0.754**	0.435**	<i>ref</i>	0.293	1.150**
Quadratic: cars per driver's license	-0.773**	-0.535**	<i>ref</i>	-0.468**	-1.270**
Travel distance	-0.381**	-0.308**	-0.078*	0.173**	-0.090**
Quadratic: travel distance	0.013**	0.008**	-0.003	-0.008**	0.000
Interaction: distance & age 0-17	<i>ref</i>	<i>ref</i>	-	-	-
Interaction: distance & age 18-40	-0.073**	-	-	-	-
Interaction: distance & age 41-66	-0.053*	-	-	-	-
Interaction: distance & age 67-100	-0.139**	-0.057**	-	-	-
Interaction: distance & gender	-0.061**	-0.050**	-	-	-
Travel speed	-	-	0.011*	-0.422	10.300**
Car parking (Origin)	0.767**	0.790*	<i>ref</i>	0.836*	1.290*
Quadratic: car parking (Origin)	-0.071	-0.253	<i>ref</i>	-0.098	-1.070
Interaction: car parking (Origin) & residential zip code	0.241**	0.314*	<i>ref</i>	0.928**	0.715**
Car parking (Destination)	0.728**	0.640*	<i>ref</i>	0.558	1.690**
Quadratic: car parking (Destination)	-0.245	-0.238	<i>ref</i>	-0.084	-1.660**
Interaction: car parking (Destination) & residential zip code	0.267**	0.199	<i>ref</i>	1.030**	0.826**
Education: primary education	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Education: vmbo/mavo	0.181	0.085	<i>ref</i>	0.615**	0.044
Education: havo/vwo	0.237*	-0.045	<i>ref</i>	0.438**	0.143
Education: hbo/university	0.529**	0.090	<i>ref</i>	0.224	0.303
Frequency of BTM stops (Origin)	0.085**	0.009	<i>ref</i>	0.462**	0.329**
Frequency of BTM stops (Destination)	0.195**	0.208**	<i>ref</i>	0.447**	0.322**
Gender	0.097**	0.478**	<i>ref</i>	0.133**	-0.298**
Driver's license	-0.863**	-0.634**	<i>ref</i>	-1.650**	-1.450**
Household members	0.046**	-0.054**	<i>ref</i>	0.108**	-0.003
Occupation: part-time job	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Occupation: full-time job	-0.215**	-0.255**	<i>ref</i>	-0.025	-0.230
Occupation: own housekeeping	-0.148*	-0.064	<i>ref</i>	0.053	-0.687
Occupation: student	0.862**	-0.063	<i>ref</i>	1.110**	1.100**
Occupation: unemployed / unable to work	-0.391**	-0.205*	<i>ref</i>	0.030	-1.050**
Street density	1.860**	1.220**	0.045	0.716*	-
Wealth: first 20% group	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Wealth: second 20% group	0.244**	0.267**	<i>ref</i>	0.233*	0.381*
Wealth: third 20% group	0.393**	0.310**	<i>ref</i>	0.508**	0.597**
Wealth: fourth 20% group	0.504**	0.465**	<i>ref</i>	0.400**	0.848**
Wealth: fifth 20% group	0.633**	0.451**	<i>ref</i>	0.290**	0.935**
Nest parameter: Bike & E-Bike	1.820**	-	-	-	-
Log likelihood: -30332.03	Rho-square-bar: 0.467				
<i>ref</i> = reference alternative / category	<i>red</i> ≠ significant * = $p < 0.05$ ** = $p < 0.01$				

Conclusion

Because the mixed logit has almost no change in model fit compared to the nested logit, and because of the many variables becoming insignificant by adding the panel structure, the nested logit model is chosen to be the final model for this research. With this model, the remaining steps of this research are performed. Thus, the nested logit model is used for the sensitivity analysis, uncertainty analysis, validation, and application in the upcoming chapters and sections.

5.3. SENSITIVITY & UNCERTAINTY ANALYSES

This section tests the final model (nested logit) on its sensitivity and uncertainty. The purpose of the analyses is to see how much the beta estimates and input values contribute to changes in the estimated modal split. The sensitivity analysis is done to assess the sensitivity of the model outcome as a result of the uncertainty in the estimated beta parameters. For each variable, a random value is drawn 10000 times from the normal distribution of the beta parameter for each mode. The mean and standard deviation resulting from the final model are used. These mean parameter values are then changed and the probabilities of the modes are calculated with the input values shown in Appendix F. The calculation of the probabilities is based on the occurrences of people owning an e-bike or a car. These are based on the averages of the Netherlands, which are 86% owning a car and 13% owning an e-bike. Four occurrences can then be formed based on the availability. These are an 11% chance of owning all the modes, 75% chance of owning no e-bike, 2% chance of owning no car, and 12% chance of owning no e-bike and no car. The final probability is the weighted average of these probabilities and the occurrences. The uncertainty analysis is done to assess the uncertainty of the outcome because of changes in the input values. The same calculation is done for the sensitivity analysis. The only difference is that the input values are changed instead of the beta parameters. The mean and standard deviation for the normal distribution of the input values are shown in Appendix F as well.

5.3.1. SENSITIVITY ANALYSIS

The sensitivity analysis results in Table 5.14 show that all variables do not show to have large variations or large differences from the original mean. This results from the fact that these variables are significant, which means that the beta estimates do not have large standard deviations. A draw from the distribution of a beta estimate is thus not far from the estimated mean beta and therefore it would not lead to large variations in outcome. The standard deviation is not larger than 2.5 percentage points from the calculated mean. The calculated mean of the parameter estimates is also all within 1 percentage point from the original probabilities. Therefore, the model is very robust and not very sensitive to any possible changes in the parameter estimates.

The analysis is now solely performed on changing the parameters of one variable at a time. However, second-order effects of two variables changing at the same time could lead to larger sensitivities in the model. The change of one variable generally leads to very small changes in the modal split, thus this second-order effect is analysed for the variables that show to have a variation larger than 1 percentage point in the sensitivity analysis. These are the alternative specific constants, the travel distance main variable, and the number of cars per driver's license main variable. Table 5.15 shows the results of the combination of two variables. Here it can be seen that the change in the beta parameters of two variables does not amplify the variation when they are adjusted at the same time compared to their variations separately. Only the number of cars per driver's license & travel distance increased the standard deviation by about 0.5 percentage point. The other two combinations have similar variations as the variation of the ASC alone. The calculated averages are also still very similar to the original probabilities. It can thus be seen that second-order sensitivities are not present in the model.

5.3.2. UNCERTAINTY ANALYSIS

The uncertainty analysis results in Table 5.16 show that the largest differences can be found for the travel distance. It has been found that the travel distance is thus the most influential variable for the probabilities of the modes. All other input variables do not change the original mean probabilities by more than 1 percentage point. Other notable results are the larger than average standard deviations of the age categories, car parking, and driver's license. For at least one of the modes, the standard deviation is then larger than 5 percentage points. A change in input of these variables can therefore change the probability of the modes the most.

To also analyse second-order effects in the uncertainty analysis, the mentioned variables that influence the outcome the most are combined to analyse their behaviour together. There are 21 combinations that can be made, and the results are found in Table 5.17. It shows that the largest variations can be found for the combinations with the travel distance, although the combined variation is not significantly larger than the variation of travel distance alone. The largest changes in variation can be found for the combinations of the age groups, and the combination of the car parking locations. This can be caused by the correlation between the categories of the same variable, but the changes in variation are nonetheless not greater than 3 percentage points. Moreover, the age categories of the bike and the e-bike are correlated to the ASC, which could also lead to larger variations in the outcome. Because the largest changes can be found in the travel distance, it is thus most important for the travel distance to have accurate input values.

Table 5.14: Sensitivity Analysis

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Original</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Alternative Specific Constants	35.90% (2.21)	4.08% (0.57)	53.68% (2.34)	5.89% (0.91)	0.46% (0.11)
Access to public transport (Origin)	35.88% (0.19)	4.08% (0.02)	53.75% (0.19)	5.84% (0.40)	0.45% (0.01)
Access to public transport (Destination)	35.88% (0.19)	4.08% (0.02)	53.76% (0.19)	5.83% (0.40)	0.45% (0.01)
Age: 0-17	-	-	-	-	-
Age: 18-40	35.89% (0.47)	4.08% (0.12)	53.76% (0.49)	5.83% (0.12)	0.45% (0.02)
Age: 41-66	35.89% (0.65)	4.08% (0.17)	53.75% (0.67)	5.83% (0.18)	0.45% (0.02)
Age: 67-100	35.88% (0.37)	4.08% (0.09)	53.76% (0.39)	5.82% (0.11)	0.45% (0.02)
Cars per driver's license	35.88% (1.12)	4.08% (0.29)	53.76% (1.27)	5.83% (0.43)	0.45% (0.08)
Quadratic: cars per driver's license	35.89% (0.59)	4.08% (0.15)	53.75% (0.63)	5.83% (0.25)	0.45% (0.06)
Travel distance	35.92% (1.33)	4.08% (0.17)	53.72% (1.12)	5.84% (0.65)	0.45% (0.06)
Quadratic: travel distance	35.66% (0.21)	4.06% (0.04)	54.04% (0.19)	5.79% (0.20)	0.44% (0.01)
Interaction: distance & age 0-17	-	-	-	-	-
Interaction: distance & age 18-40	35.89% (0.29)	4.08% (0.02)	53.76% (0.23)	5.82% (0.04)	0.45% (0.00)
Interaction: distance & age 41-66	35.89% (0.40)	4.08% (0.02)	53.76% (0.32)	5.82% (0.05)	0.45% (0.00)
Interaction: distance & age 67-100	35.89% (0.26)	4.08% (0.03)	53.76% (0.21)	5.82% (0.03)	0.45% (0.00)
Interaction distance & gender	35.89% (0.20)	4.08% (0.05)	53.76% (0.18)	5.82% (0.03)	0.45% (0.00)
Travel speed	35.89% (0.15)	4.08% (0.01)	53.76% (0.15)	5.83% (0.26)	0.45% (0.02)
Car parking (Origin)	35.89% (0.34)	4.08% (0.10)	53.76% (0.37)	5.83% (0.15)	0.45% (0.02)
Quadratic: car parking (Origin)	35.94% (0.05)	4.06% (0.02)	53.74% (0.06)	5.82% (0.02)	0.45% (0.00)
Interaction: car parking (Origin) & residential zip code	35.87% (0.06)	4.11% (0.02)	53.75% (0.07)	5.82% (0.02)	0.45% (0.00)
Car parking (Destination)	35.88% (0.34)	4.08% (0.10)	53.76% (0.37)	5.83% (0.15)	0.45% (0.02)
Quadratic: car parking (Destination)	35.85% (0.05)	4.05% (0.02)	53.81% (0.06)	5.84% (0.02)	0.45% (0.00)
Interaction: car parking (Destination) & residential zip code	35.87% (0.06)	4.12% (0.02)	53.74% (0.07)	5.82% (0.02)	0.45% (0.00)
Education: Primary education	-	-	-	-	-
Education: vmbo/mavo	36.03% (0.22)	4.07% (0.05)	53.63% (0.24)	5.80% (0.10)	0.46% (0.02)
Education: havo/vwo	36.62% (0.40)	4.04% (0.10)	53.15% (0.43)	5.73% (0.16)	0.47% (0.03)
Education: hbo/university	35.93% (0.33)	3.94% (0.08)	53.82% (0.35)	5.83% (0.14)	0.49% (0.02)
Frequency of BTM stops (Origin)	35.89% (0.12)	4.08% (0.04)	53.76% (0.13)	5.82% (0.04)	0.45% (0.01)
Frequency of BTM stops (Destination)	35.89% (0.12)	4.08% (0.04)	53.76% (0.13)	5.82% (0.04)	0.45% (0.01)
Gender	35.89% (0.22)	4.08% (0.06)	53.76% (0.21)	5.82% (0.08)	0.45% (0.01)
Driver's License	35.89% (0.36)	4.08% (0.10)	53.75% (0.42)	5.82% (0.12)	0.45% (0.02)
Household Members	35.89% (0.27)	4.08% (0.08)	53.77% (0.29)	5.83% (0.13)	0.42% (0.02)
Occupation: Part-time job	-	-	-	-	-
Occupation: Full-time job	35.89% (0.14)	4.08% (0.04)	53.76% (0.15)	5.82% (0.07)	0.45% (0.01)
Occupation: Own housekeeping	35.85% (0.12)	4.06% (0.03)	53.72% (0.12)	5.92% (0.06)	0.45% (0.02)
Occupation: Student	35.89% (0.09)	4.08% (0.03)	53.76% (0.10)	5.82% (0.03)	0.45% (0.00)
Occupation: Unemployed / Unable to work	35.89% (0.05)	4.08% (0.01)	53.76% (0.05)	5.83% (0.02)	0.45% (0.01)
Street Density	36.42% (0.67)	4.14% (0.18)	53.07% (0.69)	5.92% (0.34)	0.45% (0.03)
Wealth: First 20% group	-	-	-	-	-
Wealth: Second 20% group	35.89% (0.13)	4.08% (0.04)	53.76% (0.15)	5.82% (0.05)	0.45% (0.01)
Wealth: Third 20% group	35.88% (0.13)	4.08% (0.04)	53.77% (0.15)	5.82% (0.05)	0.45% (0.01)
Wealth: Fourth 20% group	35.89% (0.13)	4.08% (0.04)	53.76% (0.14)	5.82% (0.05)	0.45% (0.01)
Wealth: Fifth 20% group	35.89% (0.13)	4.08% (0.04)	53.76% (0.14)	5.82% (0.05)	0.45% (0.01)
Nest parameter: Bike & E-Bike	35.89% (0.03)	4.08% (0.02)	53.76% (0.01)	5.82% (0.00)	0.45% (0.00)

Table 5.15: Second-order sensitivity analysis

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Original</i>	35.89%	4.08%	53.76%	5.82%	0.45%
ASC & Travel distance	35.88% (1.95)	4.08% (0.58)	53.68% (2.22)	5.91% (0.99)	0.46% (0.10)
ASC & Cars per driver's license	35.90% (1.75)	4.08% (0.33)	53.70% (1.70)	5.86% (0.79)	0.46% (0.10)
Cars per driver's license & Travel distance	35.88% (2.35)	4.08% (0.60)	53.68% (2.46)	5.90% (0.99)	0.47% (0.14)

Table 5.16: Uncertainty Analysis

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Original</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Access to public transport (Origin)	35.83% (0.66)	4.08% (0.06)	53.69% (0.65)	5.96% (1.38)	0.45% (0.07)
Access to public transport (Destination)	35.80% (0.73)	4.07% (0.07)	53.67% (0.72)	6.00% (1.52)	0.46% (0.09)
Age: 0-17	-	-	-	-	-
Age: 18-40	36.46% (8.02)	4.06% (0.73)	53.26% (9.07)	5.78% (0.29)	0.45% (0.04)
Age: 41-66	36.40% (7.57)	4.01% (0.41)	53.27% (9.09)	5.85% (0.97)	0.47% (0.16)
Age: 67-100	36.80% (9.87)	3.98% (0.55)	52.89% (10.53)	5.74% (0.32)	0.60% (0.51)
Cars per driver's license	35.05% (2.56)	4.02% (0.27)	54.73% (3.22)	5.77% (0.32)	0.42% (0.08)
Travel distance	39.57% (22.27)	4.00% (1.88)	48.71% (20.95)	5.25% (3.84)	2.47% (6.68)
Travel speed	35.50% (1.32)	4.04% (0.13)	53.34% (1.50)	5.77% (0.44)	1.35% (2.99)
Car parking (Origin)	35.90% (3.13)	4.10% (0.64)	53.65% (5.18)	5.94% (1.34)	0.41% (0.10)
Dummy: residential zip code (Origin)	35.88% (0.07)	4.08% (0.03)	53.76% (0.34)	5.83% (0.29)	0.45% (0.01)
Car parking (Destination)	35.96% (3.77)	4.10% (0.63)	53.68% (5.40)	5.87% (0.90)	0.40% (0.12)
Dummy: residential zip code (Destination)	35.88% (0.11)	4.08% (0.03)	53.76% (0.41)	5.83% (0.31)	0.45% (0.02)
Education: Primary education	-	-	-	-	-
Education: vmbo/mavo	35.84% (0.49)	4.08% (0.29)	53.71% (0.46)	5.93% (1.25)	0.45% (0.01)
Education: havo/vwo	35.84% (0.30)	4.08% (0.58)	53.70% (0.26)	5.93% (1.13)	0.45% (0.00)
Education: hbo/university	35.98% (3.61)	4.04% (0.25)	53.71% (3.31)	5.81% (0.01)	0.45% (0.04)
Frequency of BTM stops (Origin)	35.86% (0.31)	4.08% (0.05)	53.70% (1.32)	5.91% (0.92)	0.45% (0.06)
Frequency of BTM stops (Destination)	35.86% (0.75)	4.08% (0.07)	53.74% (1.65)	5.87% (0.80)	0.45% (0.03)
Gender	35.95% (1.84)	4.06% (0.61)	53.71% (0.78)	5.83% (0.50)	0.45% (0.05)
Driver's License	35.82% (5.39)	4.03% (0.50)	53.21% (9.24)	6.45% (3.24)	0.48% (0.21)
Household Members	35.90% (0.95)	4.07% (0.36)	53.74% (1.01)	5.84% (0.43)	0.45% (0.01)
Occupation: Part-time job	-	-	-	-	-
Occupation: Full-time job	35.94% (2.10)	4.09% (0.34)	53.70% (2.47)	5.82% (0.01)	0.45% (0.03)
Occupation: Own housekeeping	35.89% (1.04)	4.08% (0.07)	53.74% (0.95)	5.82% (0.16)	0.47% (0.14)
Occupation: Student	36.07% (4.38)	4.00% (0.70)	53.56% (4.98)	5.93% (1.23)	0.45% (0.07)
Occupation: Unemployed / Unable to work	35.92% (2.07)	4.07% (0.08)	53.72% (1.94)	5.83% (0.32)	0.46% (0.12)
Street Density	36.03% (3.81)	4.06% (0.12)	53.62% (3.34)	5.84% (0.58)	0.45% (0.04)
Wealth: First 20% group	-	-	-	-	-
Wealth: Second 20% group	35.90% (1.36)	4.08% (0.24)	53.75% (1.79)	5.82% (0.17)	0.45% (0.03)
Wealth: Third 20% group	35.89% (2.23)	4.08% (0.28)	53.74% (3.17)	5.84% (0.64)	0.45% (0.02)
Wealth: Fourth 20% group	35.92% (2.89)	4.08% (0.43)	53.73% (3.63)	5.82% (0.23)	0.45% (0.07)
Wealth: Fifth 20% group	36.02% (4.40)	4.06% (0.33)	53.66% (4.63)	5.81% (0.15)	0.45% (0.06)

Table 5.17: Second-order uncertainty analysis

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Original</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Age: 18-40 & Age: 41-66	36.95% (10.86)	3.99% (0.78)	52.79% (12.53)	5.81% (0.98)	0.47% (0.17)
Age: 18-40 & Age: 67-100	37.03% (12.38)	3.94% (0.84)	52.74% (13.39)	5.71% (0.45)	0.58% (0.49)
Age: 18-40 & Travel distance	40.06% (22.79)	3.99% (1.92)	48.47% (21.88)	5.20% (3.78)	2.29% (6.43)
Age: 18-40 & Car parking (Origin)	36.55% (8.54)	4.08% (0.93)	53.05% (10.27)	5.90% (1.32)	0.41% (0.10)
Age: 18-40 & Car parking (Destination)	36.51% (8.80)	4.08% (0.93)	53.18% (10.41)	5.83% (0.90)	0.40% (0.13)
Age: 18-40 & Driver's License	36.40% (9.35)	4.02% (0.82)	52.69% (12.42)	6.42% (3.16)	0.48% (0.21)
Age: 41-66 & Age: 67-100	37.22% (12.15)	3.91% (0.63)	52.48% (13.49)	5.77% (1.00)	0.62% (0.58)
Age: 41-66 & Travel distance	39.86% (22.39)	3.96% (1.84)	48.49% (21.67)	5.27% (3.94)	2.43% (6.70)
Age: 41-66 & Car parking (Origin)	36.44% (8.16)	4.03% (0.73)	53.12% (10.37)	5.98% (1.64)	0.44% (0.18)
Age: 41-66 & Car parking (Destination)	36.49% (8.36)	4.03% (0.71)	53.16% (10.40)	5.90% (1.30)	0.42% (0.20)
Age: 41-66 & Driver's License	36.30% (8.83)	3.97% (0.59)	52.74% (12.34)	6.48% (3.30)	0.51% (0.28)
Age: 67-100 & Travel distance	40.06% (22.86)	3.92% (1.87)	47.96% (22.08)	5.19% (3.85)	2.87% (7.63)
Age: 67-100 & Car parking (Origin)	36.81% (10.30)	3.99% (0.80)	52.79% (11.54)	5.86% (1.33)	0.55% (0.49)
Age: 67-100 & Car parking (Destination)	36.73% (10.36)	3.99% (0.78)	52.95% (11.48)	5.80% (0.92)	0.53% (0.49)
Age: 67-100 & Driver's License	36.53% (10.89)	3.92% (0.68)	52.54% (13.42)	6.38% (3.20)	0.63% (0.62)
Travel distance & Car parking (Origin)	39.93% (22.02)	4.06% (1.90)	48.50% (21.19)	5.25% (4.04)	2.26% (6.31)
Travel distance & Car parking (Destination)	40.02% (22.22)	4.06% (1.89)	48.33% (21.23)	5.27% (4.03)	2.31% (6.57)
Travel distance & Driver's License	39.74% (22.64)	3.98% (1.87)	47.92% (21.98)	5.78% (5.29)	2.59% (7.19)
Car parking (Origin) & Car parking (Destination)	36.05% (4.74)	4.13% (0.87)	53.43% (7.27)	6.01% (1.60)	0.37% (0.14)
Car parking (Origin) & Driver's License	35.84% (6.06)	4.05% (0.78)	53.08% (10.42)	6.58% (3.59)	0.45% (0.22)
Car parking (Destination) & Driver's License	36.06% (6.29)	4.07% (0.76)	52.88% (10.34)	6.55% (3.35)	0.44% (0.23)

5.4. CONCLUSION

The conclusion of this chapter is that the final model is based on the results of the nested logit model found in Table 5.12. Its characteristics are:

- A nest for the bike and e-bike
- A rho-square-bar of 0.468 and a log likelihood of -86918.4
- Thirteen main variables:
 - Four spatial characteristics
 - Eight people's characteristics
 - One transport journey characteristic
- Four interaction effects:
 - Distance & gender
 - Distance & age
 - Distance & speed
 - Car parking & residential zip code
- Three quadratic components:
 - Travel distance
 - Car parking zones
 - Number of cars per driver's license in a household

The sensitivity analysis shows that the model is very robust and not sensitive to changes in the beta estimates, and the uncertainty analysis shows that the most influential variable is the travel distance. Second-order effects are not found for both the sensitivity and the uncertainty analysis. It shows that the final model is stable.

