Master Thesis

INTRAPERSONAL VARIABILITY IN MODE CHOICE BEHAVIOUR

A research based on data from the Dutch Mobile Mobility panel

Christian Beltman

December 2014

UNIVERSITY OF TWENTE.

Colophon

Title:	Intrapersonal variability in mode choice behaviour
Subtitle:	A research based on data from the Dutch Mobile Mobility Panel
Author:	J.C. Beltman (Christian) BSc. Civil Engineering & Management University of Twente E-mail: christianbeltman@gmail.com
Place:	Enschede
Date:	December 14th 2014
Status:	Final report

Keywords
Intrapersonal variability, mode choice

Supervisors

Prof. Dr. Ing. K.T. Geurs
Faculty of Engineering Technology
Centre for Transport Studies
University of Twente

Dr. T. Thomas
Faculty of Engineering Technology
Centre for Transport Studies
University of Twente

UNIVERSITEIT TWENTE.

PREFACE

This Master thesis concludes my study Civil Engineering and Management at the University of Twente. After eight years it is time to move on to a professional career and conquer challenges that are waiting for me in the future.

During my research I had a lot of support from several people who I like to thank. Firstly thanks to thank Joël Meijers and Sander van Weperen to provide a nice atmosphere to work in at the University of Twente. Our lunch walks and jokes during the day were a motivation for me to keep on working, even when times were tough. I wish you both the best in the future.

My gratitude also goes to my supervising committee consisting of Tom Thomas as daily supervisor and Karst Geurs as first supervisor. The extensive feedback and critical comments really helped to improve my research. The feedback and brainstorm meetings with Tom provided new insights and steps to proceed in my research every time. Your enthusiasm to explore all different directions in my research, lead me to the point where my thesis is now.

Furthermore I would like to thank Ron Michels, my badminton trainer from D.B.V. DIOK. I could drive back to Nijverdal twice every week with you, during which we had a lot of nice discussions and conversations. Being able to drive with you really kept me motivated to train a lot, which was always a nice way to relax from a busy day of work at the University.

Finally, I am very grateful for all the support of my family and friends during my graduation process.

Christian Beltman

Enschede, December 2014

SUMMARY

For years, researchers have been trying to understand the dynamics of travel behaviour. Having knowledge on an individuals' travel behaviour and its decision making process, provides insight in characteristics policy makers can influence in order to achieve a behavioural change. Analysis of travel behaviour is typically undertaken based on single-day data for each individual or household in a sample. Using single-day data does not capture the variability in the behaviour of an individual, since the measurement is performed only at one point in time. Therefore, research to intrapersonal variability in travel behaviour in the literature is scarce.

Most empirical research focussed on the intrapersonal variability of trip characteristics. Characteristics influencing intrapersonal variability in mode use have not been thoroughly researched yet. Two theories explain the dynamics in modal choice; the economists view using the utility theory and the theory of planned behaviour stating that habit plays a role in the mode choice process. The availability of a multi-day dataset collected in the Dutch Mobile Mobility Panel project makes it possible to analyse intrapersonal variability in mode choice behaviour. Therefore, the aim of this research is to:

Determine what characteristics influence intrapersonal variability in mode choice behaviour by exploring the available methods and apply this on the data collected in the Dutch Mobile Mobility Panel project.

To achieve this objective, literature was studied first to find what characteristics significantly influence the modal choice of individuals. These characteristics were grouped as individual, mode-specific, trip-specific and external characteristics resulting in the conceptual model for this research in Figure 1.

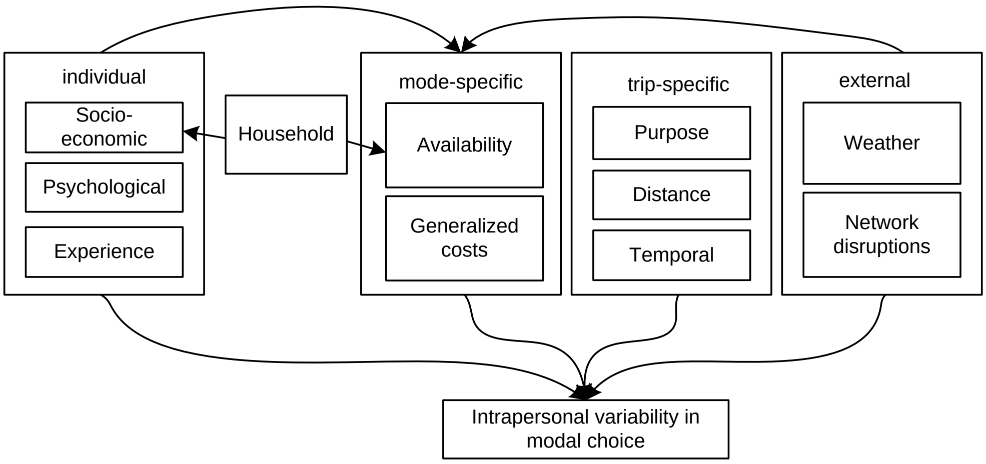

FIGURE 1: CONCEPTUAL MODEL OF CHARACTERISTICS INFLUENCING MODAL CHOICE

DATA

Data collected in the Dutch Mobile Mobility Panel project was used for this research and was collected using a smartphone application called MoveSmarter. This application tracked the travel movements of the panel members. The dataset contained the trip characteristics of the registered trips and socioeconomic characteristics of the panel members. Trip distance, weather data and location type characteristics were added using the postcodes of the registered trips. Respondents with a large share of unvalidated data, missing trips or missing postcodes were marked as unreliable. These unreliable respondents were removed from the dataset to create a valid dataset.

The validity of the resulting dataset was tested by comparing the distributions of socioeconomic characteristics to the distributions of the same characteristics of the LISS-panel and distributions of trip characteristics with trip characteristic distributions registered in OViN. Both are known representative sources of data. The dataset resulted to be representative which means that results from the analysis are valid.

ANALYSIS

An indicator of intrapersonal variability in mode choice behaviour was needed for the analysis. Kuhnimhof (2009) proposed a method expressing intrapersonal variability in mode use as a relative deviation from a state of maximum variation. This method calculates a Mode Variation Index (MIX) for every individual in a sample. Based on the values of all individuals in the sample, a cumulative distribution function is plotted, showing the intrapersonal variability in mode choice behaviour of a sample. An example of plotted cumulative distribution functions is shown in Figure 2 where the sample is split in three distance classes. Samples are compared using the median values of the cumulative distribution functions. A higher median value means a higher intrapersonal variability in mode choice behaviour of a sample

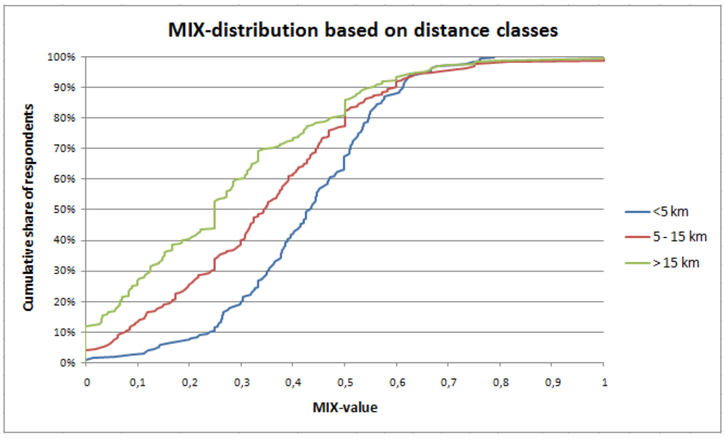

FIGURE 2: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, BASED ON TRIP DISTANCE

The effect of trip characteristics, socioeconomic characteristics and some other characteristics on the intrapersonal variability in mode use was analysed by plotting the MIX-distributions of these characteristics. Significance was tested using the two-sample Kolmogorov-Smirnov test. A significant effect on the intrapersonal variability in mode choice behaviour was found for trip distance, trip purpose, rain during trips and gender. Based on the analysis, it can be concluded that, apart from gender, socioeconomic characteristics of individuals do not influence the intrapersonal variability of a population. Additionally, using two weeks of collected data showed to be a proper quantity to capture intrapersonal variability in mode choice behaviour. No effects on the MIX-distributions were observed when extended data collection periods were used. The median values of the analysed characteristics are shown in Table 1.

Trip characteristics			**Socioeconomic characteristics**		
Purpose	Work & business	Ø=0	Urbanisation rate	Urban (1)	Ø=0,48
	Shopping	Ø=0,33		Slightly urban (2+3)	Ø=0,46
	Leisure	Ø=0,5		Rural (4+5)	Ø=0,42
Distance	0-5 km	Ø=0,43	Gender	Male	Ø=0,41
	5-15 km	Ø=0,35		Female	Ø=0,48
	>15 km	Ø=0,25	Partner status	Partner	Ø=0,43
Individual	Lower trip rates	Ø=0,42		No partner	Ø=0,47
trip rates	Higher trip rates	Ø=0,46	Age	<35 years old	Ø=0,47
Trip duration	0 - 20 min	Ø=0,40		35-55 years old	Ø=0,45
	20 - 60 min	Ø=0,43		>55 years old	Ø=0,43
	> 60 min	Ø=0,43	Education level	Higher educated	Ø=0,45
Other characteristics				Poorly educated	Ø=0,44
Rain	Rain	Ø=0,44	Income	Over modal income	Ø=0,44
	No rain	Ø=0,39		Below modal income	Ø=0,44
1 week data (1st)		Ø=0,41	Kids	No child in household	Ø=0,46
1 week data (5th)		Ø=0,40		Child in household	Ø=0,43
2 weeks data		Ø=0,44			
4 weeks data		Ø=0,44			
6 weeks data		Ø=0,44			
German mobility panel		Ø=0,34			

TABLE 1: RESULTS OF MIX-DISTRIBUTION ANALYSIS

Following the analysis of the observed intrapersonal variability in mode choice behaviour the intrapersonal variability in mode choice behaviour was simulated to test if intrapersonal variability in mode choice can be approximated by the use of a mode choice model. A multinomial logit model was estimated in BIOGEME based on the trips with the registered travel time for the car, bicycle and public transport. Travel time and trip distance were significant, as were the dummy variables for shopping trips and individuals living in a rural or slightly urban area.

A mode choice simulation was performed for the trips used to evaluate the model, using the estimated model parameters. This resulted in the availability of the simulated mode choice in addition to the observed mode choice. Following, the observed and simulated MIX-values were calculated for all individuals. MIX-distributions were plotted for characteristics expected to have an effect on the intrapersonal variability in mode choice behaviour, of which an example containing the MIX-distributions based on trip distance is shown in Figure 3.

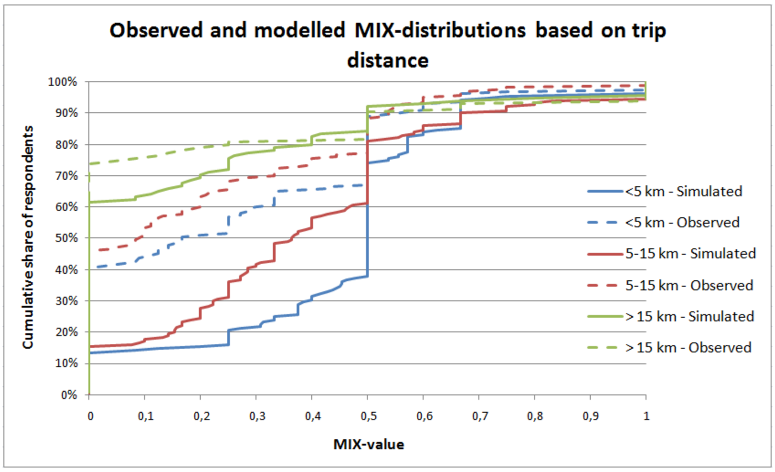

FIGURE 3: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON TRIP DISTANCE

The simulated MIX-distributions were ordered in the same way as the observed MIX-distributions. This shows that modelling using the utility theory captures the effects of characteristics on the intrapersonal variability in mode use. However, the MIX-distributions showed significant differences between the observed and simulated MIX-distributions for all analysed characteristics. A higher level of intrapersonal variability is expected according to the simulation, but is not observed. This systematic difference shows that intrapersonal variability in mode choice is difficult to model using the utility theory. Altogether this leads to the conclusion that both the economist view using the utility theory, as the theory of planned behaviour using habit, play a role in the mode choice process of individuals.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The topic of this master thesis is intrapersonal variability in mode choice behaviour. This topic is further introduced in this chapter. The first section describes the context of this research briefly. The second section describes two views regarding the dynamics in modal choice. The third section introduces the Dutch Mobile mobility panel, of which the collected data is used for this thesis. The objective and research questions for this research are defined in the fourth section followed by the research methodology in the fifth section.