6

VALIDATION

This chapter presents the validation results of the final model. It is checked if the model can indeed predict the shares of the modes in the modal split accurately enough. It is assumed that a modal split is estimated accurately if the shares of the modes are within 3 percentage points of the actual share. First, the internal validation is performed. It estimates the same model without a slice of the dataset to check the in-sample prediction. The results are compared to the final model. Also, another internal validation is performed in which the final model predicts the choices of the respondents of the used dataset to see how well it can estimate its own modal split. However, the data used in this validation is already used in the model estimation. Therefore, an external validation is also performed. The mode choice is determined for each respondent of the validation dataset with trips to and/or from Haarlem. The modal split generated from the estimated choice per respondent is compared to the original modal split of the dataset to check how well the model can predict out-of-sample. Lastly, the performance is further analysed for both internal validation and external validation by calculating the accuracy, precision, recall, and F1-score.

6.1. INTERNAL VALIDATION

The first internal validation is performed by a function in Biogeme. The dataset is split randomly into 5 smaller datasets with which the model is validated. One slice is used as the validation set and with the other slices, the model is re-estimated. This is done 5 times so that every slice has been a validation set. The results are that the estimated values for the parameters are completely the same for all 5 re-estimations. The log likelihood has dropped from -86918.4 of the final model to -87043.1 of the re-estimations. The rho-square-bar has changed from 0.468 to 0.467. The results of the application of these values on the validation set are shown in Table 6.1. The log likelihoods are all estimated similarly because they are all within a range of 200. It can thus be seen that by removing a slice of the data, the model with its factors is still estimated the same as the original model and the validation slice is also similarly estimated. This means that the model is not very sensitive to the size of data that is needed, or small differences in data. Overall, the model is thus robust.

Table 6.1: Log likelihood of validation sets

	<i>Validation set 1</i>	<i>Validation set 2</i>	<i>Validation set 3</i>	<i>Validation set 4</i>	<i>Validation set 5</i>
Log likelihood	-17411.9	-17448.7	-17452.6	-17455.9	-17274.1

Besides re-estimating the model and validating a slice of the data, the model can also be used on its own data to see how well it predicts its own modal split. It is expected that this will be quite accurate, as the model is estimated with this data. The percentage of the trips where the mode is estimated correctly for each individual is 66%. Although multiple factors are assessed, people can still have different behaviours. This correctly estimated value is therefore a decent result. However, the estimated modal split is the most important outcome that must be accurate. These results can be found in Figure 6.1.

The modal split of the dataset in Figure 6.1a is different from Figure 4.4 in the data analysis, because in this section it is the modal split of the filtered dataset. The car is thus used in half of the trips, and the bike is still the second-largest chosen mode. The modal split of the choices is calculated as follows: the mode with the highest probability is determined to be chosen, and the modal split of the respondents generated from these choices is then the outcome. This result is shown in Figure 6.1b, which is not very accurate and does not fall within the 3 percentage points accuracy margin.

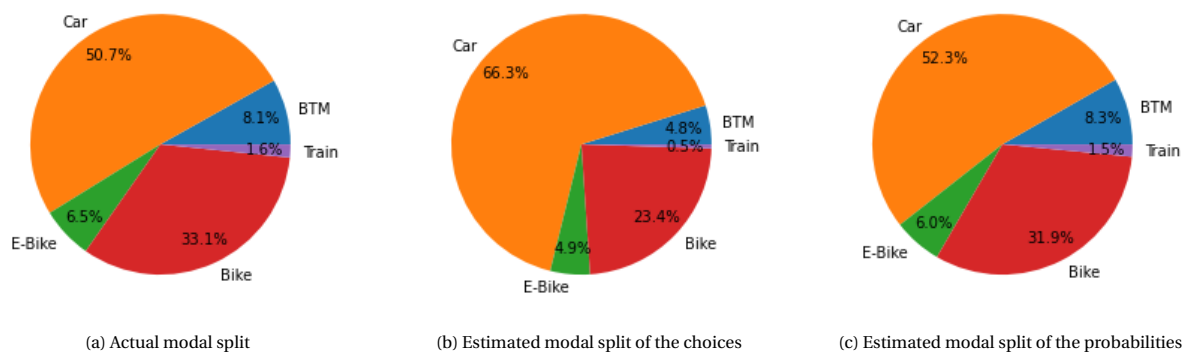


Figure 6.1: Modal split ODIN

The car is very much overestimated, and the other modes are underestimated. This can be caused by the fact that when the probabilities of the modes for a respondent are close, the higher one is still chosen. The probability of the car is more often high, thus the car is often picked as the predicted chosen alternative if the actual choice is another mode. The average of all probabilities that are calculated in Figure 6.1c is showing a better result and does fall within the 3 percentage points accuracy margin. It is very close to the actual modal split, and the model thus works best if the modal split of the probabilities is used as the predicted outcome.

Performance

To further analyse the results of the internal validation, the performance of the model is evaluated by three measures and a weighted average measure. In Table 6.2, it can be seen that the accuracy of the model is very high for all modes. The model thus correctly predicts if a mode is chosen or not for each trip. However, the precision is a bit lower. A lower value for precision means that the model incorrectly states that more trips are being taken by a certain mode than what is true. Especially for the e-bike, this value is low. More than half of the estimated trips taken by e-bike are thus actually not taken by e-bike. The model thus overestimates the e-bike choice when this option is available for a trip. The recall has even lower values for all modes except the car. A low value for recall means that from all trips actually taken by a certain mode, the model has trouble predicting those as taken by that mode. Especially for the train, it is very low, and thus the trips taken by train are often not estimated as being taken by train. This could be caused by the lower accuracy of the calculation of the travel distance as shown in Figure 4.8. The F1-score is the highest for the car because of the good scores for precision and recall. The lowest is for the train, because of its very low score for recall.

Table 6.2: Internal performance of the model

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Bike	0.752	0.678	0.480	0.562
E-Bike	0.932	0.460	0.350	0.398
Car	0.725	0.675	0.883	0.765
BTM	0.925	0.562	0.334	0.419
Train	0.984	0.533	0.165	0.252

6.2. EXTERNAL VALIDATION

The external validation is performed by estimating the choice probability of each mode for each respondent with a trip to and/or from Haarlem. Applying the final model to the ODIN data from Haarlem leads to correctly estimating the choice of 62% of the trips. A comparison of the results of the external validation of Haarlem can be found in Figure 6.2.

The modal splits shown are first the actual modal split in Figure 6.2a, the modal split based on the actual predicted choice for each trip in Figure 6.2b, and the modal split based on the average of the predicted probabilities for each trip in Figure 6.2c. The figures show that for the city of Haarlem, the model overestimates the choice for the car and underestimates the choice for the bike. Predicting a modal split based on the probabilities also relates better to the actual modal split in the external validation, but the bike is still much underestimated in this case as the difference is larger than 3 percentage points from the original.

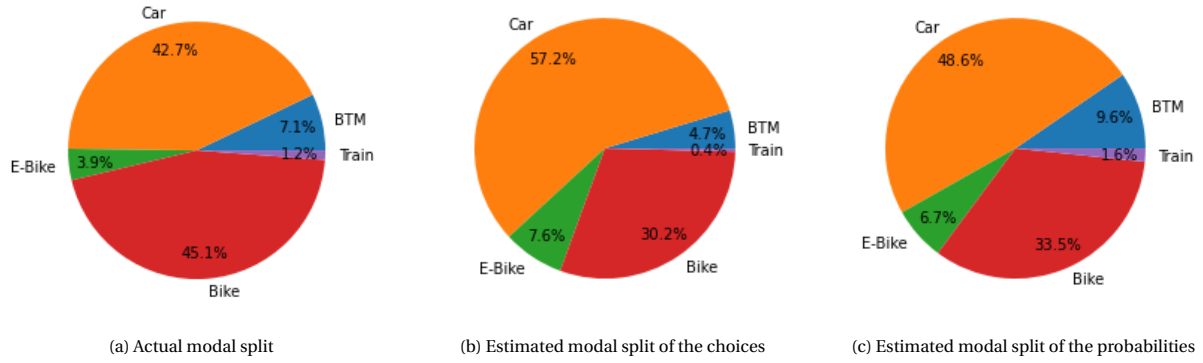


Figure 6.2: Modal split Haarlem

Based on the data, it is already concluded in Section 4.4 that the respondents of Haarlem show to have a larger preference for the bike compared to the average preference of the ODIN dataset. This behaviour of people in Haarlem can be modelled by changing the ASC's which show the average preference of a mode. The ASC of the car is fixed to 1. Increasing the ASC of the bike from 0.76 to 1.40 by trial-and-error would show that the bike has a preference over the car. This value leads to a modal split that relates closer to the original modal split and correctly estimates the mode choice of 72% of the trips (see Figure 6.3). Both the modal split of the choices in Figure 6.3b and the modal split of the probabilities in Figure 6.3c are within the 3 percentage points accuracy margin compared to the actual modal split in Figure 6.3a. Overall, the modal split of the probabilities is closer to the actual modal split than the modal split of the choices.

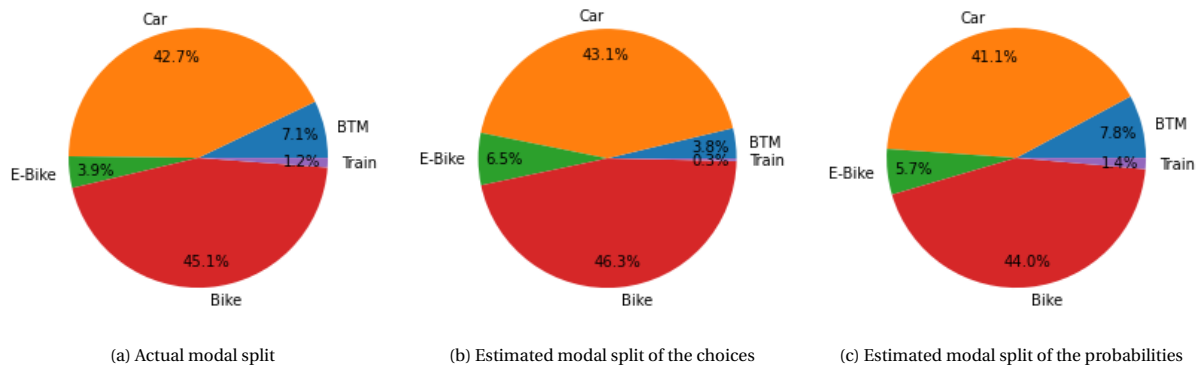


Figure 6.3: Modal split Haarlem by changing ASC Bike

Specifying the preference of the bike over the car of a given city thus leads to a better modal split prediction. The difference between the prediction of the choices and the prediction of the probabilities is small. For the prediction of the choices, differences can mostly be found in the share of the public transport modes. If these modes need to be more accurately determined, the ASC of the BTM and train could also be changed to further optimize a modal split for a given city. By making the model outcome accurate to start with, changes in the input values can better predict the modal split. However, this paper focuses on accurately predicting the bicycle shares, and changing the accuracy of the public transport modes is thus not included.

Performance

For the external validation, the performance is evaluated to further analyse its predicted outcomes. The measures calculated for the external performance in Table 6.3 show that these measures are similar to the internal performance. The accuracies are a bit lower for the bike, e-bike, and car, but a bit higher for the BTM and train. The bike scores better on precision & recall and thus has a higher F1-score in the external performance. Predicting if the bike is chosen is thus better estimated at the external performance. The e-bike shows a lower precision, but a higher recall. Of all estimated e-bike trips, less than 30% are actually taken by e-bike. However, it better estimates a trip taken by e-bike from all actual e-bike trips compared to the internal validation. The F1-score of the e-bike stays about the same. The car and train have slightly lower values for precision and

recall, thus also having a lower F1-score. The individual choice prediction of these modes is thus less exact at the external validation. The scores for BTM are similar to the internal performance and also have a similar F1-score.

Table 6.3: External performance of the model

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Bike	0.697	0.743	0.500	0.598
E-Bike	0.919	0.277	0.657	0.390
Car	0.685	0.601	0.781	0.679
BTM	0.935	0.568	0.362	0.442
Train	0.988	0.429	0.143	0.214

The differences between the internal and external performance are thus mostly present for the e-bike. The F1-score of both internal and external performance shows that the bike and car would be the better-estimated modes, while the accuracy did not. The model is thus better at predicting if a mode is chosen or not for the e-bike, BTM and train. For the bike and car, the model is better at predicting trips taken by these modes from all estimated trips taken by these modes and predicting the trips taken by these modes from all actual trips taken by these modes. However, correctly estimating the individual choice is not directly related to the goal of this research.

6.3. CONCLUSION

The internal validation shows that the model parameters are exactly the same when re-estimating without $\frac{1}{5}$ of the data. The validation slice then shows to have similar estimation results as well. Furthermore, the internal validation shows that 66% is estimated correctly and that the outcome of the estimated modal split of the probabilities gives the best result compared to the actual modal split of the data.

It can be concluded from the external validation that the model is best used in cities where the preference of the transport modes is similar to the calculated ASC's of the model. In a city like Haarlem where the bicycle is (based on the data) a more preferred mode, it will not predict the modal split accurately enough. Changing the alternative specific constants is an option, but municipalities then should perform a preliminary analysis on the preference of the bike over the car in the city.

The model performances show that the model has a high accuracy in predicting the mode choice of the trips. The modes being less accurately predicted are the car and the bike. This is possibly the case because the utility of these modes is often calculated by the model as being close to each other, and then it can be more easily predicted incorrectly. The precision and recall show less good scores and the F1-score is thus also low for some modes. However, these scores are based on individual choices and thus have to be interpreted lightly. A low score does not necessarily mean a bad modal split prediction. It shows the behaviour of the model and the individual choices underneath the predicted modal split. The modes with a low F1-score are the e-bike, BTM, and train.

7

APPLICATION

This chapter shows what the final nested logit model as found in Table 5.12 can do and how it practically can be used by analysing elasticities of variables. The final model is created to be accurate for all modes, to be able to analyse modal shifts towards the bike and e-bike effectively. While previous chapters were mostly focused on contributing to the first goal of the research, this chapter is used to support its second goal: to find significant factors that influence the modal split in order to support municipalities of Dutch urban areas to stimulate the use of the (e-)bike. First, a summation is given of projects or policies that municipalities can implement. Then, the final nested logit model of this research is used to assess the impact of these projects or policies on the modal split if that is possible.

7.1. POSSIBLE IMPLEMENTATIONS

Implementations from the municipality that can stimulate a modal shift towards (e-)bikes in urban areas can be related to the network or policies. Related to the network, municipalities can adapt the street network of the car or the (e-)bike by adding or removing links. To further facilitate the (e-)bike users, bicycle parking facilities can be added along the cycling network. Related to policies, speed limits and traffic control systems can be adjusted to give benefits to (e-)bike users. Moreover, parking restrictions can be changed to discourage the use of the car, and the accessibility of the public transport modes can be changed to stimulate sustainable mode choices. A last policy measure can be to further stimulate e-bike ownership. These measures are based on the implementations of the municipalities of Rotterdam (2022) and Amsterdam (2022). A summation of options is given for both categories of changes that could possibly stimulate (e-)bicycle use to then assess if the final model can be used to analyse its impact or not.

Network

- Adapt network of the (e-)bike
- Adapt network of the car
- Add bicycle parking facilities

Policy

- Change speed limits
- Change traffic control systems
- Adjust parking restrictions
- Change accessibility of public transport
- Stimulate e-bicycle ownership
- Stimulate shared mobility

From the network measures, changes in the street network for the (e-)bike and the car can be modelled. The factors of street density and travel distance will then be influenced. However, the bicycle parking facilities cannot be analysed as it is not a factor in the model. From the policy measures, the speed limit can be changed for the car by changing the travel speed in the model. Traffic control systems could be effective to stimulate (e-)bicycle use by prioritizing these modes at intersections. However, this is not a measure that can be changed in the input of the model and thus cannot be analysed. The parking restrictions can partly be modelled, by changing the amount of car parking in a zone. However, cost-related parking restrictions cannot be included. Changes in the accessibility of public transport can be analysed by changing the catchment area of train stations and BTM stops in a zone and adjusting the frequency of BTM modes. Stimulating e-bicycle ownership can be analysed by increasing the availability of e-bikes in the model. Finally, shared mobility would be able to

stimulate more sustainable modes such as shared e-bikes, which would reduce the number of cars in the city. However, the impact of this measure cannot be modelled with the factors that are in the final model.

The measures that are possible to model are assessed by making assumptions about possible real-life changes in the network or policies and calculating the elasticity of the variable corresponding to these changes.

7.2. APPLYING THE FINAL MODEL

This section shows how the model can be used practically in Dutch urban areas with trips up to 15 kilometers for the possible implementations that are mentioned in the previous section. It will estimate the outcome of hypothetical situations to show how the final model can be used. The base case mentioned in some implementations is based on the input in Appendix F.

Adapting the network

Positive changes for the (e-)bike in the street network can be made by increasing the bicycle network density by adding links to also decrease the cycling travel distance and closing fast car routes to increase travel distance for the car.

Increasing the bicycle network density is often not done by large changes at the same time. The short-term impact is thus assessed by increasing the network density by only 1%, but this is not expected to influence the mode choice significantly. The long-term impact could be when a municipality increases the network density through multiple projects. It is assumed that the network density could then be increased by 10%. These two options would mean that for example in the network between Rotterdam Noord and Rotterdam Zuid, the short-term project would increase the bicycle path length by 5 km, and the long-term project would increase the length by 50 km. These added links would also decrease the travel distance for the bike for certain O-D pairs. The hypothetical situation is formed in which the short-term impact decreases the average travel distance for O-D pairs in the implemented area with 10% and the long-term impact decreases the average travel distance in the larger implemented area with 25%. These changes can be different for other travel distances. These changes are therefore assessed for a base distance of 5 km and 10 km. It would mean that the changes in travel distance for 5 km would be 4.5 km and 3.75 km, and the changes in travel distance for 10 km would be 9 km and 7.5 km. It is expected that the short-term measure will not have a large impact, while the long-term measure is expected to have a larger significant difference in mode shares.

Table 7.1: Adapting (e-)bicycle network

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	30.87%	3.49%	56.91%	8.36%	0.38%
Adapting network 5 km short-term	33.97%	3.85%	54.11%	7.73%	0.35%
Adapting network 5 km long-term	39.10%	4.41%	49.41%	6.78%	0.30%
	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	17.69%	1.75%	62.69%	17.72%	0.15%
Adapting network 10 km short-term	21.12%	2.16%	60.19%	16.39%	0.14%
Adapting network 10 km long-term	27.71%	2.96%	55.09%	14.12%	0.12%

Table 7.1 first shows a base case in which the travel distance is 5 or 10 kilometers for the (e-)bike, and the other factors are based on Appendix F. The other rows show that the changes in the bicycle street network are expected to increase the bicycle shares. The resistance for the bike is higher at a larger travel distance and thus the initial shares for the bike and e-bike are lower at a 10-kilometer travel distance. For the long-term measure of the initial 10-kilometer travel distance, the difference is expected to be over 10 percentage points with the base case. At the initial 5-kilometer travel distance, this difference is less than 9 percentage points. A similar change in travel distance and street density is thus a bit more valuable at a larger initial travel distance.

Increasing the car travel distance can be achieved by closing short car routes through the city to force car drivers to drive around the city. The travel speed could also be lower, but this is not always necessary as roads around the city can have a higher speed limit. Closing links will decrease the street density of the car a little.

However, this factor is not significant in the final model and it is thus not included in this measure. A possible hypothetical measure is assessed in which the travel distance is increased by 10% or 25%. To remain focused on the changes in (e-)bicycle shares and to be able to compare this measure with the changes in the bicycle network, these changes are assessed for an initial travel distance for the (e-)bike of 5 km and 10 km. The initial average car travel distance is then 5.3 km and 10.5 km. These two values are then increased with 10% and 25% travel distance.

Table 7.2: Adapting car network

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	30.87%	3.49%	56.91%	8.36%	0.38%
Adapting network 5 km (+10%)	31.91%	3.60%	55.46%	8.64%	0.39%
Adapting network 5 km (+25%)	33.51%	3.76%	53.24%	9.07%	0.41%

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	17.69%	1.75%	62.69%	17.72%	0.15%
Adapting network 10 km (+10%)	18.93%	1.87%	60.08%	18.95%	0.16%
Adapting network 10 km (+25%)	20.92%	2.06%	55.89%	20.95%	0.18%

Table 7.2 again first shows a base case in which the travel distance is 5 or 10 kilometers for the (e-)bike, and the other factors are based on Appendix F. The other rows show that removing a link is expected to be less influential than adding a link. Decreasing the travel distance of the bike by 10% (short-term) probably has about the same effect on the bike and e-bike as increasing the travel distance of the car by 25%. However, removing a link in the car network and adding a link to the bicycle network probably has a similar effect to the car share for the initial 10-kilometer travel distance. At 5 kilometers, adding to the bicycle network is probably more effective than removing a link in the car network for the car.

Implement policies

Instead of adapting the bicycle and car network, changing policies could also increase the modal shift towards the (e-)bicycle. One measure is to decrease the average speed of cars in the network. Discussions to decrease the standard speed of cars from 50 km/h to 30 km/h in cities are already being held (NOS, 2021), and this implementation can be assessed by the model.

Table 7.3: Decreasing car travel speed

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Car travel speed 50 km/h	35.96%	4.09%	53.66%	5.84%	0.45%
Car travel speed 30 km/h	36.01%	4.09%	53.61%	5.84%	0.45%

Table 7.3 shows that decreasing the travel speed is not expected to lead to very large changes in the modal split. The model is thus not sensitive to these changes and it could mean that the implementation of decreasing the car speed in the cities is not useful for a large change in the modal split.

The next policy measure is to implement parking restrictions, which can be done by short-term or long-term decisions. A short-term decision is to increase the area of paid car parking zones. On average, 15% of the area of an urban zip code has paid parking zones. However, in an urban zip code, there is often either no paid parking or almost the whole zip code is a paid parking zone. Therefore, the changes from 0%, 10%, 50%, 90%, and 100% are assessed to see how this influences the bicycle mode choice compared to each other. It is expected that it would lead to a significant difference in the modal split when the paid car parking is present in more than half of the area of the location.

Table 7.4: Increasing paid car parking zones

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Car parking origin 0%	34.36%	3.77%	56.26%	5.22%	0.39%
Car parking origin 10%	35.37%	3.98%	54.61%	5.62%	0.43%
Car parking origin 50%	39.52%	4.84%	47.65%	7.48%	0.51%
Car parking origin 90%	43.55%	5.72%	40.48%	9.83%	0.42%
Car parking origin 100%	44.51%	5.93%	38.69%	10.50%	0.37%
	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Car parking destination 0%	34.10%	3.78%	56.35%	5.41%	0.36%
Car parking destination 10%	35.29%	3.98%	54.63%	5.68%	0.42%
Car parking destination 50%	40.25%	4.82%	47.50%	6.89%	0.54%
Car parking destination 90%	45.43%	5.68%	40.24%	8.26%	0.39%
Car parking destination 100%	46.71%	5.88%	38.45%	8.63%	0.33%

Table 7.4 shows that the implementation of car parking zones is expected to have more effect for the bike at the destination than at the origin. For the e-bike, this effect is about equal. The model increases the bicycle and e-bicycle share significantly compared to other measures if the location is changed from no paid parking to completely paid parking. The effect of paid parking is also expected to be considerable when 50% of the location is paid parking. Thus, this measure is likely to be effective for municipalities to use in their policy-making.

A long-term decision in parking restrictions is to try to decrease car ownership. Discouraging owning multiple cars can be done by removing parking places, or reducing the number of parking spaces in new housing projects (Schouten, 2019). These types of policies are modelled by assuming they will be effective and decrease the average car ownership. It is expected that it would not change the modal split at small changes, but that it would have a significant effect at a 50% decrease. The average in the Netherlands is that a household has 0.85 cars per license. The effects are assessed when the car ownership would decrease by 10%, 25%, and 50%.

Table 7.5: Decreasing car ownership

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Car ownership -10%	36.57%	4.16%	52.88%	5.92%	0.47%
Car ownership -25%	37.24%	4.25%	51.97%	6.04%	0.50%
Car ownership -50%	37.42%	4.34%	51.56%	6.18%	0.50%

Table 7.5 shows that the effect compared to the base case is actually not expected to be very significant for any of the changes. The car can therefore be discouraged in city planning, but it probably should then lead to very large changes in car ownership will it have any significant effect.

Changing public transport accessibility is often not done as a policy to increase bicycle use, because public transport is also viewed as an environmentally friendly travel option. However, changes in these factors are possibly influencing bicycle shares and are thus also assessed on their impact. An increase in accessibility would decrease the bicycle mode share. However, decreasing the accessibility would mean removing BTM or train stops. These kinds of policies are not performed in practice and it is thus not analysed.

An increase in frequency would increase the bicycle mode share in the modal split. This option is therefore analysed as it is beneficial for all modelled sustainable modes: (e-)bicycle and public transport modes. The frequency of the BTM modes is on average 39 buses/trams/metros per hour in an urban zip code. The effects are modelled in which the frequency is increased by 25%, 50%, 75% or 100% to see at which stage the impact is significant. It is expected that it actually will not be a significant change for the bike, but mostly a change for the BTM mode choice.

Table 7.6: Increasing the frequency of BTM

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Frequency of BTM origin +25%	35.96%	4.09%	53.46%	6.03%	0.46%
Frequency of BTM origin +50%	36.02%	4.10%	53.16%	6.25%	0.47%
Frequency of BTM origin +75%	36.08%	4.11%	52.85%	6.47%	0.49%
Frequency of BTM origin +100%	36.14%	4.12%	52.53%	6.70%	0.50%
	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	35.89%	4.08%	53.76%	5.82%	0.45%
Frequency of BTM destination +25%	36.06%	4.10%	53.38%	6.01%	0.45%
Frequency of BTM destination +50%	36.23%	4.11%	53.00%	6.20%	0.46%
Frequency of BTM destination +75%	36.40%	4.13%	52.61%	6.39%	0.47%
Frequency of BTM destination +100%	36.56%	4.14%	52.23%	6.59%	0.48%

Table 7.6 shows that indeed the bicycle mode share is not expected to change much compared to the base case. Unexpectedly, the BTM mode share also did not change significantly. The model is thus not sensitive to changes in this factor and it could mean that increasing the frequency of BTM modes is not a useful policy if the goal is to induce a modal shift towards sustainable modes.

In the last years, e-bike sales have grown rapidly (RTL nieuws, 2021). To analyse this effect on the modal split, the availability of the e-bike is increased. It was found that on average 13% of the people own an e-bike in 2018 (12%) and 2019 (14%). By changing this parameter, the effects of this trend can be seen on the modal split. In 2021, the estimated percentage of e-bikes has grown to 17% and it is expected to grow even further. Based on this trend, in 2025 there could be an estimated percentage of 24%, in 2030 an estimated percentage of 32% and in 2040 an estimated percentage of 50%. To see how this trend would affect the modal split, the availability of the e-bike is changed to these percentages. The expectation is that availability has a lot of influence on the choice of the e-bike because the model shows that when an e-bike is available, it is often chosen. However, it is not certain that this trend will continue as it is seen in the last few years, or that the travel behaviour as captured in the model will be the same when more people own an e-bike. The application of this change thus has to be interpreted lightly.

Table 7.7: Increasing e-bike ownership

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
<i>Base case</i>	35.89%	4.08%	53.76%	5.82%	0.45%
E-bike ownership 2021	35.15%	5.13%	53.47%	5.80%	0.45%
E-bike ownership 2025	33.50%	7.30%	53.03%	5.73%	0.44%
E-bike ownership 2030	31.65%	9.74%	52.51%	5.66%	0.43%
E-bike ownership 2040	27.23%	15.46%	51.41%	5.47%	0.42%

Table 7.7 shows that the e-bike indeed is expected to be more often chosen compared to the base case when more e-bikes are generally available. It is surprising to see that the increasing share for the e-bike leads to the model expecting a decreasing share for the bike, while the car share remains to be close to the base case. The combined share of a bike and e-bike increases from 40% to 43% based on the model. It is thus expected to lead to an increase in cycling overall.

7.3. CONCLUSION

The conclusion of this chapter is that the model produces overall logical results in the application of the analysed implementations. It can be used practically by municipalities to form expectations in the modal shift for changes in their networks or policies. Results from this chapter are not validated, and thus no hard conclusions can be formed based on the impact of a measure. However, the model does show what can be expected to change and how sensitive the model reacts to a measure to see if it would be effective. The model shows that the most effective measures to increase cycling are expected to be decreasing travel distances and discouraging the use of the car by increasing parking restrictions and travel distances for the car. Increasing the mode share of the e-bike is expected to mostly be accomplished by increasing the ownership of the e-bike.

8

CONCLUSION

This chapter presents the conclusion of the research. It will provide an overview of the research steps, an answer to the four sub-research questions, and finally, the main research question will be answered.

The objective of this research is to find significant factors for the mode choice to create a model that predicts a more accurate modal split with the bike, e-bike, car, BTM, and train that includes multiple factors for urban areas in the Netherlands. A model is assumed to be accurate when it can estimate the modal split shares within 3 percentage points of the actual modal split. Moreover, its goal is to find significant factors for the mode choice that influence the modal split, in order to support municipalities of Dutch urban areas to stimulate the use of the (e-)bike. To achieve this, the research consists of 4 phases: a literature review, data collection & analysis, modelling, and validation. The literature review is used to have a selection of factors known to possibly influence the mode choice for the car, bike, e-bike, and public transport. Data is gathered from ODIN with additions from OSM, NDOV, CBS, RDW, and CROW. The factors for which data can be found are analysed and expectations are formed for the modelling phase. Main variables, interaction effects, and quadratic components are modelled and tested on their significance and usefulness for the goal of the research. Finally, the final model is chosen and validated. The validation consists of an internal validation and an external validation which is based on another location. The model is also tested on its practical use for municipalities for the second goal of the research. The application shows that it can measure policy and network changes so that municipalities can have an insight into possible modal shifts when introducing new network or policy measures.

The first subquestion is: *Which factors known in literature could influence the choice for the (e-)bike, car, and public transport in urban areas?*

The factors found in literature that are known to influence the mode choice are categorized into three groups: spatial characteristics, characteristics of people, and transport journey characteristics. Spatial and transport journey characteristics are known to influence the mode choice more effectively (Ton *et al.*, 2019). The characteristics of people often have more factors that are significant, although they influence the mode choice less. Exceptions are the habit and attitude factors, which can also influence the mode choice effectively. However, the attitudes are too time-consuming to model and thus not included in the research. Table 8.1 shows the factors found in literature per category. The factors shown in grey are the ones that are not included in this research, because of a lack of data or because it would be too time-consuming to gather the data.

Table 8.1: Significant factors based on literature

Spatial characteristics	Characteristics of people	Travel journey characteristics
Separate bicycle lanes	Population density	Travel distance
Access to public transport	Gender	Travel time
Frequency of public transport	Age	Travel costs
Access to bicycle parking	Occupation	Travel group size
Density of network	Education	Travel motives
Connectivity of network	Driver's license	Departure day
Car parking zones	Wealth	Departure time
Comfort level	No. of household members	
Safety level	No. of cars per driving person	
Season	E-bike availability	
Weather	Health	
Hilliness / Slopes	Habit	
	Self-efficacy	
	Subjective norm	
	Attitudes	

The second subquestion is: *What function including which factors will describe the utility to the mode choice for the (e-)bike, car and public transport in urban areas?*

The final model is based on a nested logit with a nest for the bike and e-bike. The formula for calculating the utility of mode i for variables m is as follows:

$$Utility_i = ASC_i + \sum_m \beta_{i,m} * x_{i,m} \quad (8.1)$$

The utility for each mode is the alternative specific constant plus the sum of the estimated betas multiplied by the variable inputs. The nested logit model contains 13 main variables, 3 quadratic components, and 4 interaction effects. The beta estimates of the final model can be found in Table 5.12. The utility is only calculated with the variables from which the beta estimates are significant. The nest of the bike and e-bike is included in the calculation of the probabilities of the modes. For this, the utility of the nest is also needed. The formula for the nest utility of the modes i and j for individual n is:

$$Utility_{nest} = \frac{1}{2} * (Utility_i + Utility_j) + \frac{1}{\beta_{nest}} * \log \sum_n (e^{\beta_{nest} * Utility_{i,n}} + e^{\beta_{nest} * Utility_{j,n}}) \quad (8.2)$$

The calculation of the probability of an alternative *outside* the nest is similar to the calculation of the probability of an alternative in an MNL model. The probability that mode i is chosen for individual n given the utilities of the alternatives outside the nest and the nest utility equals the following formula:

$$P_{n,i} = \frac{e^{V_{n,i}}}{\sum_j e^{V_{n,j}}} = \frac{e^{\sum_m \beta_m * x_{i,m} + \epsilon_n}}{\sum_j e^{\sum_m \beta_m * x_{j,m} + \epsilon_n}} \quad (8.3)$$

The probability of an alternative *in* the nest is the probability of the nest (B) times the probability of the alternative in the nest. The formula for the probability that mode i is chosen for individual n is:

$$P_{n,i} = P[n, i | n, i \in B(n, i)] * P[n, i \in B(n, i)] \quad (8.4)$$

With these formulas, the utility of the alternatives and of the nest can be calculated, with which the probabilities of the modes can be calculated for each trip.

The third subquestion is: *How much do the analysed factors influence the modal split of urban areas?*

The model is not sensitive to changes in the beta estimates. The uncertainty analysis shows that the most influential factor is the travel distance. The change in modal split can then be about 25 percentage points of the choice for the bike or car. Another influential factor is the age, which can influence the modal split with about 10 percentage points for the bike and car. However, this factor is correlated to the alternative specific constant and can therefore show these deviations from the mean modal split because of the correlation. Having a driver's license can influence the modal split by changes of about 10 percentage points for the car and 5 percentage points for the bike. Finally, car parking can influence the modal split by changes of about 5 percentage points for the car and 3 percentage points for the bike. The changes in public transport are not large. Mostly this is influenced by the travel distance which can be about 4 percentage points for the BTM and 7 percentage points for the train. The changes in e-bike mode choice can be influenced by the travel distance with changes of about 2 percentage points. The e-bike choice is mostly influenced by the availability of the e-bike. If an e-bike is available, it is then often also the chosen option and thus increases its share in the modal split.

The fourth subquestion is: *How accurate can the model predict the modal split of urban areas?*

The internal validation estimates the modal split for the data also used in modelling. The best way of predicting the modal split is to use the average of the calculated probabilities for each mode. The difference between the actual modal split and the calculated modal split is then at most a deviation of 1.6 percentage points for the car. Estimating the modal split for an area similar to the average Dutch urban area (based on ODiN) is thus very accurate. The external validation estimates the modal split for data of Haarlem, which is not used in modelling. Estimating the modal split for this data shows larger differences from the actual modal split. The largest difference is a deviation of 11.6 percentage points for the bike. Haarlem is an area in which the data shows that cycling is on average more preferred. By changing the ASC of the bike to a higher value, this higher preference will be included in estimating the modal split. The results are then much more similar to the actual modal split. By increasing the ASC of the bike from 0.76 to 1.40, the largest difference is a deviation of 1.8 percentage points for the e-bike. The conclusion of the validation is thus that it can accurately predict the modal split in average Dutch urban areas. In non-average Dutch urban areas, an analysis of the preferences of the modes needs to be conducted to change the ASC's in the model to also be able to generate an accurate modal split.

The main research question is: *What mode choice model estimates an accurate modal split for Dutch urban areas including multiple influencing factors, which is also useful for analysing changes to (e-)bicycle policies and networks?*

The mode choice can be modelled by using the formula for the probability for each mode in Equation 2.2. The utility for each mode needed for that equation can be calculated with Equation 8.1. The beta estimates needed for the utility can be found in the results of the nested logit model in Table 5.12. Using this model, changes in (e-)bicycle policies and networks can be analysed for urban trips up to 15 kilometers by altering the input of variables that are possible to change in urban areas. These are the street density, travel distance, travel speed, car parking zones, frequency of BTM, car ownership, and e-bike ownership. Based on the model, it is expected that the largest changes in modal split can be found by reasonably altering the input of the travel distance for the bike and car, and the area of the paid car parking zones. In the long-term, changes in e-bike ownership are also expected to highly influence the modal split.

To conclude, the final model from Table 5.12 can thus predict the modal split accurately of Dutch urban areas for trips up to 15 kilometers, and it can also be used by municipalities to get an insight into the mode choice based on certain changes in their (e-)bicycle policies or networks. The contribution of this thesis is thus that it provides a mode choice model which has been made more accurate for calculating the bicycle share in Dutch urban areas with trips up to 15 kilometers. Moreover, it is to my knowledge the first model that can be used to model the mode choice in Dutch urban areas with an e-bike share separately. The mode choice model can be used in practice for a module in the Fietsmonitor to better estimate the modal split for O-D pairs, which in turn helps to advise municipalities on new bicycle policies and networks.

9

DISCUSSION

In this chapter, the limitations to this research are discussed and the recommendations for future research are shown. The limitations discuss which aspects of the research need to be considered when reading the results. The recommendations for future research are based on the results found of the factors used in the modelling phase, and on the limitations that are discussed in this chapter.

9.1. LIMITATIONS TO THE RESEARCH

This section will discuss the limitations to the research. First, it will discuss the assumptions made. Second, the limitations to the data collection. And third, the limitations to the discrete choice models. Each part will explain the known limitations and their effect on the results of this thesis.

Limitations to assumptions

The limitations to assumptions include the assumption that mode choice is an independent choice based on the four-stage transport model, and the assumption that travel behaviour did not change because of covid-19.

This thesis focuses solely on the third step in the four-stage transport model, which is the mode choice. Literature shows that the mode choice could also be modelled simultaneously with the trip distribution in the four-stage transport model. This would mean that a person would consider their destination at the same time as their mode choice. An example is someone that wants to go shopping when living in Delft. They can choose to go to the centre of Delft to shop and take the bike or go to the city of Rotterdam by train. Both the destination and the transport mode are then considered simultaneously. A limitation to this thesis is to not include these considerations in the paper and thus assuming that these choices are not made at the same time. However, this phenomenon is usually only the case for leisure motives. For travel motives such as commuting and education, the destination is already fixed and then the simultaneous choice does not play a role. The results are not affected by this assumption, because the results of this research focus on explaining why a choice is taken when the destination is already fixed.

Another assumption made is that the present travel behaviour is the same as it was in 2018 and 2019. In the last two years, travel behaviour was severely influenced by the covid-19 pandemic. Private transport was preferred over public transport due to hygienic considerations, and working from home led to less congestion on the road. This period of time may have changed the way people look at travelling nowadays. Preferences might have changed, and these changes in behaviour are not captured in the results of this research.

Limitations to data

Data collection is a very important part of the research. It eventually determines the results, and therefore choices made in this phase need to be discussed. The limitations to the data collection include the representativeness of the dataset, the calculation of the infrastructural factors, and the external validation data.

The kind of people represented in ODIN forms a limitation to the research. There are differences between the kind of people in the country and the ODIN dataset. The share of young people is hugely undervalued. This age group often takes the bike because of the Dutch culture, and these trips are thus not included. The effect this can have on the outcome is that the preference for the bike is lower than what could be true. Moreover, a large difference is found between the full-time workers (higher percentage in ODIN) and the housekeepers (lower percentage in ODIN). It could lead to the model more accurately predicting the mode choice for full-time workers, and less accurate for people doing their own housekeeping. This is also true for the education level hbo/university and the higher wealth groups, where these are much higher than average. The model would thus predict better towards people with an hbo/university degree and higher wealth households.

Besides the limitations to the kind of people in the dataset, the calculation of the infrastructural factors is a limitation. The travel distance is calculated via a shortest path algorithm in python, which is accurately calculated for the car and the bike, but less accurate for the public transport modes. In the calculation of the shortest route, it is assumed that at each stop a transfer is possible if another line crosses at that stop or station. Moreover, no access or egress modes are used, but it is compensated by adding two times the distance as-the-crow-flies to the nearest train station. The travel distances of the BTM and the train are therefore often overestimated and it can cause deviations in the accuracy of predictions. This choice is justified by the fact that the focus of this research is not on modelling public transport modes. In the process of the data collection, it was assumed that this level of detail for public transport would not cause too large changes in the model outcome. In the outcomes of the final model, it can be seen that public transport is indeed predicted accurately enough for the goal of this research. However, if for the use of the model it is desired that the public transport modes are calculated more accurately, this limitation should be considered.

Another limitation based on the data of the research is not being able to validate the model with data from other years. The external validation now exists of a cross-sectional external validation, where the model is validated for another location. Another form of external validation is longitudinal external validation, where the model is validated with other years of data. The years of data used are ODIN 2018 and 2019. After these years, the travel behaviour is assumed to be different because of the covid-19 pandemic. Before these years, ODIN was called OVIN and the data gathering was performed differently. No relevant data is thus available to assess the longitudinal external validation performance, while this would have been useful for municipalities to be able to assess the usefulness of the model for analysing the modal split in future scenarios.

Limitations to discrete choice models

There are multiple types of discrete choice models, all with their advantages and disadvantages. It has been found that the mixed logit model would be the most accurate way to determine a model (Pfaff, 2019). However, by being limited in computational power, no random parameters could be estimated in this research. Every person has their unique preferences and tastes and including these different tastes could have improved the model. This difference in taste is for example visible in the external validation of Haarlem, in which the preference for the bike is different than what was found to be the average preference in Dutch urban areas. It shows that there are probably different tastes for the alternative specific constants. However, including random parameters is also known to lead to overfitting. Not having included the random parameters is thus a limitation, but it is not certain that it would improve the applicability of the model.

9.2. RECOMMENDATIONS FOR FUTURE RESEARCH

The recommendations consist of points based on the assumptions of the research, the used techniques and methods, and the results of factors.

Recommendation based on assumptions

Based on the assumptions of the research, further research could model the effect of the simultaneous choice of trip distribution and mode choice from the four-stage transport model in Dutch urban areas. The model is now made to be used in the third step of the four-stage model, which is to estimate solely the mode choice. Combining the second and the third step would also include the trip destination choice. The same steps could be taken as in this research, only the origin and destination would also determine the results of the parameter estimates from the final model. It is recommended to research the influence of the destination choice in combination with the mode choice to get further insights into the behaviour of Dutch urban travel.

Moreover, it is assumed in this research that the years of 2018 and 2019 represent normal travel behaviour. However, travelling in the last two years has been very different, and it is possible that this permanently changed our views on travel and thus also changed our travel behaviour. It is thus advised to analyse the current travel behaviour after the covid-19 pandemic, and if this would influence the results of this research.

Recommendation based on techniques and methods

Related to the modelling technique, it is recommended that the factors included in the final model are also modelled in a mixed logit model that contains the nest, panel structure, and random taste parameters. It is advised to analyse for which factors people could have a difference in taste, and re-estimate a mixed logit

model with enough computational power. The results should be compared to the nested logit to see if it indeed fits better and if the model is still well applicable to other situations.

Because the data does not accurately represent the Dutch population, it could lead to the final model not predicting as accurate for types of people that are less present in the data. This could be overcome by studying the use of the weight factors provided by ODIN based on the household type, trip type & travel type, and include this in the modelling step of this research.

The application shows how the model could be used practically. The method used is to assess the elasticities of the variables. However, there is no data to validate if the changes to bicycle policies or networks are accurate. No conclusions can be taken from the results following from the assessed measures, which would have been valuable for municipalities. It is therefore advised to gather data from before and after network or policy measures to validate the predicted modal split of the final model for changes in network or policy measures.

Recommendations based on results

The experience gained with the results of factors lead to a couple of recommendations. The first is bicycle parking, which is not a factor in the final model because the data included mostly locations close to public transport stations. It is therefore advised to further investigate the influence of bicycle parking with a more detailed data source of the bicycle parking facilities in the Netherlands. Another factor is the street connectivity which has been calculated as the number of edges divided by the number of nodes. It has shown that this is not the best representation of street connectivity, as it did not show the results that were expected. It resulted in a negative relation towards choosing a mode, which actually has also been found by [Piatkowski and Marshall, \(2015\)](#). It however did not explain this negative relation. Only that it could have been related to the fact that the spatial characteristics are calculated on a zip code level. It is advised to research other definitions for street connectivity, such as the number of cul-de-sacs or the number of intersections. Examples of papers using such definitions are: [Nello-Deakin and Harms, \(2019\)](#), [Wati and Tranter, \(2015\)](#) and [Heesch et al., \(2014\)](#). The number of intersections could also be used to increase the accuracy of the calculation of the travel speed. For the bike and e-bike, the average travel speed is 18 km/h and 25 km/h. These values are also used in the calculation of the travel time in this research. However, to be able to include travel speed in the model, there should be variation in the travel speed between trips. This may be influenced by intersections, but other infrastructural characteristics could also be of influence. Further research into this factor is thus advised.

Not only the travel speed but also the calculation of the travel time has its limitations. This paper shows that the travel time would have been a better fit with the model, although it is less accurately calculated. Thus the travel distance is used in the final model, although the accuracy of the travel distance for the public transport modes is also not optimal. No access and egress modes are included and a real shortest path is difficult to determine without knowing transfer options and schedules. It is therefore advised to find a better method to calculate the travel time and travel distance of public transport modes, and possibly include infrastructural characteristics such as intersections to be able to calculate the travel time of the (e-)bike and the car more accurately. This can then be used in further research to estimate the impact of travel time on mode choice in urban areas.