1.1. CONTEXT

A transport policy aims to influence travel choices made by citizens, for example the choice for the mode of transport. One of the reasons for inducing a travel mode change is the large amount of congestion, as can be observed almost every single day. Congestion is regarded to have a negative effect on both the environment as well as the economy. Transport policy both aims to reduce congestion by trying to spread demand over time and inducing a modal shift. Instead of using the car, policies try to promote the use of more sustainable modes of transportation (Savelberg & Korteweg, 2011). These policies provide incentives for individuals to change their behaviour. Influencing the costs of travel has been effective in the past. An example of such a policy is peak hour avoidance, where regular users of a congested area receive a financial compensation for avoiding congested motorways during peak hours. As a result, some individuals switched to alternative modes of transport (Ministerie van infrastructuur en Milieu, 2011).

For years researchers have been trying to understand the dynamics of travel behaviour in order to develop new transport policies. To achieve behavioural change, knowledge of the determinants of travel behaviour is important. Knowing why an individuals' travel behaviour is the way it is, provides insight in characteristics policy makers can influence in order to achieve a behavioural change (Scheiner & Holz-Rau, 2013; van Wee & Dijst, 2002). In addition, a better understanding of travel behaviour makes it possible to identify groups that are not resistant to change and can be effectively targeted by different policies.

The modelling and analysis of travel behaviour is typically undertaken by using single-day data for each individual or household in a sample. These samples are usually collected by the use of cross-sectional travel surveys. Using single-day data does not capture the variability in the behaviour of an individual since the measurement is only performed at one point in time. The underlying assumption is that it does not matter on which day a person is sampled, because their travel on one day will be similar to any other day (Stopher & Zhang, 2011). It is assumed that conducting these surveys with a large sample size compensates for minor deviations. Collected data on an aggregate level provides a good figure of the travel behaviour of the entire population (Kitamura, 1990; Stopher & Zhang, 2011).

An important disadvantage of collecting data at a single point in time is that dynamics of travel behaviour are hard to be derived (Thomas, Bijlsma, & Geurs, 2013). Using cross-sectional data ignores the existence of day-to-day variation in travel behaviour on an individual level (e.g. a variation of destination or in mode choice); the data only allows researchers to explain differences in behaviour between individuals. An increased understanding of factors that cause variability in the travel behaviour of individuals provides valuable insights in the effects of policies. These insights can be provided by analysis of longitudinal data (data collected over multiple days). Usage of longitudinal data provides the opportunity to analyse if the behaviour of an individual changes over time and if so, analysis to the factors that caused this observed change. Due to the high costs of collecting longitudinal data, examples of longitudinal datasets and empirical research based on them are scarce. The empirical results mostly regard intrapersonal variability in trip generation characteristics only: number of trips, trip purpose (e.g. Pas & Koppelman (1986)) and trip distance, travel time, activity time and tour duration (Stopher & Zhang, 2011). Research to the intrapersonal variability in mode choice behaviour is left out of the scope.

1.2. Dynamics in mode choice

When looking at the research performed on collected datasets, conventional research to mode choice behaviour is derived from the random utility theory (Ortúzar & Willumsen, 2001). This theory assumes that a person chooses the alternative (i.e. mode) that provides the maximum utility and he/she is able to evaluate all alternatives, based on costs, time needed, comfort and other variables.

The approach to analyse modal choice at the individual level with the utility theory, is based on the use of discrete choice models. These models are able to parameterize utility functions for all available mode alternatives based on a set of explanatory factors (Whalen, Páez, & Carrasco, 2013). The models assume that an individual knows the values of characteristics for each alternative and is capable of identifying the alternative which provides them the highest utility. According to the utility theory, all individuals would therefore make the same decision in similar situations. However, this theory fails to explain why individuals in similar situations sometimes choose a different alternative.

Another view on travel behaviour is that habit plays a role in the decision making process (Aarts, Verplanken, & Van Knippenberg, 1998; Diana & Mokhtarian, 2009; Gärling & Axhausen, 2003). It is defined by Gärling and Axhausen (2003) as the repeated performance of behavioural sequences. Habit can be seen as a behavioural mechanism that goes against the decision-making process as described in microeconomic theory and thus the utility theory as mentioned previously. The theory assumes that the cost of searching for and constructing new alternatives is generally too high and the expected gains associated with new alternatives are too uncertain to even take them into consideration.

Also, the cognitive capacities of individuals are not sufficient to consider all alternatives and individuals therefore act out of habit (van Wee & Dijst, 2002). Habitual behaviour tends to make choices less deliberate and, in turn, causes variability in travel behaviour between individuals. Other mode choice research states that past behaviour is a predictor for future behaviour and that there are good reasons for assuming that habitual processes are at least partly responsible for the effect of past on current behaviour (Thøgersen, 2006).

Both abovementioned fields of research try to explain variability in modal choice behaviour. It is apparent that both theories account for a part of the variability in a persons' behaviour. However, both theories are not able to explain all the variability in an individuals' modal choice.

1.3. Dutch Mobile Mobility Panel

Recently, the University of Twente launched a research program to collect longitudinal travel data using a smartphone application; the related research program is called the 'Dutch Mobile Mobility Panel' (DMMP). In this panel, the travel behaviour of approximately 550 participants is automatically registered by a smartphone app called MoveSmarter. Chapter 3 deliberates on the method of data collection.

Panel members were selected from the Longitudinal Internet Studies for the Social Sciences (LISS) panel of CentERdata. Trip data of the panel members has been collected for two consecutive weeks in 2013 and four consecutive weeks in 2014, while the third data wave will be collected in 2015. The aim of the DMMP project is to analyse whether GPS enabled smartphones can be used to effectively and efficiently monitor the travel behaviour of individuals for extensive time periods. While it is not the aim of the project, the collected data provides the opportunity to analyse dynamics in mode choice behaviour.

1.4. RESEARCH OBJECTIVE AND RESEARCH QUESTIONS

Previous sections defined the context and the topic for this thesis. Section 1.2 discussed two opposed views describing the dynamics in mode choice and the DMMP-project was introduced in section 1.3. Based on these sections, the objective of this study is to:

Determine what characteristics influence intrapersonal variability in mode choice behaviour by exploring the available methods and apply this on the data collected in the Dutch Mobile Mobility Panel project.

To achieve the research objective four research questions are defined and are summed up below:

1. *What are important attributes that influence modal choice according to the literature?*
2. *What are the available methods to measure intrapersonal variability in mode choice behaviour?*
3. *Which characteristics explain intrapersonal variability in mode choice behaviour?*
4. *Can the observed intrapersonal variability in mode choice be approximated using a mode choice model?*

1.5. RESEARCH METHODOLOGY

This section elaborates on the research methodology used in order to achieve the research objective and answer the research questions. A schematic representation of the necessary steps for this research is shown in Figure 4. Three research phases are identified for this research; literature review, data collection and data processing and an analysis phase.

At first, a literature study will be conducted to identify the main characteristics causing mode choice variability. Secondly, methods to quantify intrapersonal mode use variability are identified. The third subject in the literature study discusses theory of choice modelling.

The second research phase contains all processes that are related to the data collection and data processing for this research. As section 1.3 already pointed out, the main data source for this research is the data collected in the first two waves in the Dutch Mobile Mobility Panel project. This data needs to be processed and enriched, after which proper quality data is selected.

The third research phase first tests the validity of the dataset by comparing distributions of characteristics in the dataset with the distributions of these characteristics in other datasets. Following, a descriptive analysis is performed providing insight in the effect of multiple available characteristics on the modal split. The intrapersonal variability in mode choice behaviour is then researched based on the method selected in the literature study. Finally a mode choice model is constructed to test if the intrapersonal variability in mode choice behaviour can be modelled accurately based on the available data.

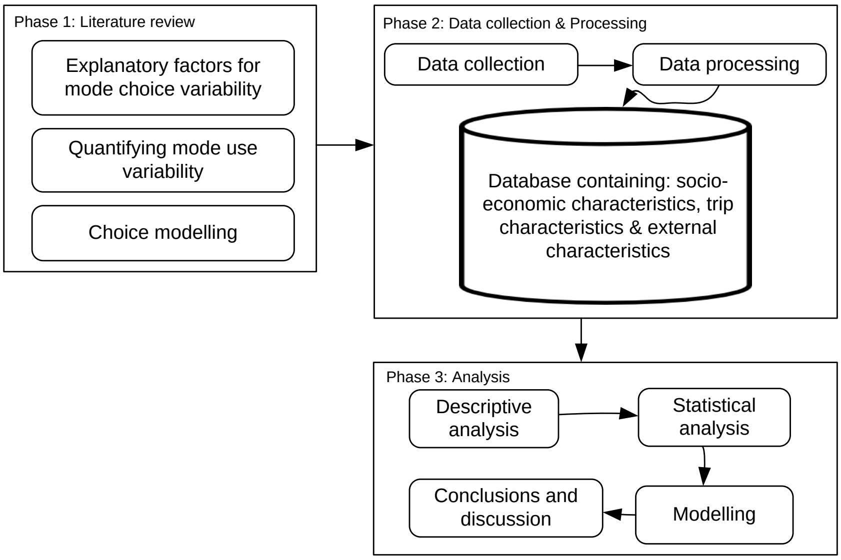

FIGURE 4: RESEARCH MODEL

1.6. CONCLUSION AND READING GUIDE

This report contains a master thesis analysing intrapersonal variability in mode choice behaviour. The research is based on the first two waves of data collected in the DMMP-project from the University of Twente, CentERdata and MobiDot. Data concerning the travel behaviour of panel members has been collected during a two-week period in 2013 and a four week period in 2014, which makes it a source for analysing intrapersonal variability in mode choice behaviour.

Chapter 2 describes the theoretical framework for this thesis. A conceptual model is constructed to show interdependencies between influential characteristics on mode use identified in the literature, and intrapersonal variability in mode use. Chapter 3 describes all processes involving the construction of a useable dataset for this research. Data collection, processing as well as data enrichment and data selection are described in this chapter. As a result of these processes, a dataset is constructed that is used for the analysis. The resulting dataset from chapter 3 needs to be representative in order to obtain valid results in the analysis phase of the research. Chapter 4 compares distributions of characteristics from the resulting dataset to distributions of these characteristics in representative datasets to make a statement of the validity of the resulting dataset.

Chapter 5 describes the analysis to the intrapersonal variability in mode choice behaviour. First, the correlation of available characteristics from the dataset with the modal split is analysed. This analysis is followed by an analysis of the dominant mode distribution. To provide insight in intrapersonal variability in mode choice behaviour, an analysis to the distribution of a variable called 'MIX' is also described. This variable will be introduced in the theoretical framework.

Following the descriptive analysis of the intrapersonal variability in mode choice behaviour, chapter 6 describes the construction of a MNL mode choice model. Using the simulated mode choices based on the estimated model parameters, a comparison is made between the observed intrapersonal variability in mode choice behaviour and the modelled intrapersonal variability. Finally, chapter 7 describes the conclusions for this research followed by the discussion on the results in chapter 8.

2. THEORETICAL FRAMEWORK

The introduction pointed out that there is a need to research intrapersonal variability in mode choice behaviour. This chapter provides insight in the current field of research and discusses a theoretical framework for this thesis. The first section explores modal choice and variability in travel behaviour based on the literature.

Modal choice is influenced by many characteristics. Therefore, the second section of this framework provides an overview of the explanatory characteristics for modal choice. Later chapters discuss what characteristics are included in this research. The third section describes the analytical framework for this thesis. Methods to quantify intrapersonal variability in mode choice behaviour are introduced followed by theory concerning mode choice modelling.

2.1. TRAVEL BEHAVIOUR AND MODAL CHOICE

Most literature describes the demand for travel as being a derived demand, based on an individual and a households' needs and desires (Ortúzar & Willumsen, 2001; Pas & Sundar, 1995; Pas, 1987; van Wee & Dijst, 2002). Individuals pursue activities at dispersed locations creating the need to make travel movements. Mokhtarian, Salomon and Redmond (2001) claim that this is not entirely true: they suggest that travel itself has an intrinsic positive utility and is therefore valued for its own sake, not just as a mean to reach a destination.

The field of study to travel behaviour has resulted in critical insights into the choices individuals and households make in their daily travel behaviour. These insights contributed in the development of increasingly sophisticated models to predict travel behaviour and changes to it (Clifton & Handy, 2001).

When considering mode choice, conventional mode choice models are discrete choice models. These models predict the probability a certain mode will be used for a trip. The probability for a mode to be selected is a function of an individual's socioeconomic characteristics in combination with the relative attractiveness of the alternatives (Ortúzar & Willumsen, 2001). The relative attractiveness is expressed in terms of utility; every individual seeks to maximize its utility and chooses the alternative with the utility level exceeding the utility level of the other available alternatives.

Diana and Mokhtarian (2009) disagree on this theory. According to their research, habit plays a role in the decision making. Habit can be seen as a behavioural mechanism that goes against the decision-making process as described in microeconomic theory, and thus the utility theory (Diana & Mokhtarian, 2009). Habit tends to make choices less deliberate, which causes variability in travel behaviour between individuals as also shown in the process model of making mode choices by Aarts et al. (1998) in Figure 5. Individuals with a weak habit consider their choice options, while individuals with a strong habit are less guided by such considerations.

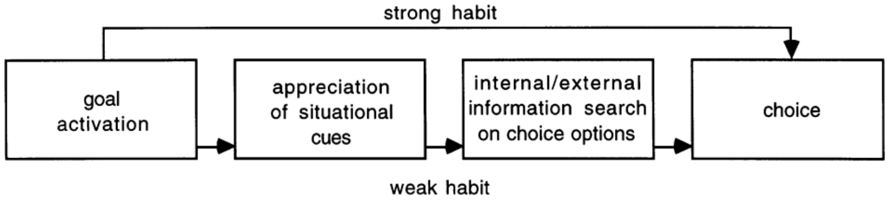

FIGURE 5: PROCESS MODEL OF MAKING CHOICES BY WEAK AND STRONG HABIT INDIVIDUALS (FROM AARTS ET AL.(1998))

2.1.1. VARIABILITY IN TRAVEL BEHAVIOUR

A distinction can be made between intrapersonal and interpersonal variability when regarding variability in travel behaviour. Intrapersonal variability is a variation in the behaviour of a single individual, while interpersonal variability describes the difference in behaviour amongst persons (Pas & Sundar, 1995). Pas (1987) identified that intrapersonal variability is either systematic (due to day-of-the-week variability) or random. The systematic, day-of-the-week, variability is explained by the fact that the demand for travel is derived from the need of individuals to perform activities at dispersed locations. For instance, most individuals do not buy groceries every day but for multiple days at once. Shopping trips are therefore not made every day. The framework by Pas is shown in Figure 6.

Research concerning intrapersonal variability has received little attention in literature. The main reason for this is that most available data sets only contain single-day data and thus research to intrapersonal variability is ruled out. The attention it did receive in literature, mostly considered intrapersonal variability in trip generation characteristics.

One example of research to intrapersonal variability is research performed by Pas. Pas (1987) examined the effect of intrapersonal variability of travel behaviour on the goodness-of-fit of least squares trip generation models. The empirical analysis, which was based on a 7-day activity survey from 1973 in Reading, showed that half of the variance is explained by within-person variability. The unit of analysis was the day-to-day variability in daily trip rates. This work was later extended to intrapersonal variability in travel time per day and travel time per trip per day (Pas & Sundar, 1995). The study also showed that intrapersonal variability explained significant portions of the overall variability.

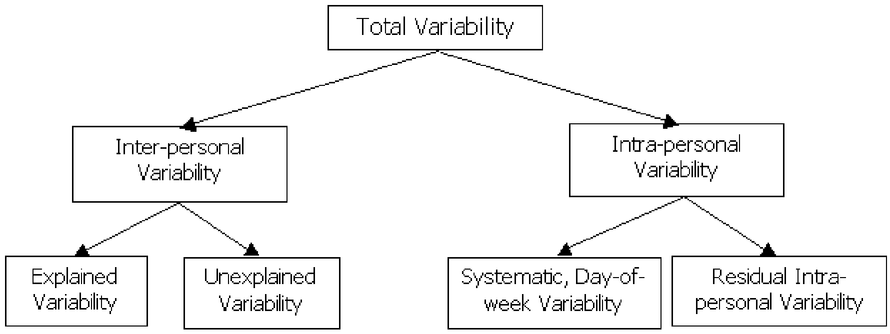

FIGURE 6: COMPONENTS OF VARIABILITY: BASIC CONCEPTS (ADAPTED FROM PAS (1987))

Later work by Pas and Koppelman (1986) found that the level of intrapersonal variability in daily trip frequency varies between demographic groups. An employed person, for example, shows lower levels of intrapersonal variability compared to individuals who are not employed outside the home.

Schlich (2001) used a sequence alignment method to analyse intrapersonal variability in travel behaviour. In his research the Mobidrive dataset was used consisting of 6-week data collected in the fall of 1999 in Germany. In this research, the unit of analysis was the daily distance covered by the respondents. The research did not take the number of trips undertaken by travellers into account.

2.1.2. Repetitiveness of travel behaviour

Stopher and Zhang (2011) analysed the repetitive patterns of travel behaviour, based on a classification into twelve different types of tours. The data used in their research came from two Australian panels where GPS-data was collected. Registered trips were reclassified into tours and during their research the number repetitions of each tour type was counted. For every tour type, a rate of repetitiveness for tour distance, travel time, activity time and tour duration was calculated. They concluded that the repetitiveness which underlies all travel demand modelling in current practice is not present. According to their work the underlying assumption that travel is repetitive from day to day is highly suspect.

Schönfelder and Axhausen (2010) analysed temporal aspects of travel based on multiple multi-week datasets. Unit of analysis was the number of new locations visited per individual. All analysed datasets showed that even after 26 weeks, still unobserved locations to that time were visited. Applying this knowledge to this research, it is expected that the observed variability during the six weeks of data collection only captures a small piece of the total variability. Furthermore, Schönfelder and Axhausen (2010) state that most of the 'new' locations were not genuinely new. Some trips are repetitive but made infrequently; for example, a visit to the dentist.

Kenyon and Lyons (2003) analysed the habitual attitude in mode use. In their research respondents were asked to state their modal choice for familiar trips at different times, with different levels of information about the trips shown every time. Their results illustrates that a majority of travellers do not consider their modal choice for the majority of the trips undertaken.

2.2. EXPLANATORY CHARACTERISTICS FOR VARIABILITY IN MODE CHOICE BEHAVIOUR

Modal choice is influenced by many different characteristics. Several studies have analysed modal choice and its explanatory characteristics, including Thøgersen (2006); Kroesen (2014); Chikaraishi, Fujiwara, Zhang, & Axhausen (2011); Ortúzar & Willumsen (2001); Sabir, Koetse, & Rietveld (2008); Scheiner (2010); Van Exel & Rietveld (2010) and Bergström & Magnusson (2003). Figure 7 shows a conceptual model of the characteristics influencing modal choice, based on these studies. These characteristics need to be incorporated in the analysis and should be corrected for in the modelling stage of this study. Ortúzar and Willumsen (2001) divided these factors in three main groups; characteristics of the trip maker, characteristics of the journey and characteristics of the transport facility. These three main groups do not account for external characteristics such as weather. The characteristics found to influence modal choice are therefore categorized in personal characteristics, trip characteristics, mode characteristics and external characteristics.

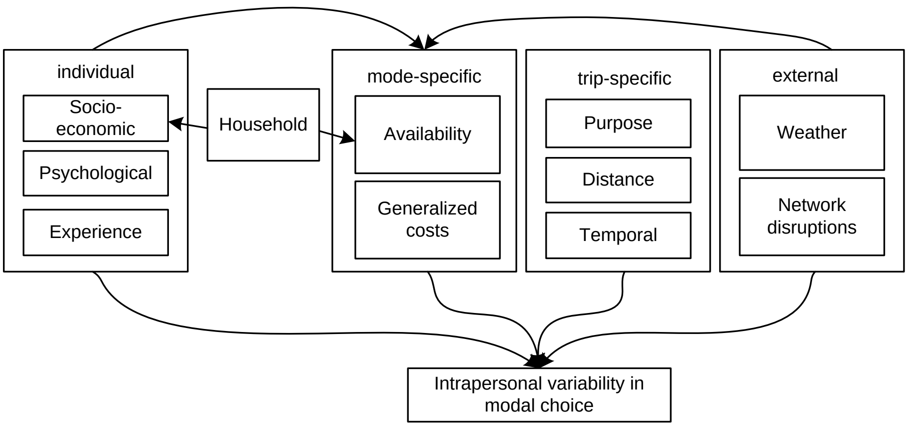

FIGURE 7: CONCEPTUAL MODEL OF CHARACTERISTICS INFLUENCING INDIVIDUAL MODE CHOICE

Figure 7 shows that a household influences the socioeconomic characteristics as well as the mode availability; for example, a household can have a double income. These characteristics increase the mobility budget for this household (and thus the individual). Additionally, having children influences the travel behaviour of their parents as well. Also, intra-household interactions influence mode availability. Individuals organize themselves in households in order to share resources, such as transport modes (car, bicycles) and play different roles in the household. The following characteristics are derived from the literature:

- Car ownership;
- Household interactions – Dependent on car ownership. When a double household has one car, this car needs to be shared. Sharing the vehicle causes variability in an individuals' mode choice behaviour;
- Household income - the probability of using the car is high for individuals with high household incomes. Even so, the probability of selecting walking and biking reduces with increasing household income (Sabir et al., 2008).

2.2.1 INDIVIDUAL CHARACTERISTICS

Many studies used individuals' socioeconomic characteristics as control variables in their research. Socioeconomic characteristics are mainly used as variables to control for. These variables are trivial and are therefore not further explained. Psychological variables as perception and intention and are further explained in this section. The following characteristics are used for this research:

- Age;
- Gender;
- Education;
- Occupation.

2.2.1.1. PERCEPTION

The perception of an individual plays a role in their mode choice behaviour. A car drivers' perception of the quality of alternative travel modes is a barrier for including these alternatives in their personal modal choice sets. Research by van Exel and Rietveld (2010) showed that the car users' perceptions of public transport travel time deviates substantially from the real travel time. The ratio between the perceived travel time by public transport and the reported travel time by the respondents was 1 : 2.3, which is a substantial difference. These deviations are partly explained by the familiarity of a car user with characteristics of the car trip and the unfamiliarity with public transport system.

Van Exel and Rietveld (2010) also asked respondents whether public transport could have been used for the same trip. Three response categories were distinguished: no, yes but rarely do and yes regularly. Their results showed that when public transport is not an option, the ratio of the perceived travel time to travel time by car is 1:2.5, for moderate users this is 1:2.0 and regular users 1:1.6. Familiarity with the public transport system therefore decreases the perceived travel time by public transportation and makes public transportation a more attractive alternative to consider in the modal choice.

Bamberg et al. (2003) showed that modal choice is dependent on perception and knowledge of alternatives. Their study targeted individuals with plans to relocate. The respondents received detailed information explaining the public transportation services in their new living area as well as a free ticket valid for one day to test the public transport system. Six months after the relocation, the travel behaviour of the respondents was evaluated. Results showed an increase in public transport use from 12.8% to 29.3% of the trips. Furthermore, the share of private car use declined from 55.5% to 41.8% of the trips. The intervention to provide information and provide a free one-day ticket showed to be effective in increasing public transport, with only the number of trips as variable. No corrections were made for urban density and public transport frequencies at either the former residence or the new one. The intervention during this research showed to be effective in inducing modal shift. The research stated that individuals need to reconsider their travel patterns when moving residences; individuals are open to change to other modes of transportation during this reconsideration.