BIBLIOGRAPHY

- [1] KiM, *Cycling Facts 2018*, (2018).
- [2] Bicycle Dutch, *Dutch cycling figures*, (2018).
- [3] Nederland Fietsland, *LF-routes*, (2021).
- [4] CBS, *Wijk- en buurtkaart 2021*, (2021).
- [5] Rijkswaterstaat, *Cycling and Dutch national infrastructure book*, (2020).
- [6] G. Santos, H. Maoh, D. Potoglou, and T. V. Brunn, *Factors influencing modal split of commuting journeys in medium-size European cities*, *Journal of Transport Geography* **30**, 127 (2013).
- [7] J. Holmgren and P. Ivehammar, *Mode choice in home-to-work travel in mid-size towns: The competitiveness of public transport when bicycling and walking are viable options*, *Transportation Research Procedia* **48**, 1635 (2020).
- [8] P. Rietveld and V. Daniel, *Determinants of bicycle use: do municipal policies matter?* *Transportation Research Part A: Policy and Practice* **38**, 531 (2004).
- [9] J. Pucher and R. Buehler, *Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany*, *Transport Reviews* **28**, 495 (2008).
- [10] E. Heinen, B. V. Wee, and K. Maat, *Commuting by Bicycle: An Overview of the Literature*, *Transport Reviews* **30**, 59 (2010).
- [11] A. D. Witte, J. Hollevoet, F. Dobruszkes, M. Hubert, and C. Macharis, *Linking modal choice to motility: A comprehensive review*, *Transportation Research Part A: Policy and Practice* **49**, 329 (2013).
- [12] Y. Wang, C. K. Chau, W. Y. Ng, and T. M. Leung, *A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods*, *Cities* **50**, 1 (2016).
- [13] Witteveen+Bos, *FietsMonitor*, (2021).
- [14] Witteveen+Bos, *FietsMonitor Haarlem*, (2021).
- [15] CROW, *Stimuleren Fietsgebruik*, (2021).
- [16] L. Yang, C. Ding, Y. Ju, and B. Yu, *Driving as a commuting travel mode choice of car owners in urban China: Roles of the built environment*, *Cities* **112** (2021), 10.1016/j.cities.2021.103114.
- [17] J. N. Gonzalez, J. Perez-Doval, J. Gomez, and J. M. Vassallo, *What impact do private vehicle restrictions in urban areas have on car ownership? Empirical evidence from the city of Madrid*, *Cities* **116** (2021), 10.1016/j.cities.2021.103301.
- [18] L. J. Basso, M. Navarro, and H. E. Silva, *Public transport and urban structure*, *Economics of Transportation* **28** (2021), <https://doi.org/10.1016/j.ecotra.2021.100232>.
- [19] F. Schneider, A. F. Jensen, W. Daamen, and S. Hoogendoorn, *Bicycle accessibility: What can we learn from best-practice examples?* *Journal of Transport Geography* (2020).
- [20] M. J. N. Keijer and P. Rietveld, *How do people get to the railway station? The dutch experience*, *Transportation Planning and Technology* **23**, 215 (2000).
- [21] A. Barberan, J. D. A. E. Silva, and A. Monzon, *Factors influencing bicycle use: a binary choice model with panel data*, *Transportation Research Procedia* **27**, 253 (2017).

- [22] J. Scheiner, *Interrelations between travel mode choice and trip distance: trends in Germany 1976–2002*, [Journal of Transport Geography](#) **18**, 75 (2010).
- [23] S. Müller, S. Tscharaktschiew, and K. Haase, *Travel-to-school mode choice modelling and patterns of school choice in urban areas*, [Journal of Transport Geography](#) **16**, 342 (2008).
- [24] M. Sabir, *Weather and travel behaviour*, Tinbergen Institute/Thela Thesis (2011).
- [25] E. Heinen, K. Maat, and B. V. Wee, *The effect of work-related factors on the bicycle commute mode choice in the Netherlands*, [Transportation](#) **40**, 23 (2012).
- [26] Exsilio Solutions, *Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures*, (2016).
- [27] J. de Dios Ortúzar and L. G. Willumsen, *Modelling Transport* (John Wiley & Sons, 2011).
- [28] R. Buehler, *Determinants of bicycle commuting in the Washington, DC region: The role of bicycle parking, cyclist showers, and free car parking at work*, [Transportation Research Part D: Transport and Environment](#) **17**, 525 (2012).
- [29] K. Wati and P. J. Tranter, *Spatial and socio-demographic determinants of South East Queensland students' school cycling*, [Journal of Transport Geography](#) **47**, 23 (2015).
- [30] G. Eldeeb, M. Mohamed, and A. Páez, *Built for active travel? Investigating the contextual effects of the built environment on transportation mode choice*, [Journal of Transport Geography](#) **96**, 103158 (2021).
- [31] R. Buehler and J. Dill, *Bikeway Networks: A review of effects on cycling*, [Transport Reviews](#) **36**, 9 (2015).
- [32] J. Ko, S. Lee, and M. Byun, *Exploring factors associated with commute mode choice: An application of city-level general social survey data*, [Transport Policy](#) **75**, 36 (2019).
- [33] D. P. Piatkowski and W. E. Marshall, *Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting*, [Travel Behaviour and Society](#) **2**, 166 (2015).
- [34] H. Badland, M. Knuiman, P. Hooper, and B. Giles-Corti, *Socio-ecological predictors of the uptake of cycling for recreation and transport in adults: Results from the RESIDE study*, [Preventive Medicine](#) **57**, 396 (2013).
- [35] K. C. Heesch, B. Giles-Corti, and G. Turrell, *Cycling for transport and recreation: Associations with socio-economic position, environmental perceptions, and psychological disposition*, [Preventive Medicine](#) **63**, 29 (2014).
- [36] S. Titze, W. J. Stronegger, S. Janschitz, and P. Oja, *Association of built-environment, social-environment and personal factors with bicycling as a mode of transportation among Austrian city dwellers*, [Preventive Medicine](#) **47**, 252 (2008).
- [37] M. Orozco-Fontalvo, A. Arévalo-Támara, T. Guerrero-Barbosa, and M. Gutiérrez-Torres, *Bicycle choice modeling: A study of university trips in a small Colombian city*, [Journal of Transport & Health](#) **9**, 264 (2018).
- [38] M. Kamargianni and A. Polydoropoulou, *Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior*, [Transportation Research Record](#) , 151 (2013).
- [39] M. Scorrano and R. Danielis, *Active mobility in an Italian city: Mode choice determinants and attitudes before and during the Covid-19 emergency*, [Research in Transportation Economics](#) **86** (2021), [10.1016/j.retrec.2021.101031](#).
- [40] H. Charreire, C. Roda, T. Feuillet, A. Piombini, H. Bardos, H. Rutter, S. Compernelle, J. D. Mackenbach, J. Lakerveld, J. M. Oppert, and et al., *Walking, cycling, and public transport for commuting and non-commuting travels across 5 European urban regions: Modal choice correlates and motivations*, [Journal of Transport Geography](#) **96**, 103196 (2021).
- [41] M. Hallberg, T. K. Rasmussen, and J. Rich, *Modelling the impact of cycle superhighways and electric bicycles*, [Transportation Research Part A: Policy and Practice](#) **149**, 397 (2021).

- [42] K. Mattisson, A. O. Idris, E. Cromley, C. Håkansson, P. O. Östergren, and K. Jakobsson, *Modelling the association between health indicators and commute mode choice: a cross-sectional study in southern Sweden*, *Journal of Transport & Health* **11**, 110 (2018).
- [43] M. J. Nieuwenhuijsen and H. Khreis, *Car free cities: Pathway to healthy urban living*, *Environment International* **94**, 251 (2016).
- [44] S. Nello-Deakin and L. Harms, *Assessing the relationship between neighbourhood characteristics and cycling: Findings from Amsterdam*, *Transportation Research Procedia* **41**, 17 (2019).
- [45] M. Kroesen, *To what extent do e-bikes substitute travel by other modes? Evidence from the Netherlands*, *Transportation Research Part D: Transport and Environment* **53**, 377 (2017).
- [46] L. M. Wen and C. Rissel, *Inverse associations between cycling to work, public transport, and overweight and obesity: Findings from a population based study in Australia*, *Preventive Medicine* **46**, 29 (2008).
- [47] S. Das, A. Boruah, A. Banerjee, R. Raoniar, S. Nama, and A. K. Maurya, *Impact of COVID-19: A radical modal shift from public to private transport mode*, *Transport Policy* **109**, 1 (2021).
- [48] D. Ton, S. Bekhor, O. Cats, D. C. Duives, S. Hoogendoorn-Lanser, and S. P. Hoogendoorn, *The experienced mode choice set and its determinants: Commuting trips in the Netherlands*, *Transportation Research Part A: Policy and Practice* **132**, 744 (2020).
- [49] D. Ton, D. C. Duives, O. Cats, S. Hoogendoorn-Lanser, and S. P. Hoogendoorn, *Cycling or walking? Determinants of mode choice in the Netherlands*, *Transportation Research Part A: Policy and Practice* **123**, 7 (2019).
- [50] S. von Behren, R. Schubert, and B. Chlond, *International comparison of psychological factors and their influence on travel behavior in hybrid cities*, *Research in Transportation Business and Management* **36** (2020), 10.1016/j.rtbm.2020.100497.
- [51] A. Rodriguez-Valencia, D. Rosas-Satizabal, R. Unda, and S. Handy, *The decision to start commuting by bicycle in Bogotá, Colombia: Motivations and influences*, *Travel Behaviour and Society* **24**, 57 (2021).
- [52] O. Simsekoglu and C. A. Klöckner, *The role of psychological and socio-demographical factors for electric bike use in Norway*, *International Journal of Sustainable Transportation* **13**, 315 (2019).
- [53] E. Heinen, K. Maat, and B. V. Wee, *The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances*, *Transportation Research Part D: Transport and Environment* **16**, 102 (2011).
- [54] H. Kim, H. Seok, S. I. You, and C. Lee, *An Empirical Analysis for Mode Choice in a Short-Distance Trip with Personal Rapid Transit*, *Journal of Advanced Transportation* **2020**, 1 (2020).
- [55] M. Winters, M. Brauer, E. M. Setton, and K. Teschke, *Built environment influences on healthy transportation choices: Bicycling versus driving*, *Journal of Urban Health* **87**, 969 (2010).
- [56] B. V. Wee, P. Rietveld, and H. Meurs, *Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time*, *Journal of Transport Geography* **14**, 109 (2006).
- [57] R. Espino, J. d. D. Ortúzar, and L. I. Rizzi, *The Value of Security, Access Time, Waiting Time, and Transfers in Public Transport*, *International Encyclopedia of Transportation*, 122 (2021).
- [58] P. Astegiano, F. Fermi, and A. Martino, *Investigating the impact of e-bikes on modal share and greenhouse emissions: a system dynamic approach*, *Transportation Research Procedia* **37**, 163 (2019).
- [59] A. J. Lopez, P. Astegiano, S. Gautama, D. Ochoa, C. Tampère, and C. Beckx, *Unveiling E-Bike Potential for Commuting Trips from GPS Traces*, *ISPRS International Journal of Geo-Information* **6** (2017), 10.3390/ijgi6070190.
- [60] C. Gorenflo, I. Rios, L. Golab, and S. Keshav, *Usage Patterns of Electric Bicycles: An Analysis of the WeBike Project*, *Journal of Advanced Transportation* **2017**, 1 (2017).

- [61] D. Ton and D. Duives, *Understanding long-term changes in commuter mode use of a pilot featuring free e-bike trials*, [Transport Policy](#) **105**, 134 (2021).
- [62] S. Cairns, F. Behrendt, D. Raffo, C. Beaumont, and C. Kiefer, *Electrically-assisted bikes: Potential impacts on travel behaviour*, [Transportation Research Part A: Policy and Practice](#) **103**, 327 (2017).
- [63] C. Moser, Y. Blumer, and S. L. Hille, *E-bike trials' potential to promote sustained changes in car owners mobility habits*, [Environmental Research Letters](#) **13** (2018), 10.1088/1748-9326/aaad73.
- [64] J. Brand, s. Niels van Oort, S. Hoogendoorn, and t. Bart Schalkwijk Vervoerregio Amsterdam, *Modelling Multimodal Transit Networks Integration of bus networks with walking and cycling*, Tech. Rep. (2017).
- [65] L. Rijsman, N. Van Oort, D. Ton, S. Hoogendoorn, E. Molin, and T. Teijl, *Walking and bicycle catchment areas of tram stops: factors and insights* (2019).
- [66] Gemeente Rotterdam, [Vervoer](#), (2022).
- [67] Gemeente Amsterdam, [Volg het beleid: Verkeer en vervoer](#), (2022).
- [68] NOS, [Overall 30 km/u in bebouwde kom? Als het aan de vier grootste steden ligt wel](#), (2021).
- [69] A. Schouten, *How to reduce car ownership in neighbourhoods?* MSc Thesis Wageningen University (2019).
- [70] RTL nieuws, [Fietsenverkoop door het dak: helpt van nieuwe tweewielers elektrisch](#), (2021).
- [71] L. Pfaff, *Comparison of Multinomial Logit and Mixed Logit*, Tech. Rep. (2019).
- [72] P. Arbués, J. F. Baños, M. Mayor, and P. Suárez, *Determinants of ground transport modal choice in long-distance trips in Spain*, [Transportation Research Part A: Policy and Practice](#) **84**, 131 (2016).
- [73] E. B. Lunke, N. Fearnley, and J. Aarhaug, *Public transport competitiveness vs. the car: Impact of relative journey time and service attributes*, [Research in Transportation Economics](#) **90**, 101098 (2021).
- [74] R. F. Allard and F. Moura, *Effect of transport transfer quality on intercity passenger mode choice*, [Transportation Research Part A: Policy and Practice](#) **109**, 89 (2018).
- [75] D. A. Hensher and J. M. Rose, *Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study*, [Transportation Research Part A: Policy and Practice](#) **41**, 428 (2007).
- [76] R. Schakenbos, L. L. Paix, S. Nijenstein, and K. T. Geurs, *Valuation of a transfer in a multimodal public transport trip*, [Transport Policy](#) **46**, 72 (2016).
- [77] CBS, [StatLine](#), (2022).

A

APPENDIX A - OVERVIEW OF FOUND SIGNIFICANCE OF FACTORS

This appendix shows an overview of the literature found about the significance of factors. These are divided into the categories infrastructural factors (Table A.1), natural environmental factors (Table A.2), socio-demographic factors - personal traits (Table A.3), socio-demographic factors - household traits (Table A.4), socio-demographic factors - neighbourhood/city traits (Table A.5), psychological factors (Table A.6) and transport journey factors (Table A.7).

When a transport mode is colored purple in a table, it means that it is a significant value and grey means it is not. When it is significant, green means a positive correlation and red is a negative correlation. The color stays purple when the factor used is categorical and therefore cannot directly mean a positive or negative correlation when the value of a variable increases or decreases. The type of values relates to the results in the paper being odds ratios (OR) or parameter values (PV).

Table A.1: Overview infrastructural factors

Reference	Type of values	Length / Density of network	No. of road intersections	No. of route options	Accessibility of destinations	Separate bicycle lanes	Safety level	Satisfaction / Comfort	Connectivity	Traffic volume	Access to bicycle parking	Access to PT	Car parking
Infrastructural factors													
Badland, 2013 [34]	OR		B					B	B	B			
Buehler, 2012 [28]	OR	B									B		B
Charreire, 2021 [40]	OR											B PT	
Eldeeb, 2021 [30]	PV	C B PT											
Hallberg, 2021 [41]	PV										PT		
Heesch, 2014 [35]	OR		B	B	B	B	B	B	B				
Heinen, 2012 [25]	PV									B	B	B	B
Kamargianni, 2013 [38]	PV				B					B	PT		
Ko, 2019 [32]	OR	C B PT	C B PT			C B PT					C B PT	C B PT	C B PT
Mattisson, 2018 [42]	PV										PT		
Nello-Deakin, 2019 [44]	PV	B		B						B	B		
Orozco, 2018 [37]	PV				B								
Piatkowski, 2015 [33]	PV		B			B	B	B					
Rietveld, 2004 [8]	PV					B	B						
Santos, 2013 [6]	PV	B			B						PT		
Scorrano, 2021 [39]	PV				B								
Titze, 2008 [36]	OR					B	B	B					
von Behren, 2020 [50]	PV												C B PT
Wati, 2015 [29]	PV	B	B										
Winters, 2010 [55]	OR		B		B								

Table A.2: Overview natural environmental factors

Reference	Type of values	Rainfall	Temperature	Wind (strong)	Season (winter)	Hills/Slopes
Natural environmental factors						
Buehler, 2012 [28]	OR				B	
Hallberg, 2021 [41]	PV				B	
Heesch, 2014 [35]	OR					B
Holmgren, 2020 [7]	PV				B	
Kamargianni, 2013 [38]	PV	C PT				
Müller, 2008 [23]	PV				B PT	
Rietveld, 2004 [8]	PV	B	B	B		B
Sabir, 2011 [24]	PV	C B PT	C B PT	C B PT		
Santos, 2013 [6]	PV	B PT				
Scorrano, 2021 [39]	PV	B				
Titze, 2008 [36]	OR					B
Wati, 2015 [29]	PV					B
Winters, 2010	OR					B

Table A.3: Overview socio-demographic factors - Personal traits

Reference	Type of values	Gender (male)	Age	Occupation	Education	Driver's License	Health
Socio-demographic factors - Personal traits							
Arbués, 2016 [72]	OR	C PT	C PT				
Barberan, 2017 [21]	PV		B				
Buehler, 2012 [28]	OR	B	B				
Charreire, 2021 [40]	OR	B PT	B PT		B PT		B PT
Eldeeb, 2021 [30]	PV	C B PT	C B PT	C B PT		C B PT	
Hallberg, 2021 [41]	PV	C B PT C B PT	C B PT				
Heesch, 2014 [35]	OR			B			
Heinen, 2012 [25]	PV	B	B				
Holmgren, 2020 [7]	PV	C B PT			B PT		
Kamargianni, 2013 [38]	PV	B					
Ko, 2019 [32]	OR	C B PT	PT	C B PT	C B PT		C B PT
Kroesen, 2017 [45]	PV	EB C B PT	EB C B PT	EB C B PT	EB C B PT	EB C B PT	
Lunke, 2021 [73]	PV	PT	PT		PT		
Mattisson, 2018 [42]	PV	B PT	B PT		B PT		B PT
Nello-Deakin, 2019 [44]	PV	B	B		B		
Piatkowski, 2015 [33]	PV	B	B		B		
Rodriguez, 2021 [51]	PV	B	B				
Santos, 2013 [6]	PV		PT				
Scorrano, 2021 [39]	PV	B	B	B			
Simsekoglu, 2019 [52]	PV	EB	EB				EB
Ton, 2019 [49]	PV			PT	B		
Ton, 2020 [48]	PV	C	C PT		C B PT	C B PT	
von Behren, 2020 [50]	PV	C B PT	C B PT	C B PT	C B PT		
Wati, 2015 [29]	PV	B	B				
Winters, 2010 [55]	OR	B	B		B		

Table A.4: Overview socio-demographic factors - Household traits

Reference	Type of values	Household members	Children	Income	Bike availability	Car availability	E-bike availability
Socio-demographic factors - Household traits							
Arbués, 2016 [72]	OR			C PT			
Barberan, 2017 [21]	PV					B	
Buehler, 2012 [28]	OR			B	B	B	
Charreire, 2021 [40]	OR		B PT	B PT		B PT	
Eldeeb, 2021 [30]	PV					C B PT	
Hallberg, 2021 [41]	PV				B	C B PT	
Heesch, 2014 [35]	OR			B			
Heinen, 2012 [25]	PV				B	B	
Holmgren, 2020 [7]	PV	B PT	PT			C B PT	
Kamargianni, 2013 [38]	PV			B			
Ko, 2019 [32]	OR		C B PT	C B PT			
Kroesen, 2017 [45]	PV	EB C B PT		EB C B PT	EB C B PT	EB C B PT	EB C B PT
Lunke, 2021 [73]	PV			PT			
Mattisson, 2018 [42]	PV		B PT				
Müller, 2008 [23]	PV					C B PT	
Nello-Deakin, 2019 [44]	PV			B		B	
Piatkowski, 2015 [33]	PV	B				B	
Rodriguez, 2021 [51]	PV			B			
Santos, 2013 [6]	PV		PT				
Simsekoglu, 2019 [52]	PV			EB			
Ton, 2019 [49]	PV	B PT	PT				
Ton, 2020 [48]	PV		C		B	C PT	
von Behren, 2020 [50]	PV	C B PT		C B PT	B PT	C B PT	
Wati, 2015 [29]	PV			B	B	B	
Winters, 2010 [55]	OR			B			

Table A.5: Overview socio-demographic factors - Neighbourhood/City traits

Reference	Type of values	Population	% young people	Cars per capita	GDP per capita	Land-use mix
Socio-demographic factors -Neighbourhood/City traits						
Arbués, 2016 [72]	OR	C PT				
Badland, 2013 [34]	OR	B				B
Buehler, 2012 [28]	OR	B				
Charreire, 2021 [40]	OR					B PT
Eldeeb, 2021 [30]	PV					C B PT
Hallberg, 2021 [41]	PV	C B				
Ko, 2019 [32]	OR					C B PT
Kroesen, 2017 [45]	PV	EB C B PT				
Nello-Deakin, 2019 [44]	PV	B				B
Rietveld, 2004 [8]	PV	B	B	B		
Santos, 2013 [6]	PV	PT		C	C PT	
Titze, 2008 [36]	OR					B
Ton, 2020 [48]	PV	B PT				
Winters, 2010 [55]	OR	B				B

Table A.6: Overview psychological factors

Reference	Type of values	Psychological factors							
		Attitude - efficient	Attitude - environment	Attitude - health	Attitude - pleasant	Attitude - comfortable & safe	Self-efficacy	Subjective norm	Habit
Barberan, 2017 [21]	PV	B	B	B	B	B	B	B	
Heesch, 2014 [35]	OR	B			B		B		B
Heinen, 2011 [53]	PV	B	B	B	B	B		B	B
Heinen, 2012 [25]	PV				B			B	
Kamargianni, 2013 [38]	PV	B	B	B					
Ko, 2019 [32]	OR		C B PT	C B PT					
Piatkowski, 2015 [33]	PV				B	B			
Rodriguez, 2021 [51]	PV	B	B	B	B	B		B	
Simsekoglu, 2019 [52]	PV	EB	EB	EB				EB	
Ton, 2019 [49]	PV								C B PT
von Behren, 2020 [50]	PV				C B PT			C B PT	

Table A.7: Overview characteristics of the transport journey

Reference	Type of values	Transport journey characteristics								
		Travel distance	Travel time	Travel costs	Travel motive	Travel group size	Departure day	Departure time	Number of transfers	Out-of-vehicle time
Allard, 2018 [74]	PV		PT	PT						PT
Arbués, 2016 [72]	OR				C PT					
Barberan, 2017 [21]	PV	B	B							
Buehler, 2012 [28]	OR	B								
Hallberg, 2021 [41]	PV		C B PT		C B PT					
Heinen, 2012 [25]	PV	B								
Hensher, 2007 [75]	PV		C PT	C PT				PT		PT
Holmgren, 2020 [7]	PV		C B PT	C B PT				PT		
Kamargianni, 2013 [38]	PV		C B PT	C PT						
Ko, 2019 [32]	OR		C B PT							
Lunke, 2021 [73]	PV		PT					PT		PT
Mattisson, 2018 [42]	PV	B	C PT							
Müller, 2008 [23]	PV	C B PT								
Orozco, 2018 [37]	PV		C B PT	C B PT						
Piatkowski, 2015 [33]	PV	B								
Santos, 2013 [6]	PV			PT						
Schakenbos, 2016 [76]	PV		PT	PT						PT
Scorrano, 2021 [39]	PV		C B PT	C B PT						
Ton, 2019 [49]	PV				C B PT	B PT	B	B		
Wati, 2015 [29]	PV	B								
Winters, 2010 [55]	OR	B								

B

APPENDIX B - OVERVIEW AND INFORMATION OF CHOSEN VARIABLES

This appendix gives a summary of the variables included in the data analysis and the modelling. Table B.1 gives a description of the variables and their unit. Table B.2 shows the data source, the type of variable in modelling, how it is measured, and the expectations of the results. A $\times$ means an expectation of not being significant or relevant and a $\checkmark$ means expecting it to be significant. When a variable is expected to be significant, an expectation can also be formed for the kind of correlation. A + is shown expecting it to be positively correlated and a - if it is expected to be negatively correlated.