2.2.1.2. INTENTION

Another psychological factor in mode choice behaviour is the intention of an individual. A theory adapted from psychology is the theory of planned behaviour (Azjen, 1991). The theory of planned behaviour emphasizes the deliberate character of the choice of an individual (Aarts et al., 1998) and holds that behavioural intent has three predictors; a schematic representation of this theory is shown in Figure 8:

- Attitude - The degree to which a person has a favourable or unfavourable evaluation or appraisal of the behaviour in question;
- Subjective norm – The perceived social pressure to perform or not perform a certain behaviour;
- Perceived behavioural control – An individuals' perception of the ease or difficulty to perform the behaviour.

The following general rule applies: the more favourable the attitude, subjective norm and perceived control, the stronger the person's intention to perform the behaviour. Applying this theory on mode choice behaviour leads to the following statement:

> *"Individuals who believe that they have neither the skills nor the opportunities to drive a car are unlikely to form intentions to engage in this behaviour, even if they hold favourable attitudes and experience social pressure to use the car"* - Aarts et al. (1998)

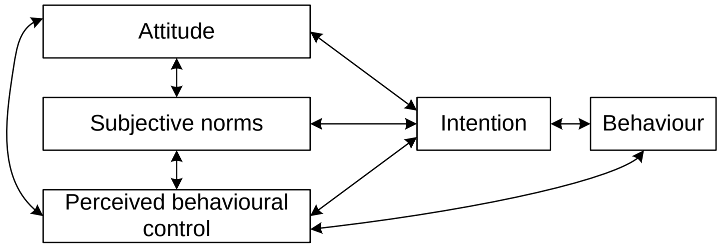

FIGURE 8: THE THEORY OF PLANNED BEHAVIOUR (ADAPTED FROM (AZJEN, 1991))

2.2.2 TRIP-SPECIFIC CHARACTERISTICS

Ortúzar and Willumsen (2001) distinguished two trip-specific characteristics which influence mode choice:

- Trip purpose;
- Time of day.

In addition, Mintesnot and Takano (2005) researched seasonal influences on modal choice of high school students in snowy regions in Japan. The research showed a significant modal shift between summer and winter for high school students. During the summer, the majority of the students (85%) used the bicycle for trips to and from school. The bicycle share decreased during the winter (to 0.35%) and students shifted to the bus (increase from 5% to 44%), walking (increase from 4% to 23%) and car passenger (increase from 3% to 24%). The research showed that mode choice is not only dependent on time of day; time of year also influences mode choice. However, mode choice variability due to seasonal effects exists due to the related weather conditions. Weather conditions are considered as external characteristics in this study and are discussed in section 2.2.4.

Furthermore, interrelationships between travel mode choice and trip distance were researched by Scheiner (2010). His research was based on the KONTIV data of the years 1976, 1982, 1989 and 2002. The research pointed out that mode choice is influenced by the trip distance. Since data at multiple instances was used, the temporal component was researched as well. It was shown that the modal split changes over time within the same distance category and it was concluded that car use has considerably increased over time.

Finally, Scheiner (2010) analysed the effect of city size on the modal split also based on KONTIV data. The research showed a decrease in car use with an increasing city size. Both the non-motorised transport modes and public transport shares increased with an increasing city size.

2.2.3 MODE-SPECIFIC CHARACTERISTICS

Ortúzar and Willumsen (2001) divide mode-specific characteristics which cause variability in mode choice behaviour in quantitative and qualitative characteristics. First, they identified quantitative characteristics:

- Relative travel time: in-vehicle times, waiting times and walking times by each mode;
- Relative monetary costs;
- Availability and cost of parking.

Second, they identified qualitative characteristics which are less easy to measure:

- Comfort and convenience;
- Reliability and regularity;
- Protection, security.

Additionally, mode availability influences modal choice. Mode availability includes the ease of accessibility of a mode to the trip maker. For public transport, the routing system and distances to activity locations from the public transport system need to be considered. A dense network with high frequencies provides a relatively higher availability compared to lower frequencies or a sparser network. Sheth (1976) states that private transportation modes provide a higher relative availability factor compared to public transport. Consequently, the functional and situational utility of the private car is higher.

2.2.4 EXTERNAL CHARACTERISTICS

Sabir et al. (2008) noted a marginal effect of urbanization on modal choice: the private car is preferred to all other modes in rural areas. No explanation for this marginal effect was provided in their research. However, it can be argued that low urbanization rates lead to more long-distance trips, which in turn leads to a decrease in walking and bicycle trips. Also, the public transport network is less dense and has lower frequencies in rural areas compared to the frequencies in cities. Altogether, this makes using public transport a less favourable mode alternative.

Another external characteristic influencing mode choice behaviour is the weather condition. Sabir et al. (2008) analysed the impact of weather conditions on mode decisions on an individual level. Their research was based on a travel diary conducted amongst 600.000 individuals in the Netherlands in 1996 and weather data from the Royal Netherlands Meteorological Institute (KNMI). These two datasets were linked in such a way that each trip observation was assigned the weather conditions of the hour in which the trip took place, with data from the nearest weather station. It was assumed that the trip maker based the mode choice on the weather conditions at the moment of departure. Four Multinomial Logit Models were estimated based on this data. One model was estimated to analyse the impact of weather conditions on mode choice, four other models were estimated for different trip purposes.

The results showed an increasing share of bicycle trips with an increase in temperature. Also, precipitation of over 1 mm per hour induced a significant modal shift from the bicycle towards the car. Added to that, the effect of wind was analysed by adding a dummy variable for wind strengths above 6 BFT. The significant effects showed that bicycle use decreased by 3.8%, which shifted evenly to walking and the car.

Additionally, Bergström and Magnusson (2003) analysed the attitude towards cycling during the winter. This research was based on surveys conducted in 1998 and in 2000 among employees at four major companies in two Swedish cities, Luleå in the north of Sweden and Linköping in the south. Large seasonal differences were observed in the modal split. During the winter an increase of 27% in car use was observed and a decrease of bicycle trips by 47%. By improving winter maintenance service levels on cycle ways, the potential increase in bicycle use was estimated to be 18%, representing a decrease of 6% in the number of car trips.

Concluding from this research the following characteristics influence modal choice:

- Precipitation;
- Temperature;
- Wind strength.

Disruptions in the transport network also influence an individual's mode choice. Zhu and Levinson (2012) reviewed empirical studies on traffic- and behavioural impacts of network disruptions. Different studies they reviewed showed that network disruptions changed travel behaviour. Most individuals changed route and departed earlier to make the trip. Also, the public transport ridership increased marginally when the road network was disrupted.

Van Exel and Rietveld (2009) analysed the effects of a one-day, pre-announced rail strike in the Netherlands. The complete strike led to a reorganization of travel behaviour, including a switch of train users to the car. For the travellers intending to go by train, 24% switched to car as a driver, 14% switched to another mode (as passenger), 18% stayed with the train and rescheduled the planned activity to another day. In addition, Klöckner and Friedrichsmeier (2011) state that major disruptions, which are partly known before the trip is started, lead to a reorganization of travel behaviour.

2.3 ANALYTICAL FRAMEWORK

This section describes the analytical framework for this research. First, quantification of mode use variability is discussed. The second subsection describes the theory of the multinomial logit model and describes how different estimated models can be assessed and compared.

2.3.1. QUANTIFYING VARIABILITY IN MODE USE

As stated in the introduction, intrapersonal variability in mode choice behaviour is analysed in this thesis. Other intrapersonal variability studies in the context of transportation (i.e. Pas (1987)) analysed the intrapersonal variability of continuous trip characteristics as trip duration and trip production. Intrapersonal variability was expressed as being the standard deviation from the average observed behaviour. Since mode use is not continuous but a frequency variable, this method of measuring intrapersonal variability cannot be applied to this research. Two methods of expressing intrapersonal variability of mode use were found in the literature and are discussed next.

Studying determinants of multimodal travel in the USA, Buehler and Hamre (2013) distinguished four mode use categories for individuals; monomodal car, multimodal car, monomodal green and multimodal green. Individuals were categorized by the group they belonged to based on observed data from the 2001 and 2009 National Household Travel Surveys in the United States. Intrapersonal variability in mode use was expressed as the share of respondents in one of the four categories, based on different observation time spans. For instance, the share of monomodal car users decreased when the observation period was increased. As an effect, the share of multimodal car users increased when the amount of time observed increased. These results were logical since the probability of observing the use of another mode increases in time. While this measure shows that multimodality increases with an increasing observation time span, this measure did not show how often other modes were used.

Kuhnimhof (2009) used a different calculation method of intrapersonal variability of mode use, resulting in a Mode Variation Index (MIX) for every individual. This method calculates the difference between observed mode use frequencies and modelled mode use frequencies in a state of maximum variation. A MIX-value of 0 for an individual means that only one used mode was observed. A value of 1 means that the observed frequencies match the frequencies as in a state of maximum variation. The method to calculate the MIX-value is described in Appendix A.

A cumulative distribution function is constructed based on the MIX-values of all individuals which show the distribution of the MIX-values across the sample population. In order to compare samples and situations, the median values of the MIX-distributions are used. Based on data from the German Mobility Panel, the MIX-value distribution for the German context was constructed as displayed in Figure 9. Lower MIX-values were associated with trips made predominantly in a rural area by car, while higher values were associated with trips predominantly made in urban areas by public transport or by bike.

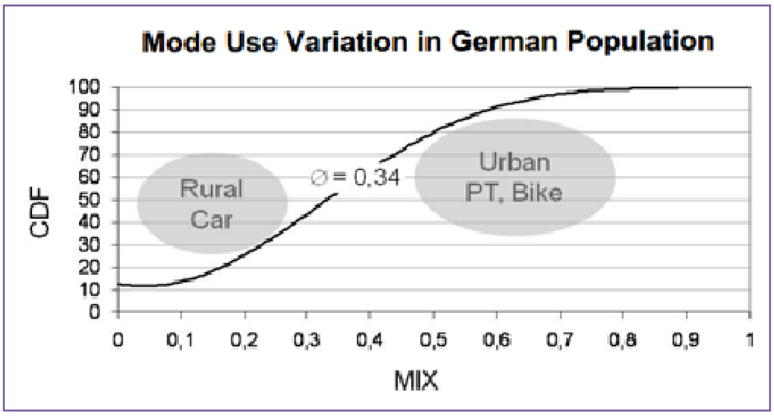

FIGURE 9: CUMULATIVE DISTRIBUTION FUNCTION OF THE MIX-VARIABLE

2.3.2. MULTINOMIAL LOGIT MODEL

Discrete choice models are used to describe a decision-makers' choices among alternatives. These decision makers are assumed to choose from a finite set of mutually exclusive alternatives. According to the random utility theory an individual n chooses the alternative j which maximizes the utility or minimizes their disutility. The utility U_{nj} of an individual n with chosen alternative j can be expressed by the following equation:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad \forall j$$

where V_{nj} is the systematic component containing the observable portion of the utility for the modeller, and ε_{nj} is a random error component. The systematic component is often specified to be linear in parameters with a constant:

$$V_{nj} = \alpha_j + \beta' x'_{nj} \;\; \forall j$$

where $x'_{nj}$is a vector of characteristics for alternative j for the individual n, β are the parameters to be estimated and describe the relative importance of the parameter and α_j is a constant specific for alternative *j.* This thesis concerns the travel mode choice for individuals. Using the previous equations the probability an individual chooses travel mode *j* is:

$$P(Y_n = j) = P\big(U_{nj} > U_{nk}\big) = P\big(V_{nj} + \varepsilon_{nj} > V_{nk} + \varepsilon_{nk}\big) \text{ for all } k \neq j$$

Multiple models are available to estimate the travellers' utilities, but the multinomial logit model (MNL) is considered the simplest and most popular discrete choice model. The MNL model assumes that the residuals are independent and Gumbel distributed (Ortúzar & Willumsen, 2001; Train, 2002). With this assumption the probability an individual chooses alternative j is calculated by:

$$P(Y_n = j) = \frac{e^{(V_{nj})}}{\sum_{k=1}^{K} e^{(V_{nk})}} \quad \text{with } K \text{ the number of alternatives.}$$

The maximum likelihood method is most often used to estimate the parameters of a MNL model. This method is based on the idea that although a sample could originate from several populations, a particular sample has a higher probability of having been drawn from a certain population than from others (Ortúzar & Willumsen, 2001). The ML estimates are the set of parameters which will generate the observed sample most often. The maximum likelihood estimations for this research are performed using the software package BIOGEME (Bierlaire, 2003)

2.3.2.1. Model goodness of fit

Multinomial Logit Models are constructed iteratively. First, a simple model is estimated that contains only a few attributes. Following, other MNL models are estimated in which other parameters are introduced one by one. The addition of extra parameters does not necessarily improve the model. To determine how well the estimated model fits the data, a statistic called the likelihood ratio index is often used (Train, 2002). This statistic is defined as:

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

Where $LL(\hat{\beta})$ is the value of the-log likelihood function with the estimated parameters and $LL(0)$ is its value of the log-likelihood function when all parameters are set to zero (Train, 2002). The ρ^2-values range from zero to on. A value of one indicates that the estimated parameters predict the choices of the sample population perfectly. Values of ρ^2 between 0.2 and 0.4 are considered to be indicative of extremely good model fits (Louviere, Henscher, & Swait, 2000).