Table B.1: Definition of variables

<i>Description</i>		<i>Unit</i>
Spatial variables		
Density of network	Density of the network at route level.	km/km^2
Connectivity of network	Number of links divided by number of nodes at route level.	
Separate bicycle lanes	Separate lanes relative to total bicycle lanes at route level.	%
Car parking zones	Percentage of location in a car parking zone. Measuring at origin and destination.	%
Access to bicycle parking	Location close to bicycle parking. Measuring at origin and destination.	-
Access to public transport	Percentage of location within 400m catchment area of BTM stops. Measuring at origin and destination.	%
Frequency of public transport	Average frequency per hour to BTM stops. Measuring at origin and destination.	Amount
Season	Meteorological season in which the trip is taken.	-
People's characteristics variables		
Gender	Male or female.	-
Age	Age of the individual.	-
Occupation	Occupation of the individual.	-
Education	Highest completed education of individual.	-
Driver's License	Individual in possession of a driver's license.	-
No. of household members	Number of household members of the individual.	Amount
Wealth	Standardized income + net worth of the household.	Amount
No. of cars per driving person	Number of cars available per household member with a driver's license.	Amount
E-bike availability	Availability of an e-bike in the household.	-
Address density	Address density of the zip code. Measuring at origin and destination.	$Amount/km^2$
Habit	Frequency in which a transport mode is used.	-
Transport journey variables		
Travel distance	Distance of a trip.	Kilometers (km)
Travel time	Time duration of a trip.	Minutes (min)
Calculated travel distance	Calculated travel distance of a trip for each mode.	Kilometers (km)
Calculated travel time	Calculated time duration of a trip for each mode.	Minutes (min)
Departure time	Time during the day at which a trip is taken.	-
Departure day	Day of the week in which a trip is taken.	-

Table B.2: Overview of characteristics of variables

	<i>Data source</i>	<i>Type</i>	<i>How to measure</i>	<i>Expectation</i>
Spatial variables				
Density of network	OSM, NDOV	Continuous	0...∞	All: +
Connectivity of network	OSM, NDOV	Continuous	0...∞	(E-)bike & Car: +
Separate bicycle lanes	OSM	Continuous	0...100	(E-)bike: + Car & PT: ×
Car parking zone	RDW	Continuous	0...100	Car: - (E-)bike & PT: +
Access to bicycle parking	CROW	Categorical	1: Security & paid, 2: Security & free	E-Bike: + Bike & car & PT: ×
Access to public transport	NDOV	Continuous	0...100	PT: + (E-)bike & car: ×
Frequency of public transport	NDOV	Continuous	0...∞	PT: + (E-)bike & car: ×
Season	ODiN: <i>Maand</i>	Categorical	1: Spring, 2: Summer, 3: Autumn, 4: Winter	(E-)bike: ✓ Car & PT: ×
People's characteristics variables				
Gender	ODiN: <i>Geslacht</i>	Binary	0: male, 1: female	All: ✓
Age	ODiN: <i>Leeftijd</i>	Categorical	1: 0-17, 2: 18-40, 3: 41-66, 4: 67-100	All: ✓
Occupation	ODiN: <i>MaatsPart</i>	Categorical	1: part-time job (12-30 hours), 2: full-time job (30+ hours), 3: own housekeeping, 4: student, 5: unemployed or unable to work, 6: retired	All: ✓
Education	ODiN: <i>Opleiding</i>	Categorical	1: primary education, 2: vmbo/mavo, 3: havo/vwo, 4: hbo/university	All: ✓
Driver's License	ODiN: <i>OPRijbewijsAu</i>	Binary	0: no license, 1: license	Car: + (E-)bike & PT: -
No. of household members	ODiN: <i>HHPers</i>	Continuous	0...9	All: ✓
Wealth	ODiN: <i>HHWelvG</i>	Categorical	1: first 20% group, 2: second 20% group, 3: third 20% group, 4: fourth 20% group, 5: fifth 20% group	All: ✓
No. of cars per driving person	ODiN: <i>HHAuto</i> <i>HHRijbewijsAu</i>	Continuous	0...2	Car: + (E-)bike & PT: -
E-bike availability	ODiN: <i>HHEFiets</i>	Binary	0: no e-bike, 1: e-bike	E-bike: + Bike & car & PT: -
Address density	CBS	Continuous	0...∞	(E-)bike & PT: + Car: -
Habit	ODiN: Starting with <i>Fq</i>	Categorical	1: never, 2: every year, 3: every month, 4: every week, 5: (almost) every day	All: ✓
Transport journey variables				
Travel distance	ODiN: <i>AfstV</i>	Continuous	0...∞	All: -
Travel time	ODiN: <i>Reisduur</i>	Continuous	0...∞	All: -
Calculated travel distance	OSM, NDOV	Continuous	0...∞	All: -
Calculated travel time	OSM, NDOV	Continuous	0...∞	All: -
Departure time	ODiN: <i>KVertTijd</i>	Categorical	1: 0:00-7:00, 2: 7:00-9:00, 3: 9:00-16:00, 4: 16:00-19:00, 5: 19:00-24:00	Car & PT: ✓ (E-)bike: x
Departure day	ODiN: <i>Weekdag</i>	Categorical	1: Sunday, 2: Monday, 3: Tuesday, 4: Wednesday, 5: Thursday, 6: Friday, 7: Saturday	All: ✓

C

APPENDIX C - ADDITIONS TO THE DATA ANALYSIS

This appendix first shows the representativeness of the ODIN sample compared to the population. Then the bar plots are shown of the factors that are analysed in Chapter 4 in which the ODIN factors are shown and the factors added to the ODIN dataset.

C.1. REPRESENTATIVENESS OF ODIN DATA

The representativeness of the ODIN data sample is shown in Table C.1. Generally, the ODIN dataset is close to the national distribution of these factors.

Table C.1: Representativeness of ODIN data sample (CBS, 2022)

Variable	Category	National share	ODiN share
Gender	<i>men</i>	49.64%	47.38%
	<i>women</i>	50.36%	52.62%
Age	<i>average</i>	41.9	46.18
	<i>0-20</i>	22.05%	1.68%
	<i>20-40</i>	24.85%	39.27%
	<i>40-65</i>	34.05%	43.47%
	<i>65+</i>	19.00%	15.58%
Occupation	<i>part-time</i>	11.16%	19.54%
	<i>full-time</i>	34.81%	43.98%
	<i>own housekeeping</i>	15.30%	3.63%
	<i>student</i>	12.99%	10.74%
	<i>unemployed</i>	1.38%	
	<i>unable to work</i>	4.35%	4.53%
	<i>retired</i>	20.01%	17.57%
Education	<i>primary education</i>	9.96%	2.53%
	<i>vmbo/mavo</i>	20.75%	14.14%
	<i>havo/vwo</i>	37.23%	35.35%
	<i>hbo/university</i>	30.57%	47.97%
	<i>not known</i>	1.49%	0.00%
Driver's License	<i>no</i>	35.45%	13.14%
	<i>yes</i>	64.55%	86.86%
Cars per driver's license	<i>average</i>	0.79	0.686
Household members	<i>average</i>	2.15	2.66
Wealth	<i>first 10% group</i>	10.00%	
	<i>second 10% group</i>	10.00%	10.95%
	<i>third 10% group</i>	10.00%	
	<i>fourth 10% group</i>	10.00%	15.94%
	<i>fifth 10% group</i>	10.00%	
	<i>sixth 10% group</i>	10.00%	19.88%
	<i>seventh 10% group</i>	10.00%	
	<i>eight 10% group</i>	10.00%	24.76%
	<i>ninth 10% group</i>	10.00%	
	<i>tenth 10% group</i>	10.00%	28.46%
E-bike ownership	<i>no</i>	86.90%	76.75%
	<i>yes</i>	13.10%	23.25%

Some differences are that the younger and older generation are less included in the ODIN dataset than the national share. Moreover, the share of retirees is therefore also lower than what can be seen nationally. The share of respondents doing their own housekeeping is very low compared to average, and the full-time workers are more represented than the national share. For the shares of education, havo/vwo should be the largest share, while in the ODIN dataset it is the hbo/university category. The ODIN dataset also includes a larger share of people with a driver's license. This could also be because the minimal age of respondents in ODIN is 6 years old. Furthermore, the richer households are more represented in the ODIN dataset than what is seen nationally, and the poorer households are thus less represented. Lastly, e-bike ownership has a slightly higher share compared to the national share. Although the ODIN dataset generally represents the population well, the differences need to be considered when interpreting the results of the models.