2.3.2.2. Comparison of models

The likelihood ratio test is used in statistics to compare the fit of two models, one of which is nested within the other (Ortúzar & Willumsen, 2001; Train, 2002). This test analyses if the addition of an extra parameter significantly improves the model. For this test, the log-likelihood of the base model ($LL_{basemodel}$) is compared with the log-likelihood of the model with the added parameter (LL_{added}). The LR statistic

$$-2(LL_{addedP} - LL_{basemodel})$$

is asymptotically distributed χ^2 with r degrees of freedom, where r is the number of added parameters. A higher value of the LR-statistic than the χ^2 values with r degrees of freedom means a rejection of the null hypothesis. The model with the added parameter is then considered to be significantly better than the original model.

2.4 Conclusion

This chapter described the theoretical framework for this thesis. The first section described modal choice, variability in mode choice behaviour and current research in that field. The second section described characteristics that, according to the literature, influence the modal choice of individuals. A conceptual model showing the interrelationships between these characteristics was shown in Figure 7. The third section described the analytical framework for this thesis where first a method to quantify mode use variation was described, followed by theory of MNL modelling.

3. DATA COLLECTION AND DATA PROCESSING

This chapter elaborates on the data collection and processing process. The first section describes the data collection process used in the DMMP-project. The second section describes the process of data cleaning and noise reduction. The third section describes the data enrichment process followed by the data selection process in the fourth section. The fifth section elaborates on the addition of the dataset collected in 2014. Finally, this chapter ends with a discussion on some limitations of the constructed dataset.

3.1. DATA COLLECTION USING MOVESMARTER APP

The University of Twente has set up a research project in cooperation with Novay and Mobidot in order to study travel behaviour observed from panel members. The aim of this research project is to examine whether GPS-enabled mobile phones are an effective and efficient method to use for collecting travel behaviour data. This project also examines the variability and stability of individual travel behaviour over time. In order to investigate these issues, a three year field experiment will be conducted which started in 2013. At the moment the first two waves of data have been collected which are both used for this research. Data in this field experiment is collected over a period of multiple weeks using a smartphone app called MoveSmarter. The panel used for this research is a sample from the LISS panel of CentERdata, which is based on a true probability sample of households drawn from the population register by Statistics Netherlands.

Collecting data consists of two consecutive methods which will be further described in this section and is visualised in Figure 10. First, MoveSmarter registers movements of panel members. Second, the panel members need to verify the registered trips on a web interface from CentERdata. This research is based on the trips as verified by the panel members.

3.1.1. TRIP REGISTRATION

Data in this research is collected using various sensors of modern-day smartphones (GPS, WIFI and cell-ID) (Thomas et al., 2013). Using a smartphone for data collection has the advantage that most individuals owning a smartphone always take the phone with them and make sure it is turned on. In this way, non-response due to a forgotten or non-active GPS datalogger is likely to decrease, which has been a common source of measurement failure in previous studies (Wolf, 2000).

The MoveSmarter has been developed by Novay and Mobidot for both Android and iPhone smartphones in order to detect travel movements of the panel members. The application consists of two modules: the sensing module and the back-end module.

The sensing module uses the available sensors of the smartphone. It automatically measures all the movements the respondents make. It also uses these sensors to optimize battery performance. For example, the GPS-sensor does not need to be enabled when no movements are detected. The trips detected by the sensing module are transferred to the back-end module.

The back-end module transforms the collected data by the sensing module. It collects geo-traces and transforms these into trips. Using algorithms, the trip purpose and used mode are derived (Bohte, 2010). When multiple modes are used during one journey, this journey is split into multiple trips based on the used modes (Thomas et al., 2013).

3.1.2. TRIP VERIFICATION

Respondents are asked to verify the collected data from the trip registration process on the CentERdata web interface. Measured trips are displayed on this web-page with their trip characteristics. In case of erroneous trip registration, respondents are asked to alter the trip characteristics in order to verify the registration process. A fixed number of choices are available for the travel mode and purpose. This choice set is derived from the MON and OViN research projects. OViN is described in section 4.1. The number of possible choices in the DMMP project was smaller compared to OViN; choices with small shares which are normally added to the ‘other’ category were already considered as being ‘other’. In addition to altering trips and their characteristics, respondents have the opportunity to add, split or delete trips when MoveSmarter did not register the trips correctly.

The online review provides the respondents the opportunity to add data which might have influenced the travel behaviour in addition to verifying the trip characteristics. A list of special circumstances is provided to the respondents which consist of personal influences (e.g. illness) and other characteristics (e.g. road works, weather influences etc.). The respondents were asked if, and in what way, their travel behaviour was influenced when special circumstances occurred.

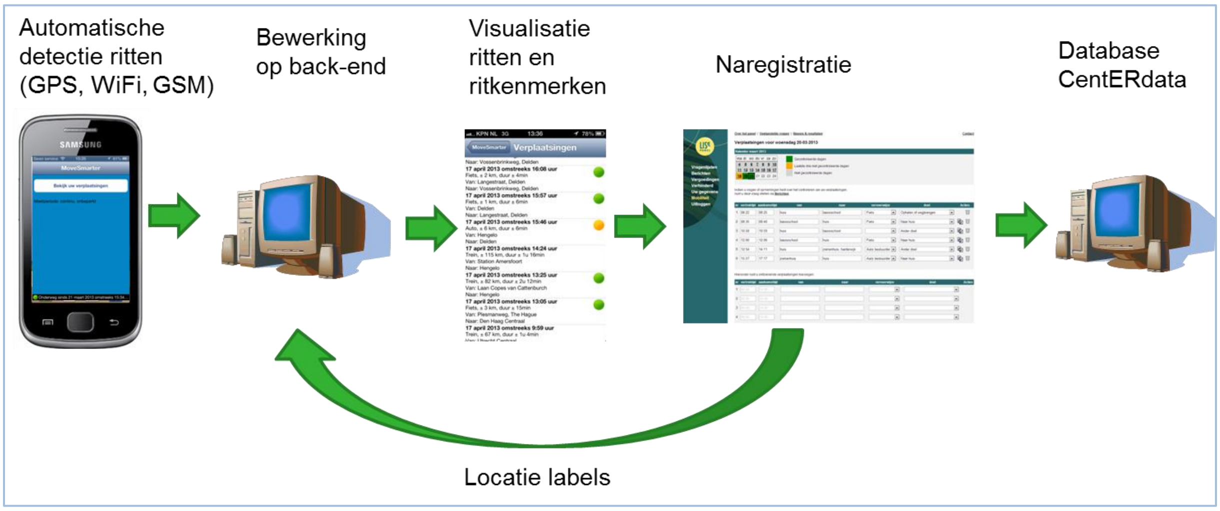

FIGURE 10: SCHEMATIC DISPLAY OF THE DATA COLLECTION PROCESS FOR THE DUTCH MOBILE MOBILITY PANEL

3.2. DATA CLEANING AND NOISE REDUCTION

The database containing all verified trips showed a great diversity in response manners by the respondents. A database containing consistent entries is needed in order to perform a good analysis. Data cleaning and noise reduction was performed to create this consistent database. Trips with possible noise were selected based on several criteria:

- ***Foreign trips***
  Foreign trips were identified in three separate ways. First, GPS-coordinates clearly outside the Netherlands were selected as being foreign. Second, postcodes in a number format which was different compared to the Dutch number format were selected as being foreign. Third, during the selection of respondents who registered trips properly, which will be described in section 3.4, yet unselected foreign trips were encountered. These trips were also appointed as being foreign.
  Foreign trips were identified and excluded since the data enrichment was performed using postcodes, as described in section 3.3, and no foreign postcodes were available in the additional datasets.
- ***Trips with a negative travel time***
  Trips in the database contained departure and arrival times. Subtracting the departure time from the arrival time resulted in the trip travel time. Analysis of the extreme values showed the presence of trips with negative travel times. These trips were not selected for further analysis.
- ***Double data entries***
  The data used for this study contained all trips verified by the respondents. Trips were saved twice on some occasions, which resulted in two trips in de dataset with exactly the same features. These double trips were deleted from the dataset.

- ***Manual encountered oddities***
  The last method to reduce noise in the dataset existed of selecting respondents who did not register their trips properly. For example, the origin of trips did consequently not match the destination of the previous trip. Multiple trips were therefore not registered and these respondents were marked as being unreliable data sources. This process was completed manually since the great diversity in manually entered origins and destination prevented automation. Section 3.4 further elaborates on this issue.

3.3. DATA ENRICHMENT

The database with the verified trips did not contain a lot of characteristics which can be of influence on the intrapersonal variability in mode choice as discussed in the theoretical framework. The dataset was enriched in order to work with these characteristics. This section discusses the data enrichment processes used.

3.3.1. INSERTED POSTCODES

The 25679 verified trips did not all contain a postcode as location data to work with. 7785 trips only contained manually entered descriptions of the origin and destination addresses. Some of these cases contained either the origin postcode or the destination postcode but not the complete O-D pair. Because other additional data could be added based on postcodes it was important to complement the list by inserting missing postcodes.

For this process a list containing all postcodes in the Netherlands on a postcode-6 was found online [www.postcodedata.nl]. This list contained the name of the street, town and the longitude and latitude of every postcode in the Netherlands. By transforming the (manually entered) descriptions of the origin and destination addresses the name of the street and town were separated into two columns. Missing postcodes were found and added to the trips by searching for the street name and the name of the town in the postcode-6 list. No matches were found when the transformations of the origin and destination did not result in a proper address. Resulting from this process, 3249 origins and 3223 destinations were added to the data resulting to 3255 additional O-D pairs with both origin and destination postcode.

Additionally, another method was used to add more missing postcodes. Individuals visited some locations multiple times during the data collection period, being either an origin location or a destination location. For the trips with a missing postcode it was analysed if the location was already visited as either an origin or a destination. The associated postcode was then added to the trip.

The same process was repeated on the trips collected in the second data wave when the data became available, to add missing postcodes for those trips as well.

3.3.2. ADDED WEATHER DATA

During the data collection period, the weather conditions for the trips were collected alongside the trip data. These weather conditions were lost when the trips were verified by the respondents. The choice was made to append KNMI-data to the trips. Data from the nearest weather station from the origin location at the hour of departure was appended to the trips. The process of adding weather data is described in more detail in Appendix B. All characteristics as described in the theoretical framework were present in the KNMI dataset.

3.3.3. ADDED LOCATION TYPES/ACCESSIBILITY OF THE LOCATION

Mode choice is dependent on the accessibility of certain modes. Accessibility of some modes is dependent on the location. For instance, the train is a more attractive alternative when the station is fairly close to both the origin and the destination. The Planbureau voor de Leefomgeving (2003) created a database containing the distance towards a railway station (station, intercity station and intercity junction), BTM station and motorway onramp for every postcode. It also contained the location type of the postcode is in this database based on ABCR classification.

3.3.4. Characteristics of mode alternatives

Mode choice is dependent on the availability of alternative travel modes. An individual makes the choice of mode use based on a set of available travel modes. Unavailable travel modes will quickly be eliminated during the mode choice process.

No questions regarding mode availability and mode ownership were asked during the data collection period of the DMMP-project. Therefore, it is unknown what modes are available to the respondents and assumptions need to be made. Since the Netherlands is a country with around eighteen million bicycles it is assumed that every household has at least one bicycle available.

Additionally, car availability is addressed since car unavailability greatly impacts the intrapersonal variability in mode choice behaviour. It was assumed that a car was available for a respondent when the individual made at least one trip as a car driver during the data collection period. This assumption resulted in a car available for 89% of the respondents. OViN 2012 does ask respondents what modes are available to them and analysis showed that 89% of the OViN respondents own a car in their household. Since the shares of car availability are the same for both the DMMP dataset as well as OViN, the assumption made regarding car availability is considered valid.

The verified trips from the first data collection wave did not contain trip characteristics for the alternative travel modes. This data can be extracted from Google Maps and added to the data as shown by Van der Kruijs (2013). Unfortunately, this was not possible since the data was stored at CentERdata.

An alternative method to add these characteristics was to create a database containing these characteristics for trips between all postcodes in the Netherlands. The data can then be added to the trip based on the registered origin and destination postcodes. The Netherlands has 4688 different postcodes on a postcode-4 level. A matrix of these postcodes would contain 22 million cells for every characteristic. Google allows individuals 2.500 data requests and businesses 100.000 data requests per day. Filling this matrix would take too long and therefore the choice was made not to add the characteristics of alternative travel modes.

During the second data collection wave the travel times for the car, bicycle and public transport were registered real-time during the data collection period and available in the dataset.

3.3.5. Inserted distances

The registered trips did not contain the trip distance, while the trip GPS location of the trip origin and destination were logged. In the theoretical framework it was pointed out that trip distance is a determinant of modal choice. The trip distances needed to be inserted in the dataset in order to include trip distance as a characteristic in the analysis.

A matrix containing distances between postcodes on a postcode-4 level was made available for this research in order to add the trip distances. This matrix was computed by a national traffic model from the Netherlands. Manual verification of trip distances compared to distances in Google Maps showed an overestimation of trip distances for trips with both the origin and destination within the same postcode-4 area. The difference between Google Maps and the inserted trip distances decreased as the trip distance increased. For short trips the inserted trip length is unreliable, but for trips where the origin postcode is not the destination postcode, the inserted distance showed to be quite reliable.

3.4. DATA SELECTION

Previous sections described the data cleaning, noise reduction and data enrichment processes used. The resulting trip database contains marked trips and unreliable respondents. This section describes the data selection process.

3.4.1. RESPONDENT SELECTION

Respondents were selected in two ways, through the day questions and through the registered trips. First, selection based on day-questions is described followed by the selection based on the trip data.

The respondents filled in a prompted recall survey (called day-questions) during the data collection period. This survey asked if special circumstances occurred that day and if so, how these circumstances affected the travel behaviour. From the 593 respondents in the trip database only 513 respondents filled in the prompted recall survey. Four respondents did fill in the prompted recall survey but did not register any trips and were therefore not selected.

Not all surveys were filled in correctly. Erroneous responses were identified and marked based on the following criteria:

- Double data entries;
- For some days respondents stated that special circumstances occurred that day. However, they did not state how these circumstances influenced their travel behaviour;
- Some respondents stated that for one day both ‘no special circumstances’ and one (or more) of the predefined special circumstances occurred. Based on this it is not possible to identify a trip as either ‘with special circumstances’ or ‘without special circumstances’.

Respondents completing the prompted recall survey for 9 or more days correctly were selected. This number was chosen to ensure that only respondents with sufficient data were selected.

Combining the respondent-ID, day and month from the selected respondents resulted in a day-ID for every day the selected respondent answered the prompted recall survey. In the trip database the trips without a registered recall survey that same day were marked in order to make it easy to exclude these trips from the analysis.

Additional selection was performed based on the registered trips following the respondent selection based on the day-questions. All trips were checked manually to see whether or not destinations of trip X matched the origins of trips X+1. When this is not the case it means that either the registration of the MoveSmarter was wrong or that a trip is missing. Respondents were considered as unreliable when the trips of a respondent consequently showed a mismatch between the destination of trip X and origin of trip X+1. This process was performed manually since the entry of origins and destinations by the respondents resulted in a many variations for the same origins and/or destinations and automation was therefore not possible. An automation attempt was done by using fuzzy logic[1], but the attempt failed. Small variations in either the origin or destination resulted in a large mismatch, which meant that the origin of trip X+1 and destination of trip X were different while actually these were the same. Using this method would result in reliable respondents being considered as unreliable.

[1] Fuzzy logic is a form within logic. It deals with approximate reasoning rather than fixed and exact reasoning. A percentage match between the destination of trip X and origin of trip X+1 was calculated using a fuzzy logic plug-in in Excel. Analysis of the results showed that 1 mistype could lead to a low match percentage. This was often lower compared to trips where the destination was completely different from the next departure. Because of this the process was classified as a failure and not used any further.

The trip data was enriched based on the postcode registered with the trip. Six respondents had no registered trips with a postcode and therefore these respondents were deleted from further analysis. Also two respondents did not fill in the prompted research survey for any day. These respondents were also deleted from further analysis.

The abovementioned selection processes resulted in 400 selected respondents with 19702 registered trips (including marked trips from section 3.2). 3025 of these trips contained a marking as to contain an error and 16677 trips were unmarked.

The choice was made to look at the share of erroneous trips of individuals instead of the number of erroneous trips to delete respondents providing unreliable data. Respondents with a few registered trips and a high share of errors would otherwise still be selected because the number of errors would be below the deletion threshold. Figure 11 shows the distribution of the shares of erroneous trips among the respondents. It can be seen that by removing twenty percent of the respondents, over seventy percent of the erroneous trips were removed as well. This threshold was chosen to not remove too much data. Also, removing more individuals would not result in the removal of many additional erroneous trips. Removing 20 percent resulted in a dataset containing the trips of 321 individuals.

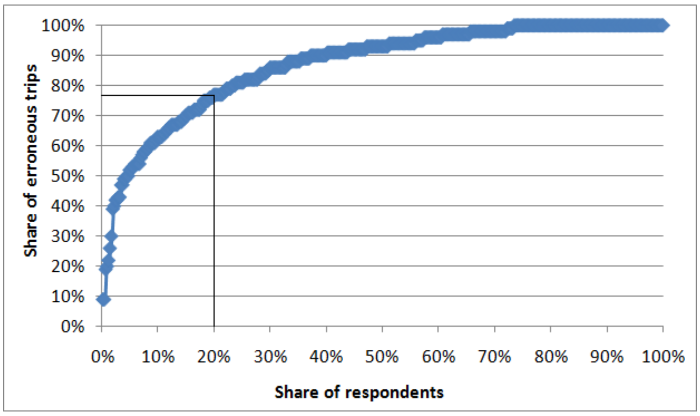

FIGURE 11: SHARE OF ERRONEOUS TRIPS VERSUS THE SHARE OF RESPONDENTS

3.5. ADDING DATA FROM WAVE 2

Section 4.1 described the data collection process as performed by Mobidot and CentERdata for the first data collection wave in 2013. The second wave was collected between April and July 2014 and the data became available in august. Respondents registered their travel behaviour for four consecutive weeks. The choice was made to include this dataset in the set resulting from the process as described by the previous sections in this chapter, since the dataset contained a lot of extra registered trips for the respondents.

The dataset of 2014 contained trips of 681 individuals. Only the trips of the respondents selected in section 3.4 were regarded, since the purpose of adding the new dataset was to increase the amount of data per selected respondent. Unfortunately, not all respondents from the first wave responded in the second wave. 201individuals registered their trips in both data collection periods.

The data processing and enrichment process as described in sections 3.3 and 3.4 were performed similarly on the added data. Only, no time was spent to find oddities manually in the data as registered by the respondents. The same algorithms as used for the first wave were in order to edit and complement the list of postcodes. As a result of not spending much time to include the second wave of data, the number of missing postcodes was higher for the second wave with 8.2% of the trips without a postcode present.

Trips in this second wave also contained the travel time in seconds for the car, the bicycle and public transport additional to the data characteristics also collected in the first wave. From all selected trips from the second wave, 4610 trips contained the travel times for these three modes.

3.6. LIMITATIONS OF THE RESULTING DATASET

Data collected in the DMMP-project is based on individual level data instead of household data. Mode choice and trip purpose are dependent on household characteristics, as argued in the theoretical framework. Reported trips might be made for the sake of the household instead of for the respondent, as well as mode choice depends on activities within the household and household characteristics. For instance: a family with one car does not have the option to make two car trips at the same time. A choice has to be made which person in the household uses the car and who uses another mode since no car is available anymore. This effect cannot be derived from the data and can influence the results of the analysis.

The available data is retrieved from a mobility panel. This panel is ought to be a representative sample of the Dutch population. The travel behaviour of people outside the panel might differ from the observed travel behaviour of the panel members, since only a small sample of the population is in this research. Chapter 4 analyses the representativeness of the data. It analyses both the representativeness of the background variables of the selected respondents as well as their travel behaviour.

The panel data has been collected before the start of this master thesis research, making it impossible to add personal characteristics in case they were missing. In the case of missing characteristics as identified from the literature, an assumption was made or the characteristic was neglected in the analysis; notes can be added to the results to state the possible effect of the missing characteristics.

3.7. CONCLUSION

Section 3.1 described the data collection process used in the DMMP project. Both the trip registration and trip verification were considered. The data collected in the DMMP project contained erroneous trips. Section 3.2 described the used data cleaning and noise reduction processes.

The theoretical framework in Chapter 2 identified characteristics that are not in the dataset. Section 3.3 described the processes used to enrich the dataset. Weather data, trip distance and location type data was added to the registered trips. Figure 12 shows what characteristics are available in the resulting dataset.

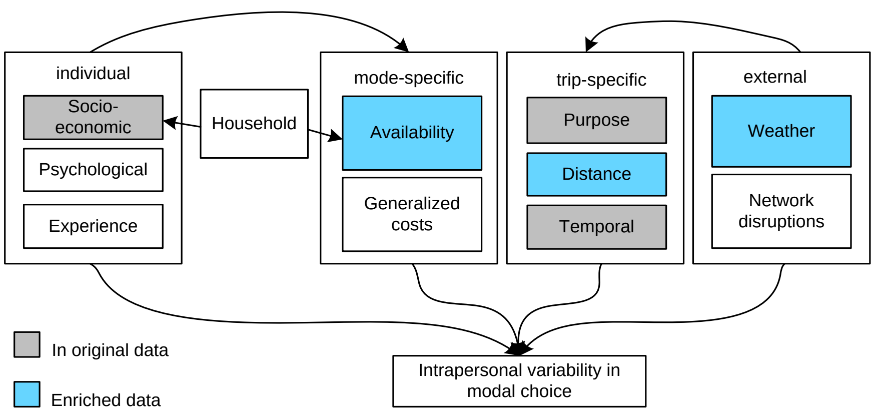

FIGURE 12: CONCEPTUAL MODEL ADAPTED TO DATA AVAILABILITY

Section 3.4 described the data selection process. Not all respondents provided valid data and respondents not fitting the set criteria were not selected for the analysis. The processes used resulted in a total of 321 selected respondents with 15606 registered trips, of which 915 trips consisted at least one of the previously mentioned errors (5.8%). Section 3.5 described the addition of the second data wave. Missing postcodes accounted for 8.2% of erroneous trips for this wave. Finally, some limitations to the resulting dataset were mentioned in section 3.6.