C.2. BAR PLOTS FROM ODIN DATA

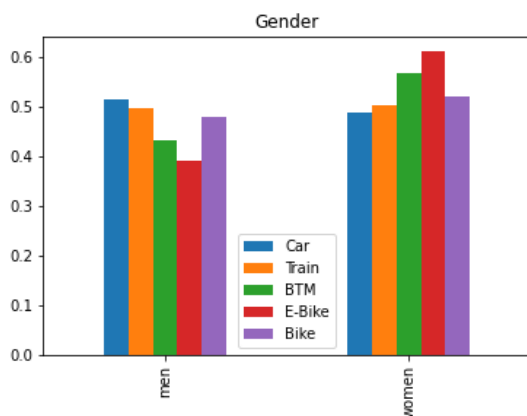


Figure C.1: Gender counts

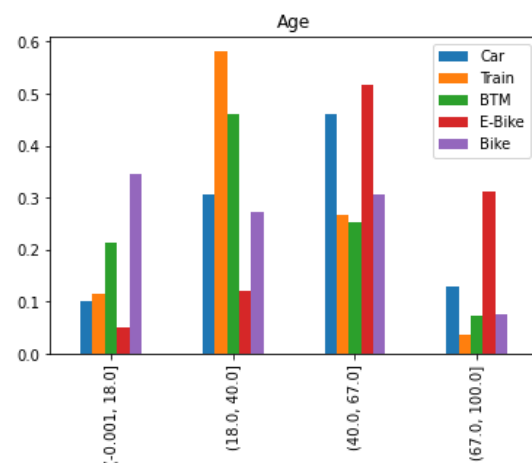


Figure C.2: Age counts

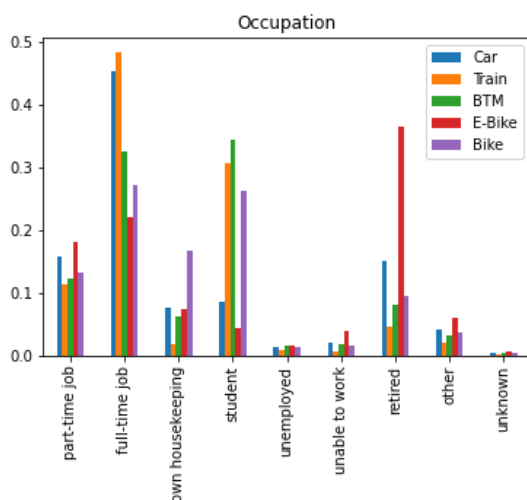


Figure C.3: Occupation counts

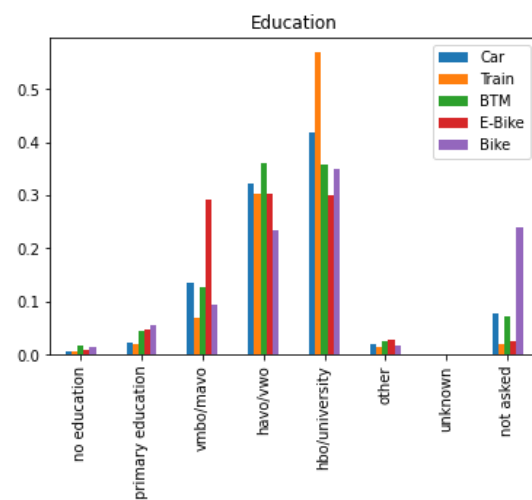


Figure C.4: Education counts

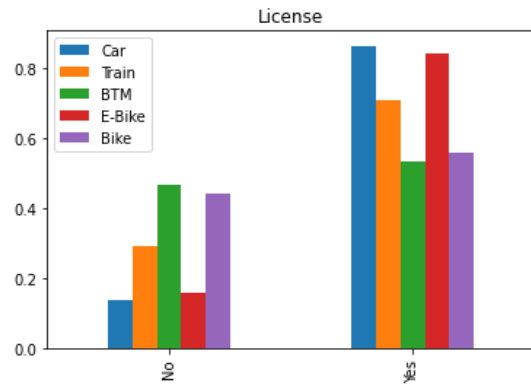


Figure C.5: Driver's license counts

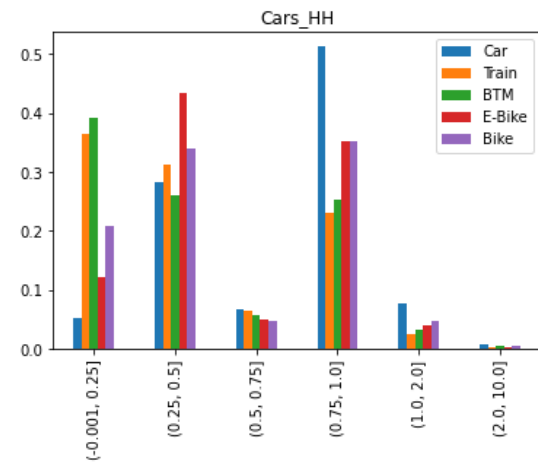


Figure C.6: No. of cars per driver's license in household counts

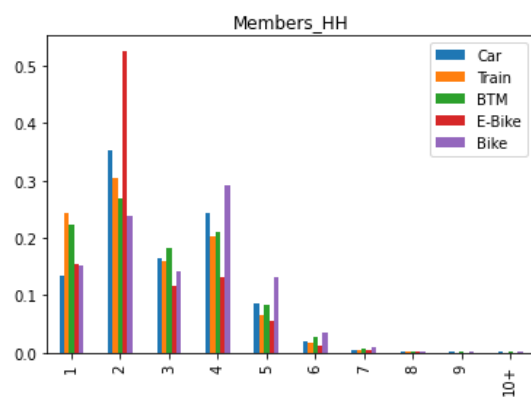


Figure C.7: No. of household members counts

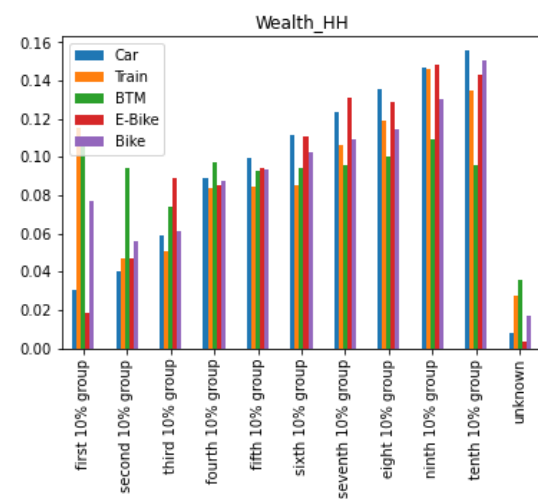


Figure C.8: Wealth counts

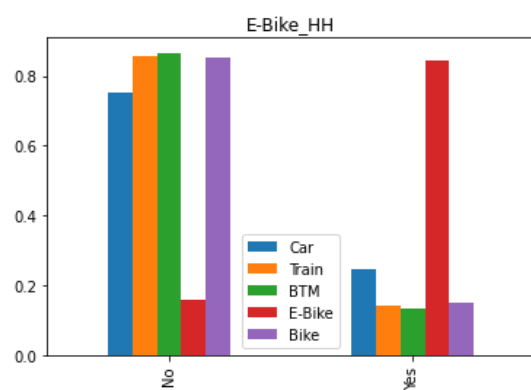


Figure C.9: Availability of E-bike counts

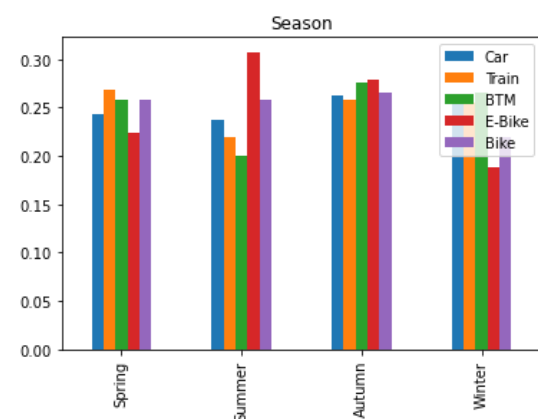


Figure C.10: Season counts

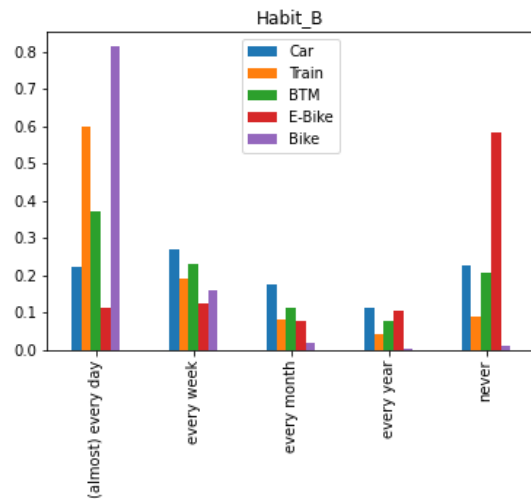


Figure C.11: Habit bike counts

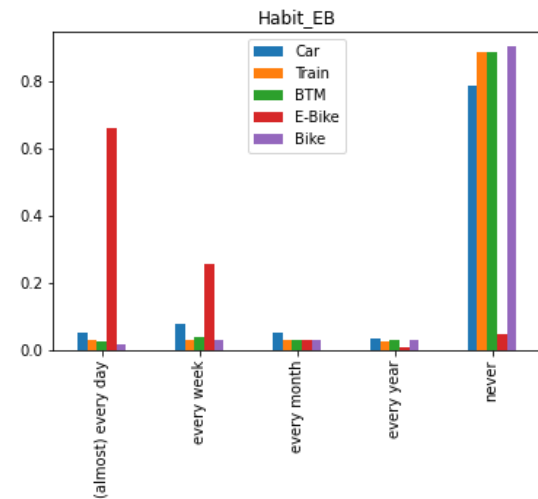


Figure C.12: Habit e-bike counts

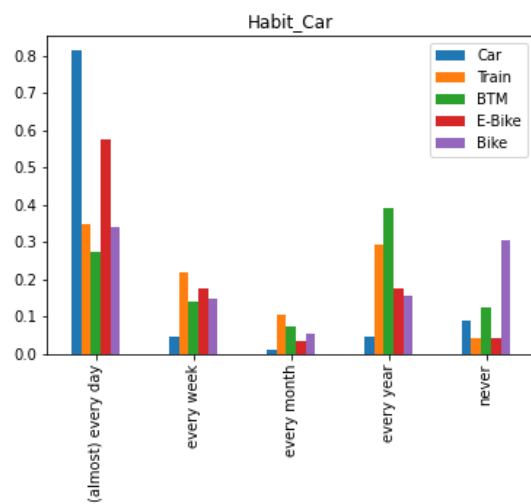


Figure C.13: Habit car counts

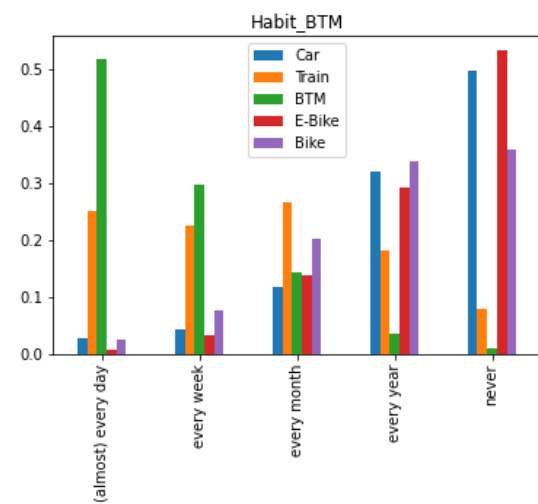


Figure C.14: Habit BTM counts

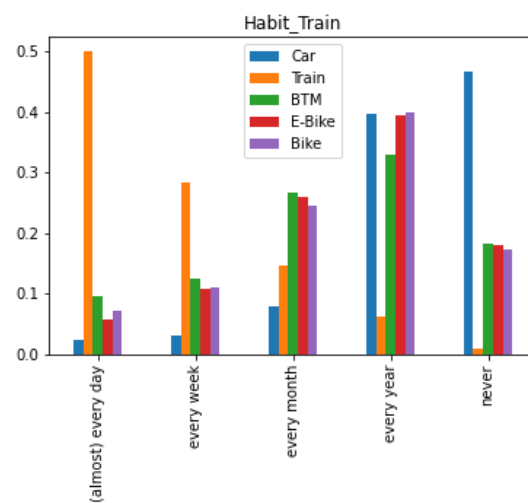


Figure C.15: Habit train counts

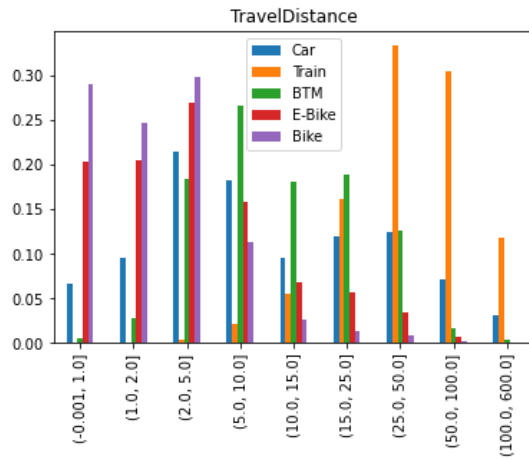


Figure C.16: Travel distance counts

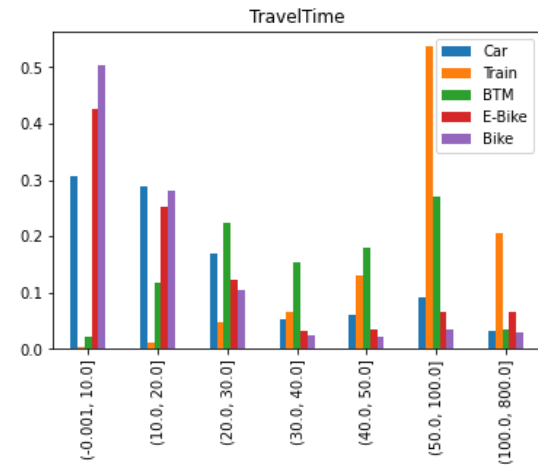


Figure C.17: Travel time counts

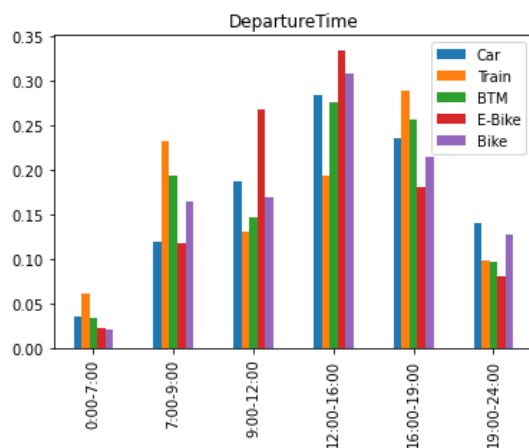


Figure C.18: Departure time counts

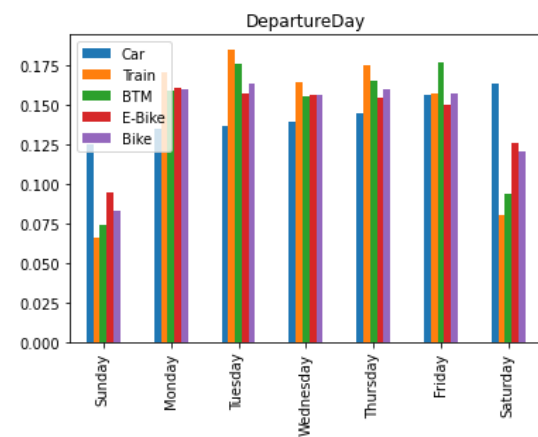


Figure C.19: Departure day counts

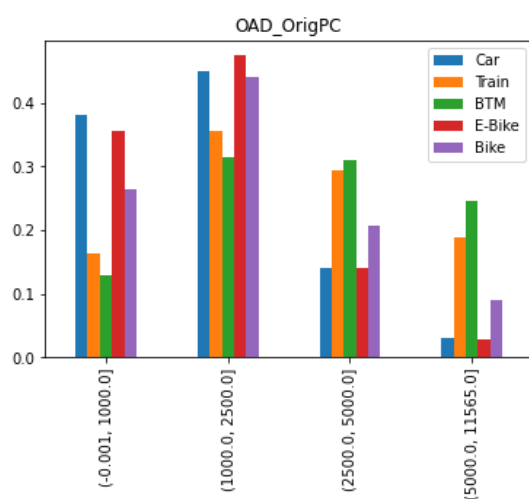


Figure C.20: OAD origin counts

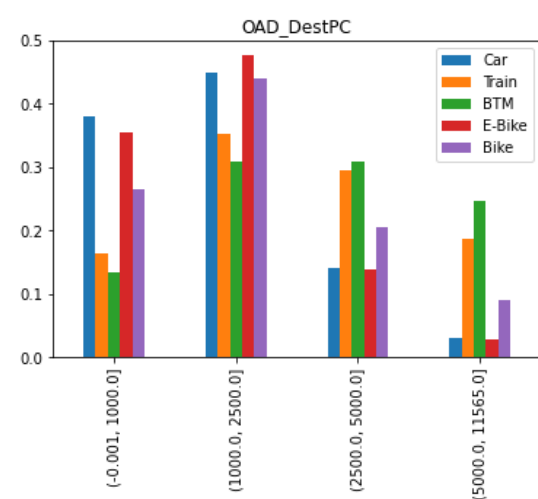


Figure C.21: OAD destination counts

C.3. BAR PLOTS FROM ADDED DATA

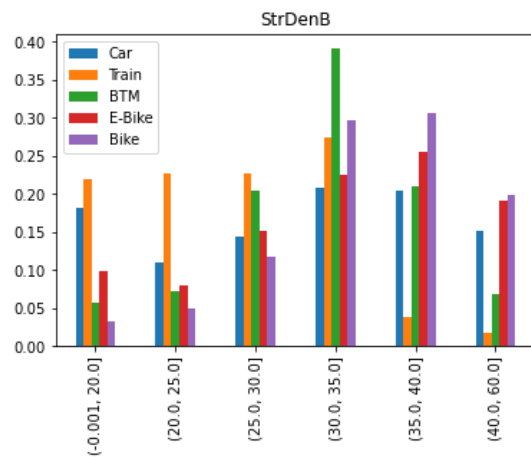


Figure C.22: Street Density Bike counts

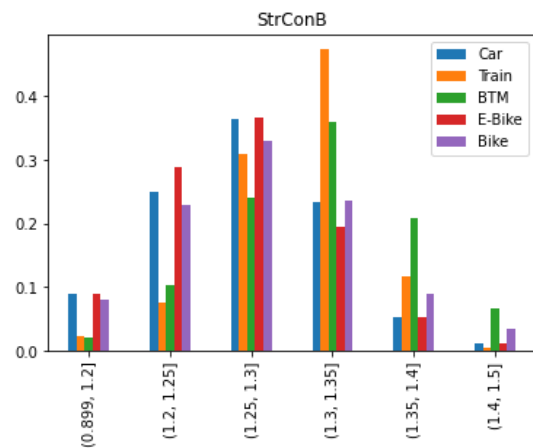


Figure C.23: Street Connectivity Bike counts

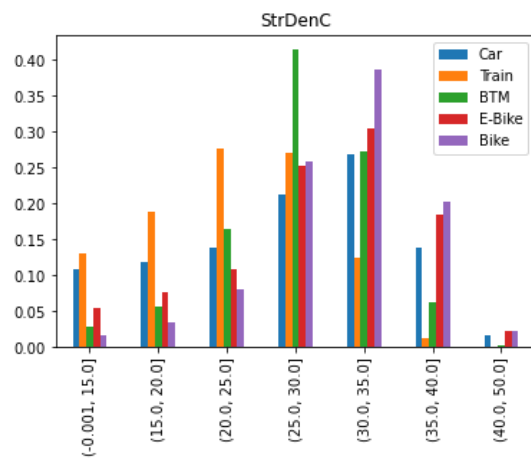


Figure C.24: Street Density Car counts

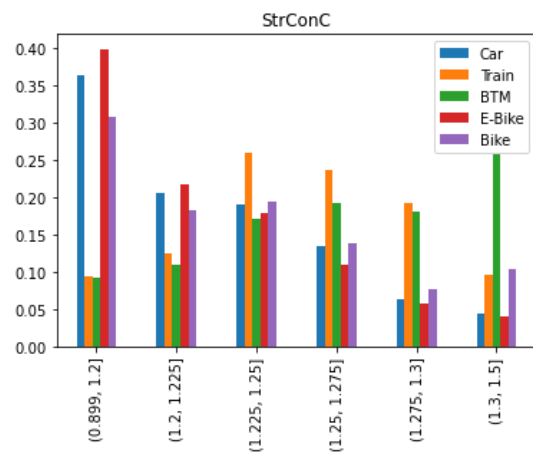


Figure C.25: Street Connectivity Car counts

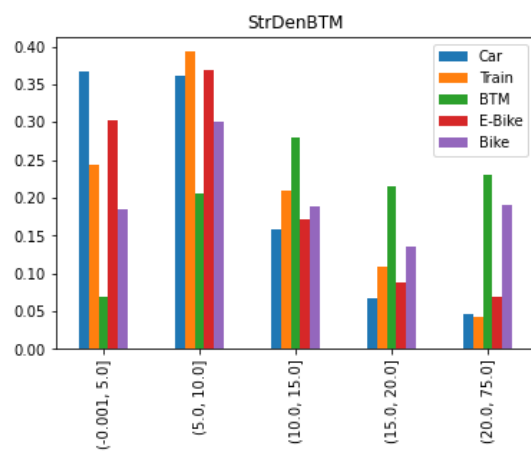


Figure C.26: Street Density BTM counts

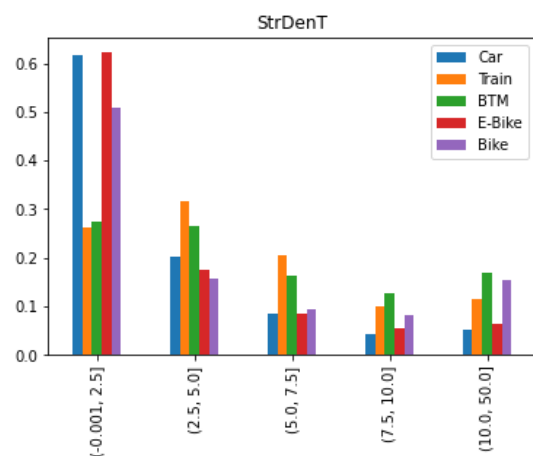


Figure C.27: Street Density Train counts

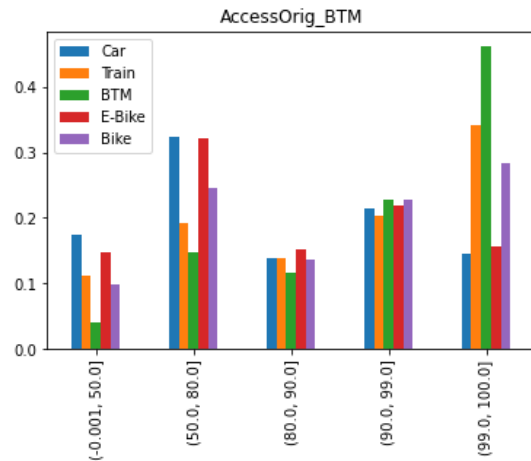


Figure C.28: Access BTM Origin counts

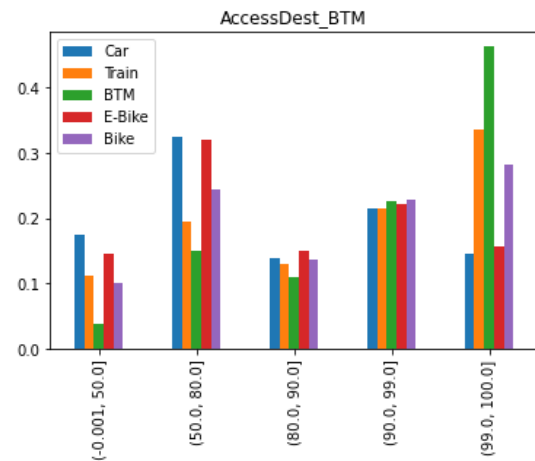


Figure C.29: Access BTM Destination counts

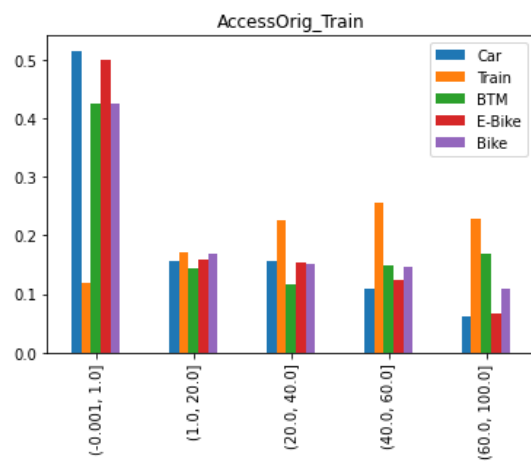


Figure C.30: Access Train Origin counts

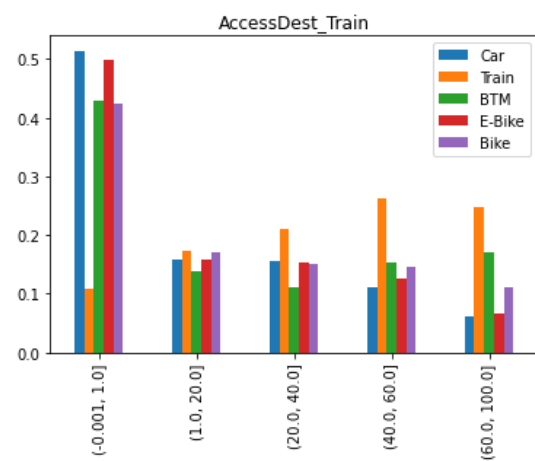


Figure C.31: Access Train Destination counts

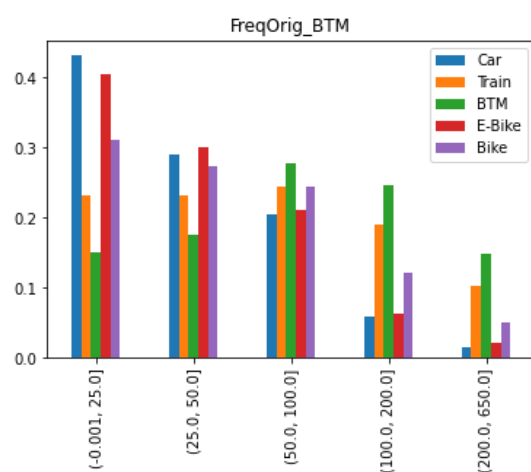


Figure C.32: Frequency BTM Origin counts

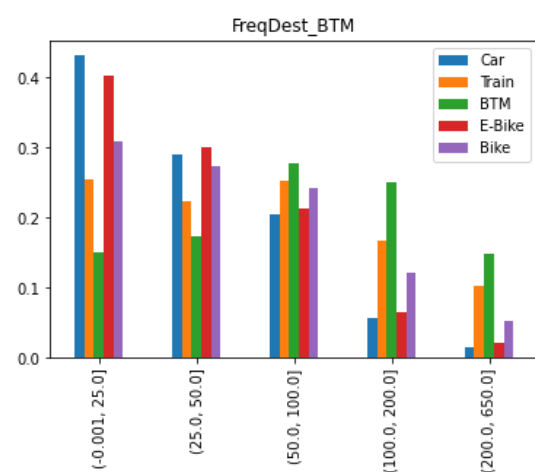


Figure C.33: Frequency BTM Destination counts

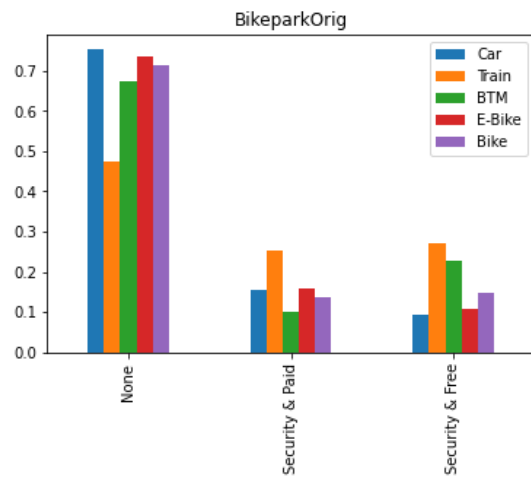


Figure C.34: Bicycle parking Origin counts

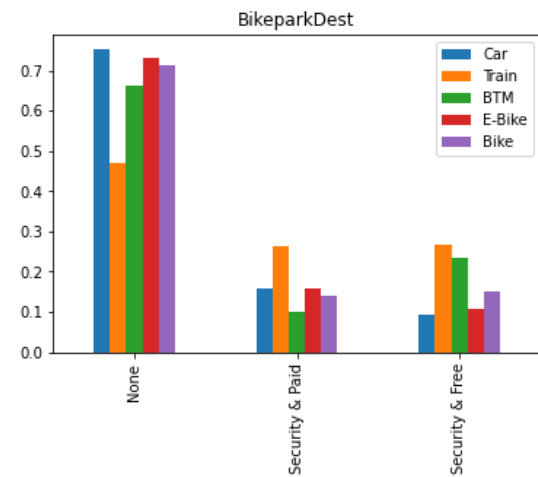


Figure C.35: Bicycle parking Destination counts

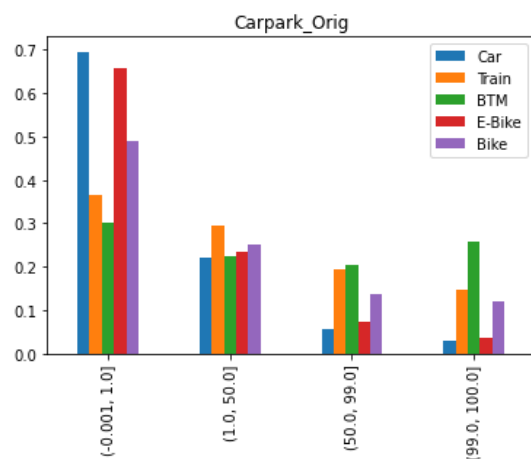


Figure C.36: Car parking zone Origin counts

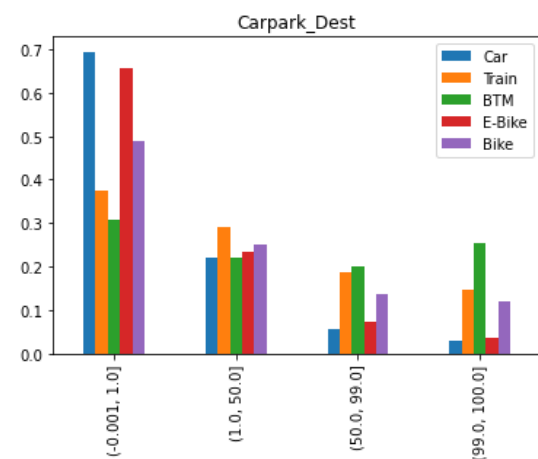


Figure C.37: Car parking zone Destination counts

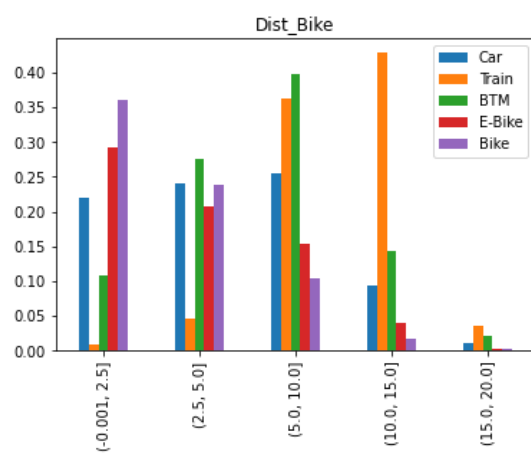


Figure C.38: Calculated travel distance bike counts

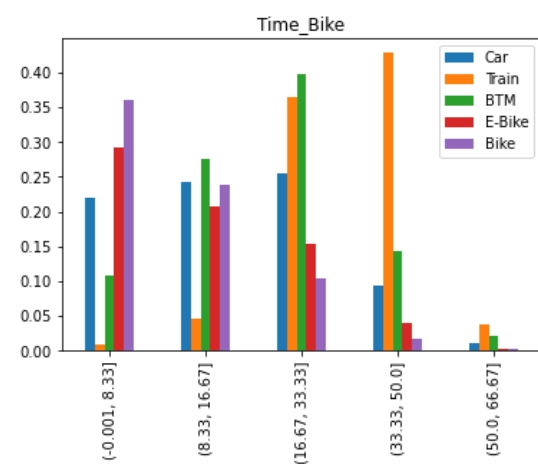


Figure C.39: Calculated travel time bike counts

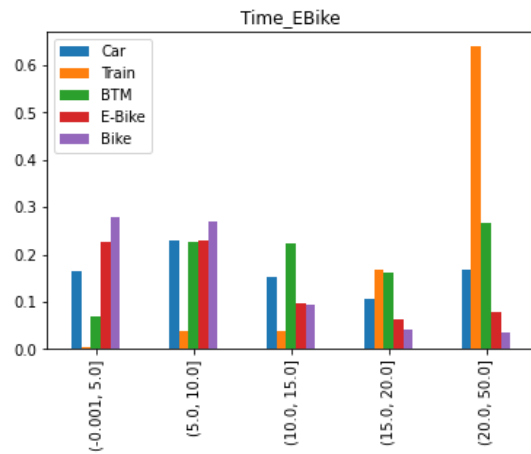


Figure C.40: Calculated travel time E-bike counts

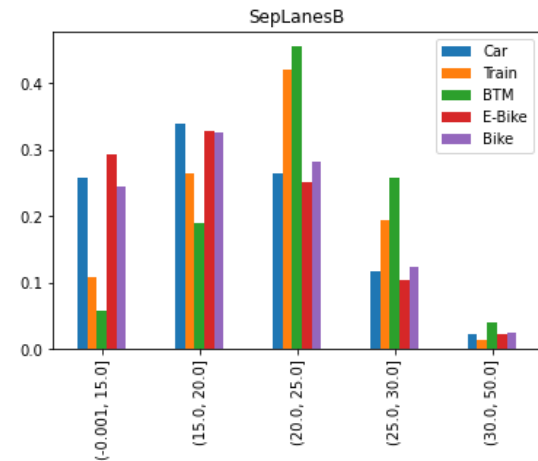


Figure C.41: Separate Bikelanes counts

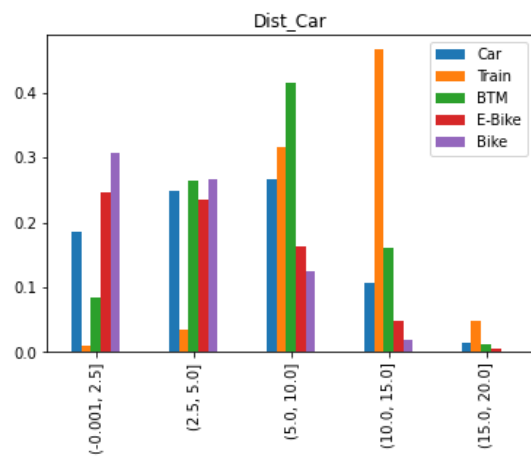


Figure C.42: Calculated travel distance car counts

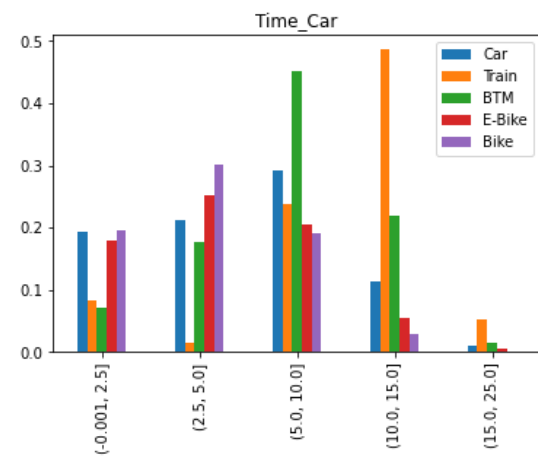


Figure C.43: Calculated travel time car counts

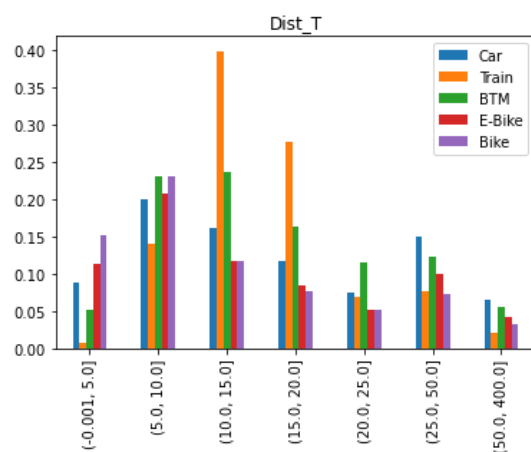


Figure C.44: Calculated travel distance train counts

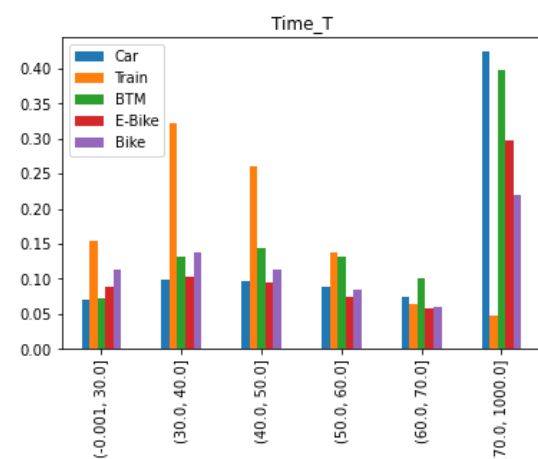


Figure C.45: Calculated travel time train counts

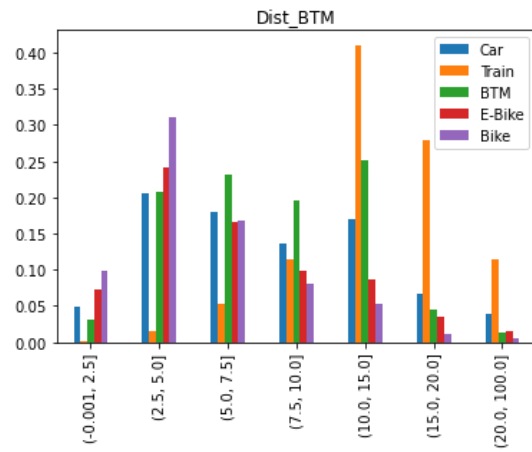


Figure C.46: Calculated travel distance BTM counts

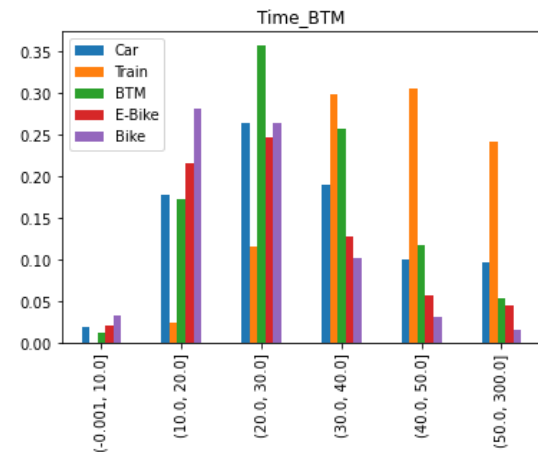


Figure C.47: Calculated travel time BTM counts

D

APPENDIX D - FILTERING STEPS

This appendix shows the filtering steps taken in the ODIN dataset that include the years 2018 and 2019. The first steps of the filtering found in Table D.1 are based on selecting the right data, and the numbering in the text refers to the filtering steps shown in the table. This is explained in Section 4.1.3. The next steps explained are based on removing data that is invalid or not needed.

Based on the plots from Appendix C, some categories from ODIN are removed from the dataset. These plots show for example that the availability of the e-bike and the availability of a driver's license do not contain an 'unknown' value in the dataset. Furthermore, the 'other', 'unknown', or 'not asked' values for multiple factors will be removed, because they do not contain specific information which can be used in modelling. This will be removed for the occupation (5), education (6), and wealth (7). Moreover, for occupation, the categories of unemployed and unable to work are combined as one category because they are quite similar in name and share between modes. Also, the categories of wealth are changed from 10% groups to 20% groups to lower the number of categories, and the least change in shares of modes is present when grouping this way. This is to limit the impact of the estimated influence of having fewer categories for this factor in modelling. Moreover, the 10+ category from the number of household members is low and does not contain the same type of information as the other categories, so this value is also removed so it can be modelled as a continuous variable (8). Furthermore, the 'unknown' values from the habit of the car (9) are removed, and the values of cars per driver's license in a household between 2.0-10.0 (10), because it does not contain much data for such a large range. The last removal is for the zero values of the zip codes and distances between the zip codes that are illogical, which are the calculated distances larger than 20 km (11). For these values, it is unsure if they are correct because the travel distance was selected as being smaller than 15 kilometers. Finally, for validation purposes after modelling, the entries that contain an origin or destination within the area of Haarlem are kept separate from the validation dataset (12). The final dataset is saved and then also split into a leisure dataset (13) and a practical dataset (14). The number of respondents in the dataset after each filtering step and the percentage change compared to the previous step are given in Table D.1.

Table D.1: Filtering steps

Filter step	Description	Number of respondents	Percentage change
(1)	Complete dataset	374329	
(2)	Selection of modes	295766	-21.0%
(3)	Selection of urbanity	236853	-19.9%
(4)	Selection of distance	160838	-32.0%
(5)	Remove from occupation	153724	-4.4%
(6)	Remove from education	127568	-17.0%
(7)	Remove from wealth	125976	-1.2%
(8)	Remove from household members	125937	-0.03%
(9)	Remove from habit car	121256	-3.7%
(10)	Remove from no. of cars	120758	-0.4%
(11)	Remove from zip codes	118566	-1.8%
(12)	Remove locations in Haarlem	116783	-1.5%
(13)	Split to leisure motives only	63852	-45.3% Compared to (12)
(14)	Split to practical motives only	29612	-74.6% Compared to (12)

E

APPENDIX E - ADDITIONS TO THE MODELLING

This appendix supports extra information about the results of the models. First, two MNL models are shown that differ in travel motive. These are estimated to analyse the difference between the general MNL as found in Section 5.2.4, and an MNL that is estimated with only trips of a certain travel motive. Furthermore, information and details about the interaction effects and quadratic components is shown, which relates to Section 5.2.3. The section about interaction effects shows the results of the parameter estimates of the interaction effects and the significance of a model with only main effects and a model with both main effects and interaction effects. The section about quadratic components shows the plots made of the results of the quadratic components.

E.1. MOTIVE MODELS

This section shows the comparison between travel motives for MNL models. The MNL model is re-estimated for trips of only practical motives and trips of only leisure motives. The newly estimated parameters are compared to the general MNL in Table 5.10 to see how the motives have possibly changed the influence of variables on mode choice. The goal of this section is to provide further insights into the influence and behaviour of the variables in the model. Municipalities can use this information to optimize their changes in bicycle networks and policies for a particular travel motive.

MNL - Practical motives

The practical motives include trips for the commute and educational purposes. Results of only using trips made for these motives can be found in Table E.1. It can be seen that there is much insignificant for the e-bike, mostly because of the fewer trips taken for these motives. The number of cars per driver's license main effect has become insignificant for the bike and train, while the quadratic components are still negative. For the e-bike, the quadratic component is insignificant while the main effect is negatively significant. There is thus still a negative relationship between the number of cars per driver's license and mode choice, but the kind of relationship is different than under general circumstances. A positive utility when fewer cars are available in a household is thus not an effect that has been found for practical motives for all transport modes. The relation between travel distance and the bike and e-bike has changed to only a significant main effect. The quadratic relation is thus not present for practical motives. The interaction of distance and age 67-100 has become more negative for the e-bike, which means there is a larger resistance for this age group when travelling a larger distance for practical motives. The travel speed has become insignificant for BTM. Speed is thus not important for practical motives when travelling with BTM modes, while the travel distance is about equally important. Car parking is less important and insignificant for the bike and e-bike, while it is more important for the public transport modes. For practical motives, public transport is thus more likely to be the alternative when paid car parking is present.

Furthermore, low-educated people are more likely to take public transport to work or school while higher-educated people more often take the train. The frequency and access of public transport are less important for practical motives. The reason could be that people tend to use the same connections each day to school or work, and thus an increased frequency or access is still relevant, but less important than usual. The gender is not significant for the bike and BTM, thus males or females do not have a preference for using these modes when travelling to school or work. Having a driver's license leads to a more negative influence on the mode choice compared to the car, thus the car is more often chosen when someone has a license. Being a student leads to a higher choice for the bike or public transport and the street density is only of importance for the (e-)bike, where it has a more positive influence on the choice of these modes. Finally, the wealth of a household is less important because many of the categories have become insignificant.

MNL - Leisure motives

The leisure motives include the trip purposes of shopping, visiting, touring, sport/hobby, and other leisure. The model result of only including data of leisure motives can be found in Table E.2. It shows that the alternative specific constant of the bike is not significant, thus there is no preference for the bike compared to the car. The e-bike has a more positive influence, and thus there is a larger preference for this mode for leisure trips. Other differences compared to the general MNL are that the number of cars per driver's license is now positively significant for the e-bike, with a more negative quadratic component. This leads to a positive utility when there are not many cars available in a household, and a negative utility when more cars are available. The positive estimation of the main effect is higher compared to the general MNL for the bike and BTM. This means that the contribution to utility stays positive for a larger number of cars per driver's license in a household. However, it is the other way around for the train. The positive main effect and negative quadratic effect are lower than in the general MNL, which leads to the contribution to the utility being positive for a smaller number of cars per driver's license in a household.

Furthermore, the travel speed is more negatively significant for BTM, which means that a higher speed is not preferred for leisure trips. It seems that people travelling for leisure do not care about even lower travel speeds than usual. Car parking has turned out to be more influential in leisure trips, by being more positively estimated. Car parking is thus most effective for people travelling for leisure purposes. The parameter estimates for the educational level show that it is generally less important for the BTM, where only education level havo/vwo is significant. The parameter estimates for the train show that havo/vwo is positively estimated for practical motives, and negative for leisure motives. This means that people with an education level of havo/vwo prefer the train to get to work or school, but in their spare time do not want to take the train. The access to public transport is higher estimated than for the general MNL, and the frequency is similar. The importance of a higher frequency at the destination is not significant in leisure trips for the train. This means that the accessibility of public transport is felt to be more important for the mode choice of public transport for leisure trips compared to general trips. The driver's license is less negatively estimated for leisure purposes, which means other modes compared to the car are more easily chosen when going on a leisure trip, while for practical motives the resistance is much higher. A change in significance can be seen for someone doing their own housekeeping. In the general MNL, it is not significant for the e-bike and BTM, while in the leisure motive MNL the bike and train are not significant. When going on a leisure trip, the e-bike and BTM are thus more often chosen. The street density is less important for the bike, not important for the e-bike, and more important for BTM mode choice. Finally, the train is more often chosen for households with a higher income compared to practical purposes.

Table E.1: MNL practical variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	2.090**	1.540**	ref: 1.0	-2.580**	-2.010**
Access to public transport (Origin)	-	-	-	1.140**	0.503**
Access to public transport (Destination)	-	-	-	1.200**	0.626**
Age: 0-17	ref	ref	ref	ref	ref
Age: 18-40	-0.819**	-0.419	ref	-0.524**	-0.784**
Age: 41-66	-0.789**	0.123	ref	-0.819**	-1.150**
Age: 67-100	-1.230**	0.206	ref	-0.750**	-2.270**
Cars per driver's license	-0.191	-1.350**	ref	-0.725**	0.337
Quadratic: cars per driver's license	-0.613**	0.158	ref	-0.243*	-1.230**
Travel distance	-0.253**	-0.160**	-0.094**	0.133**	-0.114**
Quadratic: travel distance	0.004	0.003	-0.001	-0.007**	0.000**
Interaction: distance & age 0-17	ref	ref	-	-	-
Interaction: distance & age 18-40	-0.058**	-	-	-	-
Interaction: distance & age 41-66	-0.023	-	-	-	-
Interaction: distance & age 67-100	-0.117**	-0.122**	-	-	-
Interaction: distance & gender	-0.055**	-0.098**	-	-	-
Travel speed	-	-	0.010*	-0.595	10.700**
Car parking (Origin)	0.377	0.034	ref	1.200**	1.790**
Quadratic: car parking (Origin)	0.314	0.385	ref	-0.471	-1.530**
Interaction: car parking (Origin) & residential zip code	0.149	0.407	ref	0.926**	0.562**
Car parking (Destination)	0.352	-0.064	ref	1.170**	2.350**
Quadratic: car parking (Destination)	0.301	0.437	ref	-0.482	-1.820**
Interaction: car parking (Destination) & residential zip code	0.182	0.388	ref	0.929**	0.401*
Education: primary education	ref	ref	ref	ref	ref
Education: vmbo/mavo	-0.221	0.029	ref	0.732**	0.438
Education: havo/vwo	0.020	-0.043	ref	0.554**	0.553*
Education: hbo/university	0.440**	0.168	ref	0.313	0.824**
Frequency of BTM stops (Origin)	0.136**	0.108	ref	0.406**	0.231**
Frequency of BTM stops (Destination)	0.201**	0.161*	ref	0.389**	0.247**
Gender	-0.034	1.120**	ref	0.021	-0.347**
Driver's license	-1.420**	-1.020**	ref	-2.020**	-2.050**
Household members	0.074**	-0.028	ref	0.111**	0.020
Occupation: part-time job	ref	ref	ref	ref	ref
Occupation: full-time job	-0.247**	-0.099	ref	-0.159*	-0.657**
Occupation: own housekeeping	-0.602**	-0.596*	ref	-0.333	-0.844*
Occupation: student	1.030**	-0.290	ref	1.450**	1.040**
Occupation: unemployed / unable to work	-0.641**	-0.140	ref	0.071	-1.400**
Street density	3.840**	1.920**	-0.195	-0.239	-
Wealth: first 20% group	ref	ref	ref	ref	ref
Wealth: second 20% group	-0.004	0.169	ref	0.013	0.102
Wealth: third 20% group	0.108	0.301	ref	0.388**	0.098
Wealth: fourth 20% group	0.242**	0.477*	ref	0.314**	0.560**
Wealth: fifth 20% group	0.424**	0.275	ref	0.288**	0.519**
Log likelihood: -24598.07	Rho-square-bar: 0.393				
ref= reference alternative / category	red ≠ significant		* = p < 0.05 ** = p < 0.01		

Table E.2: MNL leisure variables

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Alternative Specific Constants	-0.092	2.220**	ref: 1.0	-5.810**	-6.530**
Access to public transport (Origin)	-	-	-	1.310**	0.772**
Access to public transport (Destination)	-	-	-	1.350**	0.933**
Age: 0-17	ref	ref	ref	ref	ref
Age: 18-40	-0.603**	-1.210**	ref	-0.713**	0.281
Age: 41-66	-0.482**	-0.767*	ref	-0.716**	-0.082
Age: 67-100	-0.643**	-0.901**	ref	-0.232	-1.010*
Cars per driver's license	1.120**	0.388*	ref	0.722**	1.200**
Quadratic: cars per driver's license	-0.833**	-0.425**	ref	-0.552**	-1.010**
Travel distance	-0.377**	-0.379**	-0.126**	0.198**	-0.077**
Quadratic: travel distance	0.013**	0.008**	0.002	-0.009**	0.000
Interaction: distance & age 0-17	ref	ref	-	-	-
Interaction: distance & age 18-40	-0.069*	-	-	-	-
Interaction: distance & age 41-66	-0.087**	-	-	-	-
Interaction: distance & age 67-100	-0.142**	0.025	-	-	-
Interaction: distance & gender	-0.060**	-0.068**	-	-	-
Travel speed	-	-	0.010*	-0.890**	10.500**
Car parking (Origin)	0.934**	1.350**	ref	1.120**	1.520*
Quadratic: car parking (Origin)	-0.190	-0.684	ref	-0.345	-1.450*
Interaction: car parking (Origin) & residential zip code	0.355**	0.082	ref	1.010**	0.909**
Car parking (Destination)	0.995**	1.440**	ref	0.491	1.580*
Quadratic: car parking (Destination)	-0.370*	-0.863*	ref	0.270	-2.120**
Interaction: car parking (Destination) & residential zip code	0.296**	0.032	ref	1.040**	1.480**
Education: primary education	ref	ref	ref	ref	ref
Education: vmbo/mavo	0.182*	-0.317**	ref	0.187	-1.010**
Education: havo/vwo	0.259**	-0.489**	ref	0.250*	-0.633*
Education: hbo/university	0.641**	-0.426**	ref	0.093	-0.504
Frequency of BTM stops (Origin)	0.066**	0.016	ref	0.416**	0.397**
Frequency of BTM stops (Destination)	0.092**	-0.088	ref	0.395**	0.170
Gender	0.060*	0.448**	ref	0.166**	-0.218*
Driver's license	-0.617**	-0.473**	ref	-1.270**	-0.970**
Household members	0.030**	-0.130**	ref	0.117**	-0.072
Occupation: part-time job	ref	ref	ref	ref	ref
Occupation: full-time job	-0.341**	-0.482**	ref	-0.199**	-0.153
Occupation: own housekeeping	-0.077	0.281**	ref	0.280**	0.055
Occupation: student	0.690**	-1.140**	ref	0.565**	0.953**
Occupation: unemployed / unable to work	-0.344**	-0.058	ref	0.099	-0.526*
Street density	1.840**	0.615	-0.448	0.923**	-
Wealth: first 20% group	ref	ref	ref	ref	ref
Wealth: second 20% group	0.159**	0.101	ref	0.068	0.349
Wealth: third 20% group	0.361**	0.210*	ref	0.431**	0.805**
Wealth: fourth 20% group	0.438**	0.356**	ref	0.237**	1.020**
Wealth: fifth 20% group	0.621**	0.352**	ref	0.083	0.899**
Log likelihood: -44613.39	Rho-square-bar: 0.501				
ref= reference alternative / category	red ≠ significant		* = $p < 0.05$ ** = $p < 0.01$		

E.2. INTERACTION EFFECTS

The results of the interaction effects are given in this section. The interaction effects tested are distance & gender, distance & age, distance & speed, car parking zones & wealth, and car parking zones & residential zip code.

The interaction between travel distance and gender leads to a significantly better model fit. The results in Table E.3 show that for the bike and e-bike, males are more likely to travel longer distances.

The interaction between travel distance and age leads to a significantly better model fit. The results in Table E.4 show that for the bike and e-bike, the elderly have a larger resistance to cycling longer distances. For the bike, there is also a smaller significant resistance against cycling longer distances for the other age groups compared to the youngest age group.

The interaction between travel distance and travel speed leads to a significantly better model fit. The results in Table E.5 show that a higher speed leads to a higher choice probability. This effect is more clearly seen for public transport than for the bike and the car. It is not significant for the e-bike.

The interaction between car parking zones and wealth leads to a significantly better model fit. However, the results in Table E.6 show that none of the interaction effects for the car are significant, and thus the interaction effect is not included further.

The interaction between car parking zones and residential zip code leads to a significantly better model fit. The results in Table E.7 show that for all modes the interaction effect is significant. The interaction effect is negative for the bike and e-bike, while it is positive for the public transport modes. The negative value with a positive main effect means that if the residential zip code is at origin or destination, there is a little less preference for the bike and e-bike compared to the car. The positive value means there is more preference for the public transport modes compared to the car. Apparently, the public transport modes are thus more likely chosen when travelling to or from home with paid parking zones. Often these zones are in the city and thus easily accessible by public transport.

Table E.3: Interaction effects Distance & Gender

Main effects		Main effects + Interaction	
Log likelihood	-101623.7	Log likelihood	-101484.0
Rho-square-bar	0.379	Rho-square-bar	0.380
ASC_Bike	0.892**	ASC_Bike	0.775**
ASC_EBike	0.647**	ASC_EBike	0.494**
ASC_BTM	-1.780**	ASC_BTM	-1.810**
ASC_Train	-3.650**	ASC_Train	-3.710**
B_CDIST_B	-0.429**	B_CDIST_B	-0.396**
B_GEN_B	-0.006	B_GEN_B	0.223**
		B_CDISTGEN_B	-0.071**
B_CDIST_EB	-0.368**	B_CDIST_EB	-0.329**
B_GEN_EB	0.513**	B_GEN_EB	0.787**
		B_CDISTGEN_EB	-0.072
B_CDIST_C	-0.200**	B_CDIST_C	-0.205**
B_GEN_C	ref	B_GEN_C	ref
		B_CDISTGEN_C	0.008
B_CDIST_BTM	-0.092**	B_CDIST_BTM	-0.089**
B_GEN_BTM	0.281**	B_GEN_BTM	0.333**
		B_CDISTGEN_BTM	-0.006
B_CDIST_T	-0.023**	B_CDIST_T	-0.020**
B_GEN_T	-0.085	B_GEN_T	0.043
		B_CDISTGEN_T	-0.008

ref = reference alternative / category

* = $p < 0.05$ ** = $p < 0.01$

Table E.4: Interaction effects Distance & Age

Main effects		Main effects + Interaction	
Log likelihood	-99868.4	Log likelihood	-99770.2
Rho-square-bar	0.389	Rho-square-bar	0.390
ASC_Bike	3.190**	ASC_Bike	2.970**
ASC_EBike	1.470**	ASC_EBike	1.220**
ASC_BTM	0.838**	ASC_BTM	0.848**
ASC_Train	-1.090**	ASC_Train	-1.060**
B_CDIST_B	-0.429**	B_CDIST_B	-0.378**
B_AGE_B_1	<i>ref</i>	B_AGE_B_1	<i>ref</i>
		B_CDISTAGE_B_1	<i>ref</i>
B_AGE_B_2	-2.150**	B_AGE_B_2	-1.970**
		B_CDISTAGE_B_2	-0.045**
B_AGE_B_3	-2.310**	B_AGE_B_3	-2.130**
		B_CDISTAGE_B_3	-0.043*
B_AGE_B_4	-2.720**	B_AGE_B_4	-2.280**
		B_CDISTAGE_B_4	-0.160**
B_CDIST_EB	-0.362**	B_CDIST_EB	-0.314**
B_AGE_EB_1	<i>ref</i>	B_AGE_EB_1	<i>ref</i>
		B_CDISTAGE_EB_1	<i>ref</i>
B_AGE_EB_2	-0.934**	B_AGE_EB_2	-0.778**
		B_CDISTAGE_EB_2	-0.023
B_AGE_EB_3	-0.483**	B_AGE_EB_3	-0.272
		B_CDISTAGE_EB_3	-0.039
B_AGE_EB_4	-0.532**	B_AGE_EB_4	-0.109
		B_CDISTAGE_EB_4	-0.110*
B_CDIST_C	-0.197**	B_CDIST_C	-0.199**
B_AGE_C_1	<i>ref</i>	B_AGE_C_1	<i>ref</i>
B_CDIST_BTM	-0.091**	B_CDIST_BTM	-0.092**
B_AGE_BTM_1	<i>ref</i>	B_AGE_BTM_1	<i>ref</i>
B_AGE_BTM_2	-2.120**	B_AGE_BTM_2	-2.130**
B_AGE_BTM_3	-2.890**	B_AGE_BTM_3	-2.900**
B_AGE_BTM_4	-2.710**	B_AGE_BTM_4	-2.760**
B_CDIST_T	-0.022**	B_CDIST_T	-0.022**
B_AGE_T_1	<i>ref</i>	B_AGE_T_1	<i>ref</i>
B_AGE_T_2	-2.140**	B_AGE_T_2	-2.170**
B_AGE_T_3	-3.040**	B_AGE_T_3	-3.080**
B_AGE_T_4	-4.350**	B_AGE_T_4	-4.450**