4. Descriptive analysis of the dataset

The previous section described the data collection, enrichment and selection process. This chapter compares the dataset with the LISS-panel and OViN to see how distributions of characteristics in the selected dataset compare to other datasets.

The first section of this chapter describes what OViN is and what data it contains. The second section compares the distributions of the background characteristics with distributions of the LISS-panel. The third section discusses the trip length distribution followed by a comparison of the model split to OViN 2012 in the fourth section. The fifth section compares the trip purposes with OViN 2012. Finally, the trip distance distribution is compared to OViN in the sixth section. Because OViN contains a plurality of the number of trips registered in the DMMP-dataset, comparisons are done in this chapter based on observed shares of characteristics.

4.1. OViN

Section 3.1 mentioned a comparison of the DMMP data collection process to OViN. The following sections in this chapter compare the dataset resulting from chapter 3 to characteristics from OViN. Therefore, OViN is briefly described in this section to create a general image to what OViN is and what data it contains.

The Dutch government is interested in how and when the Dutch population participates in traffic. In order to analyse this, Statistics Netherlands is collecting data of the travel behaviour among a large pool of respondents every year. Respondents are asked to register all their travel movements during one single day. By using a large sample, the respondent burden is low (only one day of registration), but the results are ought to be representative for the entire population. Information of all movements of Dutch inhabitants can be obtained based on this research.

The OViN 2012 dataset is used for comparison in this research. This was the most recent available dataset at the time. The dataset contains three types of characteristics. The first type of characteristics contains the socioeconomic characteristics of the respondents. Characteristics regarding the household composition are among these characteristics.

The second type of characteristics in the OViN dataset describes the availability of transportation modes in the household. It is also known whether or not a respondent is in possession of a driver’s licence or a student public transport card.

Finally, the third group of characteristics are characteristics of the trips made by the respondents during the registration day. Among these characteristics are: travel mode, trip purpose and trip distance. Based on the registration of trips in OViN a distinction can be made between uni-modal trips and multimodal trips in case one trip consists of multiple trip stages.

4.2. Distribution of background characteristics

The socioeconomic characteristics of the respondents involved in the research are described Table 2. As mentioned in section 3.1, the participants in the research were selected from the LISS panel. The table shows that the distribution for the selected respondents resemble the distributions of the entire LISS panel, which is ought to be representative for the Dutch population according to CentERdata. One thing standing out is that fewer respondents of 65 years and older were selected using the processes described in the previous chapter. A possible explanation is that elderly individuals are not used to operate a smartphone and a computer. As a result, the prompted recall questions were not filled in or trips were not registered, which were used selection criteria for respondents in this research. Additionally, the shares of the first and second data wave are comparable to each other. It is observed that the sample fallout has been spread evenly over the population and the 2014 sample is representative for the population as well.

	LISS panel	Participants	Selected participants	Respondents 2014
Number of individuals	6360	655	321	201
Gender				
Male	46%	53%	53%	52%
Female	54%	47%	47%	48%
Age				
15-24	13%	10%	12%	9%
25-34	12%	14%	15%	14%
35-44	16%	19%	22%	22%
45-54	18%	18%	18%	19%
55-64	19%	22%	22%	24%
65 and older	22%	17%	12%	11%
Most important occupation				
Payed job	51%	55%	62%	63%
Looking for work	3%	6%	3%	3%
In school	11%	9%	11%	8%
Housekeeping	8%	4%	4%	3%
Retired	19%	17%	13%	14%
Incapacitated	4%	5%	5%	5%
Unpaid work	2%	2%	1%	1%
Other	1%	1%	2%	1%

TABLE 2: DISTRIBUTION OF BACKGROUND VARIABLES

4.3. TRAVEL TIME DISTRIBUTION

This section and the following sections analyse the distributions of trip characteristics of the DMMP data and compare these with OViN2012. It can be concluded that the resulting DMMP sample is representative, when the distributions are similar for OVIN and the DMMP-panel.

Table 3 shows the travel time distributions for all trips collected in the first wave, the second wave, the entire dataset and OViN. The datasets show more or less the same distributions. Short trips are registered more often in the OViN research compared to the first wave of DMMP-data, but the difference is marginal. Additionally, the second DMMP-wave compares perfectly with the distribution of OViN. Since the travel time distribution is comparable between the sample and OViN, the sample is ought to be representative based on this characteristic.

Travel time distribution (min)	DMMP 2013	DMMP 2014	DMMP 20013 & 2014	OViN 2012
0 - 14	47%	50%	49%	51%
15 - 29	26%	27%	27%	27%
30 - 59	17%	16%	16%	15%
60 - 89	5%	4%	5%	4%
90 - 149	3%	2%	2%	2%
> 150	1%	1%	1%	1%

TABLE 3: TRAVEL TIME DISTRIBUTION

4.4. MODAL SPLIT

This thesis analyses the intrapersonal variability in mode choice behaviour. In order to get valid results the modal split of the selected respondents needs to be representative. This section compares the modal split of the DMMP-dataset to OViN and the modal split observed for the Dutch Mobility Panel project (MPN) by Olde Kalter et al. (2014).

The goal of the MPN is to increase the understanding of the characteristics underlying a change in mobility behaviour of individuals. Multiple waves of data will be collected for this project and the results of the first wave as reported by Olde Kalter et al. (2014) are used for comparison here. Respondents in this project register their trips and activities in a travel diary for three days, where OViN is based on one day of data and the DMMP-data is based on two weeks.

Figure 13 shows the modal splits of the different datasets. Some differences are observed when comparing the modal splits of the samples. The share of car driver trips is 5% higher in 2014 for the DMMP-sample compared to 2013, while a lower share of walking trips are registered. Also, the share of car drivers is overrepresented for both the DMMP datasets and the MPN data when comparing to OViN. Bicycle, walking and car passenger trips are underrepresented compared to OViN.

Based on Figure 13 one could debate the representativeness of OViN regarding the modal split. Two multiday datasets show the same modal splits, which is different compared to OViN. Since the modal split of the DMMP-dataset is comparable to the MPN modal split, the DMMP dataset is considered representative on this characteristic.

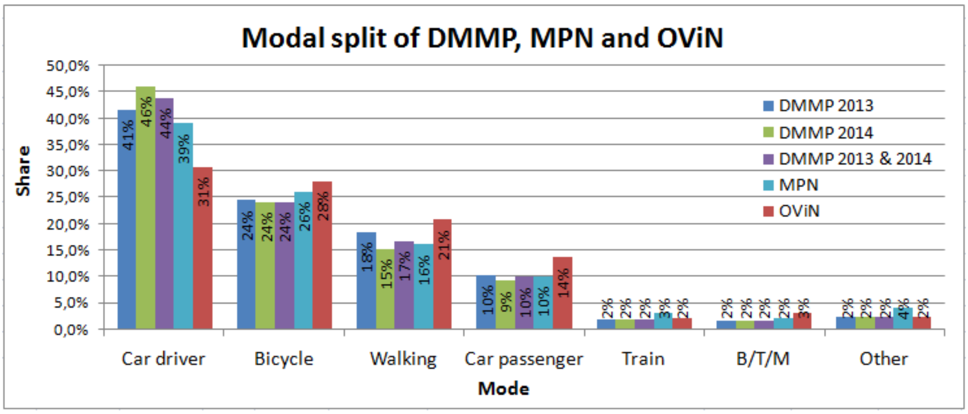

FIGURE 13: MODAL SPLIT FOR THE SELECTED DMMP DATA COMPARED TO OVIN

4.5. Trip purpose analysis

Additional to the distribution of the trip characteristics analysed in the previous two sections, trip purpose is also an important indicator to explain travel behaviour. Table 4 shows the distribution of trips among the different trip purposes for both the DMMP data and OViN. These distributions show large differences for multiple purposes. The difference can be explained by the following: OViN data is collected throughout the entire year while the DMMP only collected data during two periods of two consecutive weeks during the first data collection period, and four consecutive weeks during the second data collection wave. During the first week of the first DMMP data collection period, both the primary and the secondary schools had a holiday. Also the second Sunday of the first period was Mothers day. These circumstances are also registered in OViN, but since OViN registers trips throughout the entire year it corrects itself for these circumstances. DMMP on the other hand, does not contain the large amount of data from other days to correct for this and therefore it can be said that the trip purposes are not representative for the entire year. This explains the differences in the distributions of trip purpose as shown by Table 4.

The second wave is more representative compared to OViN than the first wave. During the second wave, the respondents registered trips for four consecutive weeks instead of two weeks as was done for the first wave. Holidays account for a lower share of days for this sample. This backs up the previous claim that the large share of holiday days caused a difference in trip purpose distribution between the first wave of DMMP-data and OViN.

Trip purpose	DMMP 2013	DMMP 2014	DMMP 2013 &	OViN 2012
To home	30,7%	32,8%	31,9%	39,6%
Leisure	14,2%	10,6%	12,2%	5,6%
Shopping	12,0%	12,1%	12,0%	11,2%
To work	8,9%	10,1%	9,5%	11,0%
Visit	6,1%	5,7%	5,9%	7,0%
Other trip purpose	4,8%	4,6%	4,7%	1,3%
Business trip	4,5%	3,9%	4,2%	1,1%
Bringing away or picking up	4,4%	5,9%	5,3%	4,4%
Sport/hobby	3,6%	3,7%	3,7%	4,4%
Hiking/sightseeing trip	2,9%	2,8%	2,9%	4,4%
Tranfer	2,8%	2,8%	2,8%	--
Going out	2,2%	2,0%	2,1%	--
Personal care	1,8%	2,0%	1,9%	2,3%
Education / seminar	0,8%	0,9%	0,9%	7,6%

TABLE 4: TRIP PURPOSE DISTRIBUTION

4.6. TRIP DISTANCE

Another distribution to analyse is the distribution of the trip distance. The theoretical framework stated that trip distance is an influential characteristic in mode choice behaviour. To get valid results from the analysis it is important that the trip distance distribution is comparable to OViN. Figure 14 shows the distributions for the first wave, the second wave, all registered trips for both waves and OViN in one figure.

The figure shows that for trips between zero and one kilometre, the share of trips is higher for OViN compared to the DMMP share. This has already been explained in section 3.3.5. The values between one and four kilometres show a correction for this effect and the difference for the first five kilometres is only three percent (respectively 57% of the DMMP trips and 60% of the OViN trips).

Additionally, OViN shows peaks at a five kilometre interval. These peaks are not observed for DMMP and are caused by the different methods of data collection. In OViN, a travel diary is filled registered in which respondents have the tendency to round the travelled amount of kilometres off. This is not the case for DMMP where the distances were inserted based on the origin and destination postcodes as described in section 3.3.5.

Taking both the differences on the shortest distance and the small peaks at a five kilometre interval into account it is to say that both distributions are quite comparable to each other and it can be concluded that the dataset resulting from the data selection and processing phase is representative considering trip distance.