ref = reference alternative / category * = $p < 0.05$ ** = $p < 0.01$

Table E.5: Interaction effects Distance & Speed

Main effects		Main effects + Interaction	
Log likelihood	-101623.7	Log likelihood	-99727.0
Rho-square-bar	0.379	Rho-square-bar	0.390
ASC_Bike	0.888**	ASC_Bike	0.753**
ASC_EBike	0.939**	ASC_EBike	0.941**
ASC_BTM	-1.620**	ASC_BTM	-2.040**
ASC_Train	-3.690**	ASC_Train	-5.240**
B_CDIST_B	-0.429**	B_CDIST_B	-0.397**
		B_CSPEED_B	1.100**
B_CDIST_EB	-0.370**	B_CDIST_EB	-0.310**
		B_CSPEED_EB	0.060
B_CDIST_C	-0.201**	B_CDIST_C	-0.156**
		B_CSPEED_C	0.051**
B_CDIST_BTM	-0.093**	B_CDIST_BTM	-0.109**
		B_CSPEED_BTM	3.570**
B_CDIST_T	-0.023**	B_CDIST_T	-0.110**
		B_CSPEED_T	11.600**

ref = reference alternative / category * = $p < 0.05$ ** = $p < 0.01$

Table E.6: Interaction effects Car parking & Wealth

Main effects		Main effects + Interaction	
Log likelihood	-104727.4	Log likelihood	-104717.3
Rho-square-bar	0.360	Rho-square-bar	0.360
ASC_Bike	-0.324**	ASC_Bike	-0.317**
ASC_EBike	0.094	ASC_EBike	0.098
ASC_BTM	-2.060**	ASC_BTM	-2.060**
ASC_Train	-3.400**	ASC_Train	-3.390**
B_WEA_B_1	ref	B_WEA_B_1	ref
B_WEA_B_2	-0.019	B_WEA_B_2	0.020
B_WEA_B_3	0.169**	B_WEA_B_3	0.176**
B_WEA_B_4	-0.012	B_WEA_B_4	-0.035
B_WEA_B_5	-0.142**	B_WEA_B_5	-0.169**
B_WEA_EB_1	ref	B_WEA_EB_1	ref
B_WEA_EB_2	0.179*	B_WEA_EB_2	0.200**
B_WEA_EB_3	0.126	B_WEA_EB_3	0.130
B_WEA_EB_4	0.189**	B_WEA_EB_4	0.176*
B_WEA_EB_5	0.038	B_WEA_EB_5	0.023
B_WEA_C	ref	B_WEA_C	ref
B_WEA_BTM_1		B_WEA_BTM_1	ref
B_WEA_BTM_2	-0.019	B_WEA_BTM_2	0.020
B_WEA_BTM_3	0.169**	B_WEA_BTM_3	0.176**
B_WEA_BTM_4	-0.012	B_WEA_BTM_4	-0.035
B_WEA_BTM_5	-0.142**	B_WEA_BTM_5	-0.169**
B_WEA_T_1	ref	B_WEA_T_1	ref
B_WEA_T_2	-0.068	B_WEA_T_2	-0.034
B_WEA_T_3	-0.100	B_WEA_T_3	-0.094
B_WEA_T_4	0.285**	B_WEA_T_4	0.265**
B_WEA_T_5	0.161*	B_WEA_T_5	0.138
<i>Origin</i>			
B_CPARKORIG_B	0.009**	B_CPARKORIG_B	0.009**
B_CPARKORIG_EB	0.007**	B_CPARKORIG_EB	0.006**
B_CPARKORIG_C	ref	B_CPARKORIG_C	ref
B_CPARKORIG_BTM	0.017**	B_CPARKORIG_BTM	0.016**
B_CPARKORIG_T	0.013**	B_CPARKORIG_T	0.012**
		B_CPARKORIGWEA_C_1	ref
		B_CPARKORIGWEA_C_2	0.001
		B_CPARKORIGWEA_C_3	0.000
		B_CPARKORIGWEA_C_4	-0.001
		B_CPARKORIGWEA_C_5	-0.001
<i>Destination</i>			
B_CPARKDEST_B	0.009**	B_CPARKDEST_B	0.009**
B_CPARKDEST_EB	0.006**	B_CPARKDEST_EB	0.006**
B_CPARKDEST_C	ref	B_CPARKDEST_C	ref
B_CPARKDEST_BTM	0.016**	B_CPARKDEST_BTM	0.016**
B_CPARKDEST_T	0.012**	B_CPARKDEST_T	0.012**
		B_CPARKDESTWEA_C_1	ref
		B_CPARKDESTWEA_C_2	0.001
		B_CPARKDESTWEA_C_3	0.000
		B_CPARKDESTWEA_C_4	0.000
		B_CPARKDESTWEA_C_5	0.000

ref = reference alternative / category

* = $p < 0.05$ ** = $p < 0.01$

Table E.7: Interaction effects Car parking & Residential zip code

Main effects		Main effects + Interaction	
Log likelihood	-105126.0	Log likelihood	-104395.2
Rho-square-bar	0.357	Rho-square-bar	0.362
ASC_Bike	-0.051**	ASC_Bike	-0.047**
ASC_EBike	0.214**	ASC_EBike	0.220**
ASC_BTM	-2.030**	ASC_BTM	-2.070**
ASC_Train	-3.290**	ASC_Train	-3.310**
<i>Origin</i>			
B_CPARKORIG_B	0.890**	B_CPARKORIG_B	0.968**
		B_DUMCPARKORIG_B	-0.166**
B_CPARKORIG_EB	0.656**	B_CPARKORIG_EB	0.946**
		B_DUMCPARKORIG_EB	-0.466**
B_CPARKORIG_C	<i>ref</i>	B_CPARKORIG_C	<i>ref</i>
		B_DUMCPARKORIG_C	<i>ref</i>
B_CPARKORIG_BTM	1.670**	B_CPARKORIG_BTM	1.100**
		B_DUMCPARKORIG_BTM	0.991**
B_CPARKORIG_T	1.260**	B_CPARKORIG_T	0.791**
		B_DUMCPARKORIG_T	0.832**
<i>Destination</i>			
B_CPARKDEST_B	0.874**	B_CPARKDEST_B	0.927**
		B_DUMCPARKDEST_B	-0.128**
B_CPARKDEST_EB	0.620**	B_CPARKDEST_EB	0.834**
		B_DUMCPARKDEST_EB	-0.368**
B_CPARKDEST_C	<i>ref</i>	B_CPARKDEST_C	<i>ref</i>
		B_DUMCPARKDEST_C	<i>ref</i>
B_CPARKDEST_BTM	1.620**	B_CPARKDEST_BTM	0.964**
		B_DUMCPARKDEST_BTM	1.080**
B_CPARKDEST_T	1.200**	B_CPARKDEST_T	0.571**
		B_DUMCPARKDEST_T	1.010**
<i>ref</i> = reference alternative / category		* = $p < 0.05$ ** = $p < 0.01$	

E.3. QUADRATIC COMPONENTS

The results of the quadratic components are given in this section. The quadratic components that are tested are travel distance, car parking zones, and the number of cars per driver's license in a household.

The plot in Figure E.1 show the relation of the travel distance for each mode. The positive quadratic components for the car, bike, and e-bike lead to a less steep line toward higher travel distances (x-axis). However, this value is very small and thus the utility (y-axis) for travel distance still gets lower for each increased step in travel distance. For the train, the quadratic component is not significant and thus it remains linear. The BTM mode has a negative quadratic component and a positive linear value. Therefore, between distances of 0 and 8.67 km the utility is positive. For travel distances larger than 8.67 km, the utility for choosing BTM will also be negative, in which the line gets steeper for a larger distance. Overall, the most resistance against travel distance can be seen as expected for the bike.

The plot in Figure E.2 show the relation of the car parking zones for each mode but the car because that is the reference alternative. The darker colored line of a mode is of the origin variable, and the lighter colored line is of the destination variable. The negative quadratic components and positive main effects lead to a curve that gets less steep for a larger value of the car parking zones. The highest points are not reached for the BTM and the bike, but it is for the train and e-bike. This means that when an area is covered with more than about 72% for the train and about 82% for the e-bike, the utility decreases and thus will be less often chosen. Generally, public transport is preferred probably because paid parking zones are more often present in the cities where public transport can also be more often found.

The plot in Figure E.3 show the relation of the number of cars per driver's license in the household for each mode but the car, because that is the reference alternative. It shows that when not many cars are available, the train and the bike are favorite options. However, the quadratic component is highly negative, which leads to a steep decline in preference when the number of cars increases. When more than about 1.75 cars per driver's license are available, the e-bike becomes the favorite option. The quadratic component for BTM is not significant and it has therefore a linear decrease in utility.

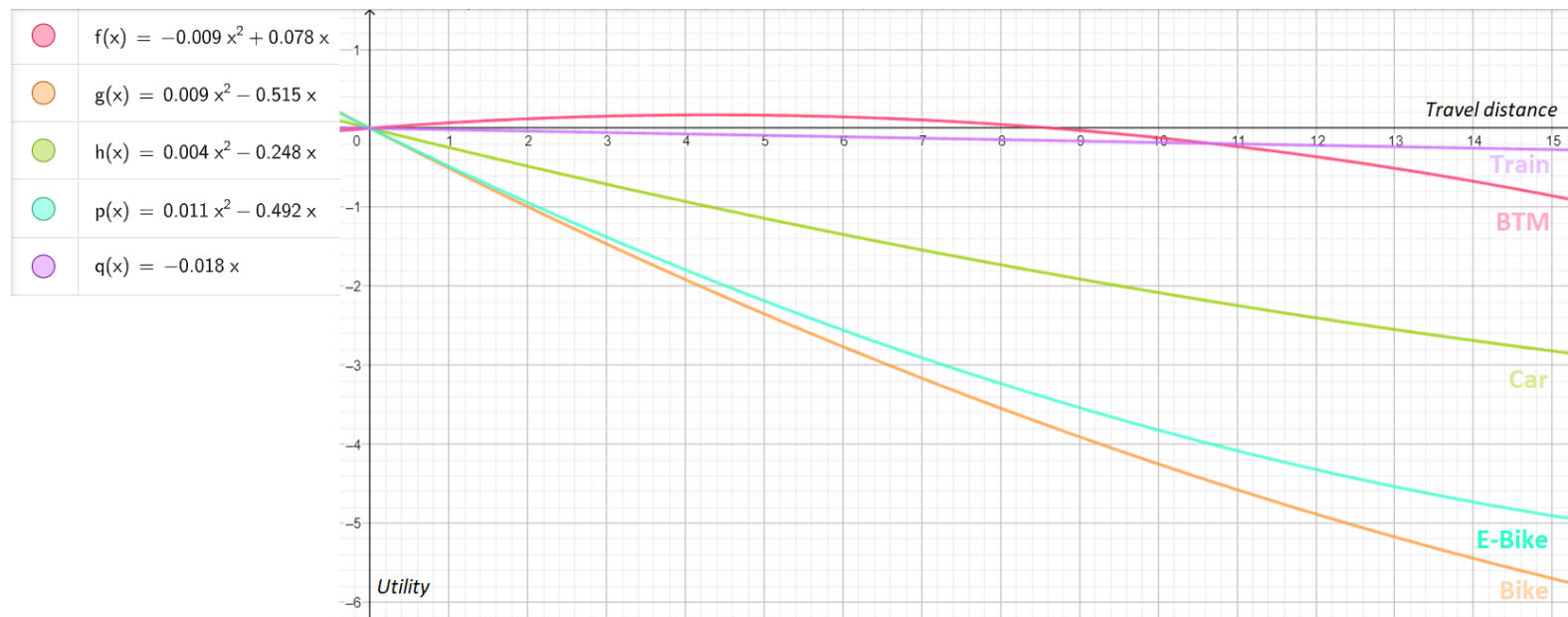


Figure E.1: Quadratic relation of the travel distance

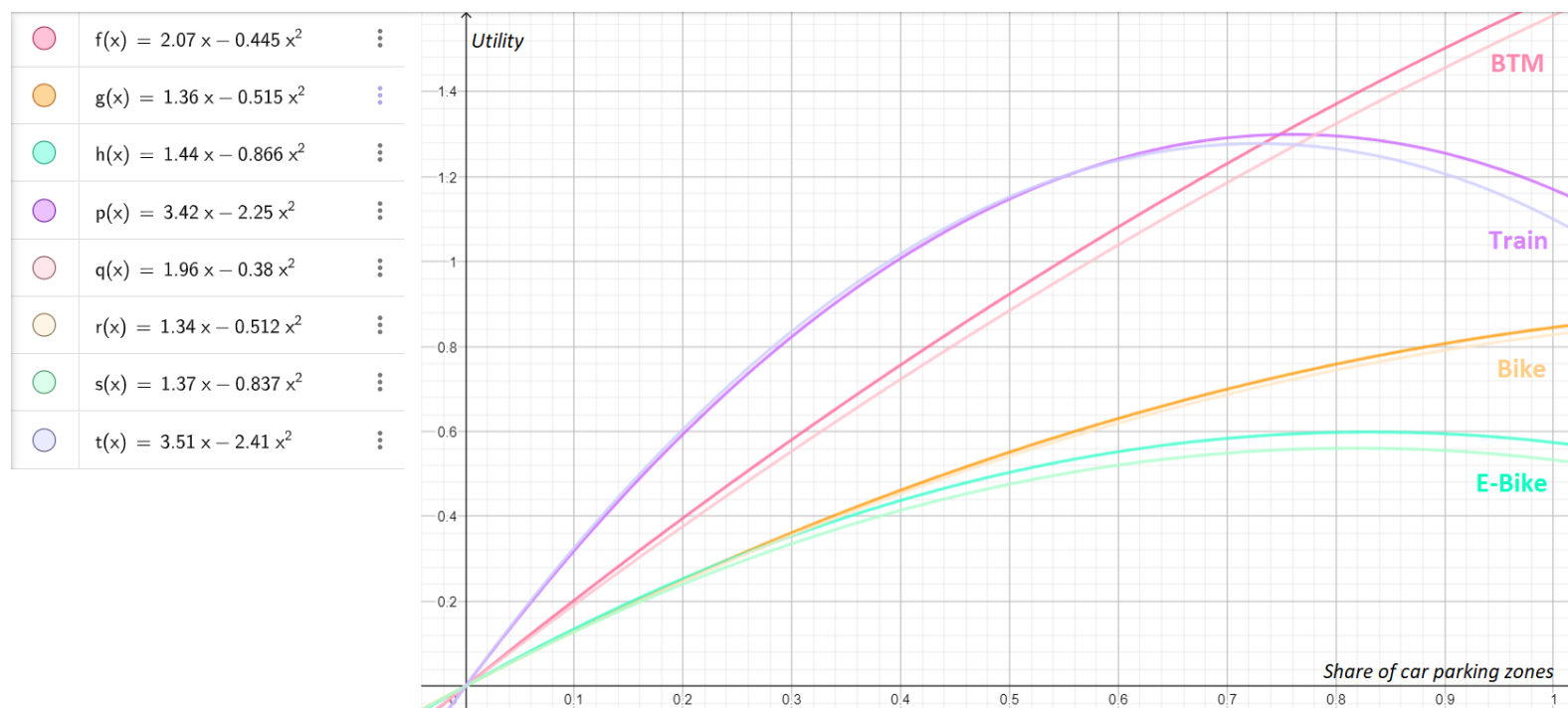


Figure E.2: Quadratic relation of the car parking zones

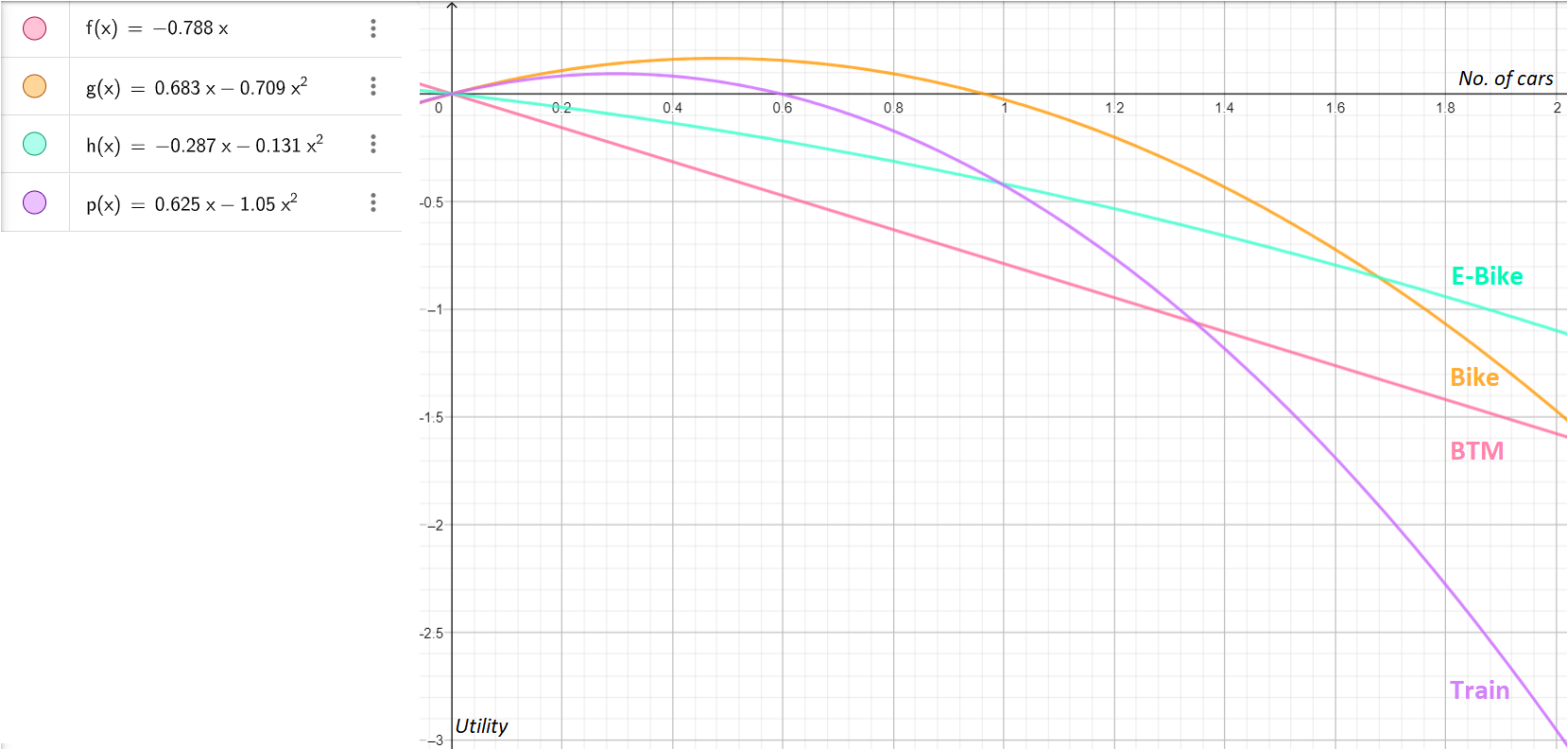


Figure E.3: Quadratic relation of the number of cars per driver's license in the household

F

APPENDIX F - BASE INPUT VALUES

This appendix shows the base input values of the variables for the calculation of the probability of the modes. These are used in the sensitivity and uncertainty analysis in Section 5.3, and in the application of Chapter 7.

Most input values in Table F1 are based on the averages and standard deviations of the population of the Netherlands (CBS, 2022). The spatial variables are based on the average and standard deviation calculated for each zip code in a Dutch urban area. Only the travel distance, travel speed, residential zip code, and street density are based on the average and standard deviation from the ODiN dataset.

Table F1: Input values for utility functions

	<i>Bike</i>	<i>E-Bike</i>	<i>Car</i>	<i>BTM</i>	<i>Train</i>
Access to public transport (Origin)	-	-	-	0.81 (0.21)	0.26 (0.32)
Access to public transport (Destination)	-	-	-	0.81 (0.21)	0.26 (0.32)
Age: 0-17	0.22 (0.42)	0.22 (0.42)	0.22 (0.42)	0.22 (0.42)	0.22 (0.42)
Age: 18-40	0.25 (0.44)	0.25 (0.44)	0.25 (0.44)	0.25 (0.44)	0.25 (0.44)
Age: 41-66	0.34 (0.48)	0.34 (0.48)	0.34 (0.48)	0.34 (0.48)	0.34 (0.48)
Age: 67-100	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)	0.19 (0.39)
Car ownership	0.85 (0.27)	0.85 (0.27)	0.85 (0.27)	0.85 (0.27)	0.85 (0.27)
Travel distance	3.84 (3.69)	3.84 (3.69)	4.07 (3.78)	6.31 (5.91)	15.31 (20.67)
Travel speed	-	-	1.43 (2.78)	0.21 (0.13)	0.18 (0.15)
Car parking (Origin)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)
Dummy: residential zip code (Origin)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)
Car parking (Destination)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)	0.15 (0.31)
Dummy: residential zip code (Destination)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)	0.48 (0.5)
Education: Primary education	0.10 (0.30)	0.10 (0.30)	0.10 (0.30)	0.10 (0.30)	0.10 (0.30)
Education: vmbo/mavo	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)
Education: havo/vwo	0.38 (0.49)	0.38 (0.49)	0.38 (0.49)	0.38 (0.49)	0.38 (0.49)
Education: hbo/university	0.31 (0.46)	0.31 (0.46)	0.31 (0.46)	0.31 (0.46)	0.31 (0.46)
Frequency of BTM stops (Origin)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)
Frequency of BTM stops (Destination)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)	0.39 (0.43)
Gender	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)
Driver's License	0.64 (0.48)	0.64 (0.48)	0.64 (0.48)	0.64 (0.48)	0.64 (0.48)
Household Members	2.15 (1.30)	2.15 (1.30)	2.15 (1.30)	2.15 (1.30)	2.15 (1.30)
Occupation: Part-time job	0.11 (0.31)	0.11 (0.31)	0.11 (0.31)	0.11 (0.31)	0.11 (0.31)
Occupation: Full-time job	0.35 (0.48)	0.35 (0.48)	0.35 (0.48)	0.35 (0.48)	0.35 (0.48)
Occupation: Own housekeeping	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
Occupation: Student	0.13 (0.34)	0.13 (0.34)	0.13 (0.34)	0.13 (0.34)	0.13 (0.34)
Occupation: Unemployed / Unable to work	0.06 (0.24)	0.06 (0.24)	0.06 (0.24)	0.06 (0.24)	0.06 (0.24)
Street Density	0.32 (0.09)	0.32 (0.09)	0.28 (0.07)	0.10 (0.08)	-
Wealth: First 20% group	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
Wealth: Second 20% group	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
Wealth: Third 20% group	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
Wealth: Fourth 20% group	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
Wealth: Fifth 20% group	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)