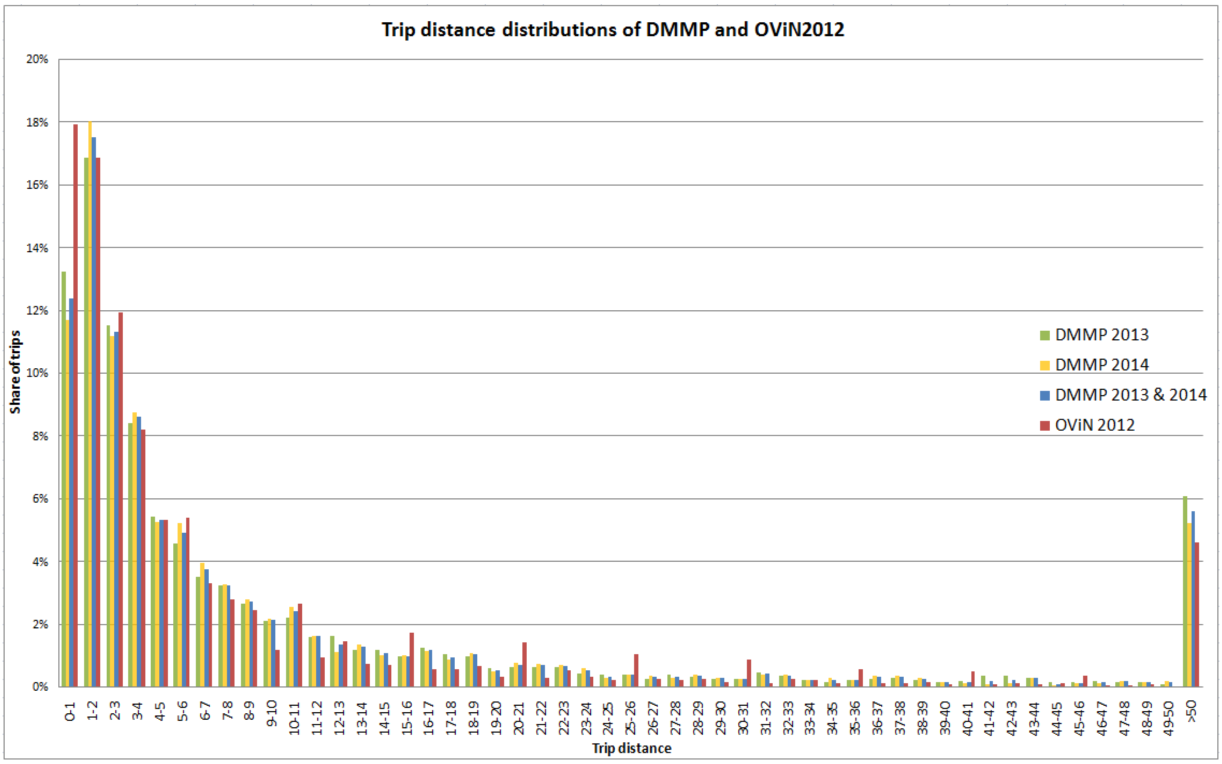

FIGURE 14: COMPARISON OF TRIP DISTANCE DISTRIBUTION BETWEEN OVIN 2012 AND THE DMMP DATASET

4.7. Conclusion

This section showed that for the background characteristics, travel time, mode use and trip distance the DMMP sample is representative for the entire population based on a comparison with the LISS panel for socioeconomic characteristics and with OViN 2012 for trip characteristics. This is not the case for trip purpose, but differences between OViN and the DMMP dataset were explained in section 4.5. Based on the analysis described in this section it can be concluded that the data sample resulting from the data selection process is a representative sample.

5. Analysis

Previous chapters described the data collection and transformation process followed by an analysis of the representativeness of the dataset. This chapter uses the resulting dataset and elaborates on the analysis phase of this research. The first section analyses the correlation between several available characteristics and the modal split. The second section briefly describes the dominant mode use analysis as performed for this research and this chapter finishes with an intrapersonal mode use variability analysis in the third section.

5.1. Correlation analysis

The dataset resulting from the processes described in the previous chapters contains a large amount of characteristics. An overview of the available characteristics is presented in Appendix C. Distributions of characteristics expecting to influence the modal split are showed and analysed in Appendix D. Graphs display the effect of the researched characteristics on the modal split. The Pearson's chi-square test is used to test whether there is a significant relationship between the modal split and these available characteristics. This test compares observed frequencies in certain categories to frequencies expected by chance. The value of χ^2 is calculated using the following formula (Field, 2009):

$$\chi^2 = \sum \frac{\left(observed_{ij} - model_{ij}\right)^2}{model_{ij}}$$

In this formula the $observed_{ij}$ are the frequencies calculated in tables containing the observed values for the DMMP-dataset. The expected frequencies are calculated based on these tables with observed frequencies, using the following formula:

$$model_{ij} = E_{ij} = \frac{row\ total_i \times column\ total_j}{n}$$

Where n is the total number of observations. The obtained χ^2 value can be checked against a distribution with known properties. Before this is done the degrees of freedom are calculated using (r-1)(c-1) where r is the number of rows and c is the number of columns. When the value of χ^2 is larger than the critical value one can say that there is a significant relationship between the tested characteristic and the modal split.

Table 5 shows the resulting χ^2 values for characteristics influencing the model split and which were included in the DMMP dataset. It is observed that all but one characteristic have a significant relationship with respect to the modal split. Hourly precipitation tested not significant because of the intervals chosen to calculate the observed frequencies. The chi-square test is sensitive to small expected frequencies in one or more of the cells in the table and there were a small amount of observations with a large hourly precipitation in table containing the observed frequencies. Summing all observations above 50mm/hour yielded in a significant relationship for the hourly precipitation characteristic ($\chi^2(36)=76.98$, $p < .05$).

Characteristic	Degrees of freedom	χ2	Critical Value	Significant? (P<.05)
Gender	6	1072	12,6	yes
Age	30	1642	43,8	yes
Housing type	24	1033	36,4	yes
Education level	30	1633	43,8	yes
Income class	84	2596	106,4	yes
Household income	66	2204,2	86	yes
Number of children	36	893,5	51	yes
Phone type	18	566,5	28,9	yes
Home based trips	6	111,9	12,6	yes
Location type trip origin	18	1224,9	28,9	yes
Degree of urbanization residence location	24	1895,3	36,4	yes
Trip distance	100	19217,3	124,3	yes
Departure time	138	1156,7	166,4	yes
Type of day	6	438,2	12,6	yes
Day of week	36	613,6	51	yes
Trip purpose	84	11934,6	106,4	yes
Amount of sunshine trip hour	66	153	86	yes
Temperature	96	276,8	119,9	yes
Sky coverage	54	168,4	72,2	yes
Rain	6	50,5	12,6	yes
Hourly precipitation	126	135,2	153,2	no
Hourly wind speed	132	238,9	159,8	yes

TABLE 5: RESULTS CORRELATION ANALYSIS

5.2. DOMINANT MODE USE

The theoretical framework mentioned the use of the methods of Buehler & Hamre (2013) and Kuhnimhof (2009) to express intrapersonal variability in mode choice behaviour. Another method was used earlier on in the research which was formulated based on knowledge instead of literature. This section briefly describes the third method used.

The third measure of intrapersonal variability in mode choice behaviour is based on the use of the dominant mode used. The dominant mode is the mode accounting for the largest share of trips, determined for every individual. This is a measure for variability in the following way: The higher the share of trips with the dominant mode, the lower the share of trips performed by other means of transport. Therefore, it is derived that a high share of trips for the dominant mode used means a low intrapersonal variability in the modal choice of an individual. This analysis was performed on both the trip frequency and the trip lengths. The analysis showed that the private car is the dominant mode for most of the respondents. When considering the trip distances the private car showed even higher dominance values. This analysis is further described in Appendix E.

5.3. INTRAPERSONAL MODE USE VARIABILITY

The importance to measure mode use variability on a personal level has been stated earlier on in this thesis. Since mode use is based on frequencies, standard deviation is not a suiting measure for this. Two quantification methods were introduced in the theoretical framework, namely the method by Buehler & Hamre (2013) and Kuhnimhof (2009). The method of Buehler and Hamre only provides shares of respondents in certain specified categories. Based on collected data individuals were either monomodal or multimodal. Using this method in the Dutch context would result in all individuals being multimodal, since the Dutch are prone to use multiple modes of transportation. Using this method would therefore not provide insight in the intrapersonal variability in mode choice behaviour in the Dutch context. Instead, the method of Kuhnimhof, who expressed the variability as a relative deviation from a state of maximum variability, is used. This method is capable to handle different time intervals of collected data. This section explores the mode use variability of the individuals in the DMMP-dataset using this method.

The analysis uses the same mode classification as Kuhnimhof (2009); car driver, car passenger, bicycle, walking and public transport & other. As stated in chapter 3, 201 respondents registered trips for both data collection waves. This analysis calculates the MIX-values for these individuals. The cumulative distribution functions of the MIX-value are constructed based on these values. Adding extra data, the effects of the trip purpose and the trip length on the MIX-distribution are analysed in this section, as well as the effect of rain and a range of available socioeconomic characteristics. Significance is tested using two sample Kolmogorov-Smirnov tests (K-S test). This test is explained in Appendix F.

5.3.1. DATA QUANTITY AND COMPARISON TO THE GERMAN MOBILITY PANEL

This section first analyses the effect of data quantity on the MIX-distribution. Based on the resulting MIX-distributions the intrapersonal variability in mode use of the DMMP-panel is compared with results from the German Mobility Panel as described in the theoretical framework.

To test the effect of data quantity on the MIX-distribution, MIX-distributions were plotted in Figure 15 for multiple data collection intervals; one week (first data collection week and the fifth data collection week), two weeks, four weeks and all six weeks of collected data. The MIX-distribution for one week of data collection was plotted twice, since the first data collection week in 2013 was a holiday week which could affect the results. Since the fifth week was a normal working week this week was included for comparison reasons.

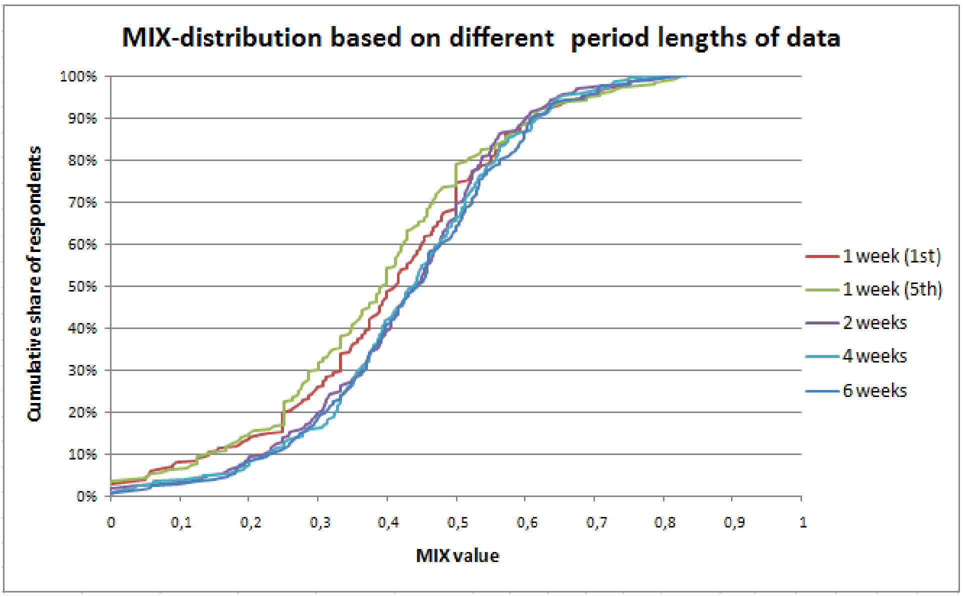

FIGURE 15: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, BASED ON DIFFERENT DATA COLLECTION INTERVALS

Figure 15 shows that there is hardly any difference between the MIX-distributions for two, four and six weeks of data (two weeks: Ø=0.44; four weeks: Ø=0.44; six weeks: Ø=0.44). A significant difference is observed between the MIX-distributions of data collected during the fifth week, and four weeks of data collection (K-S test: D=0.136, p<.05). While the MIX-distribution for only the first week is different as well, the difference is not significant.

While the MIX-distributions are the same for two, four and six weeks of data collection, this does not mean need to be the case for every individual. Since the MIX-distributions are cumulative distribution functions, individuals can show different levels of intrapersonal variability in mode choice behaviour when extra data is added, while the overall MIX-distribution is the same. Analysis showed small differences in the MIX-values of individuals when different data collection time spans were considered. Appendix G further elaborates on this issue.

From the above it is clear that two weeks of data collection is a good time span to capture intrapersonal variability in mode use by the respondents, since using more data did not yield other results. Additionally, the trip rate can also have an effect on the intrapersonal variability values. A higher number of registered trips gives a higher probability to observe other-than normal mode use, and thus results in an increase of the MIX-value of that individual. To test this effect, the group of respondents was split based on the number of registered trips at the median trip rate. The resulting MIX-distributions are shown in Figure 16.

As expected, the group of individuals with a high number of registered trips shows a higher intrapersonal variability in mode use compared to individuals with fewer registered trips (below median: Ø=0.42; above median: Ø=0.46). However, the observed difference is not significant (K-S test: D=0.19, p<.05).

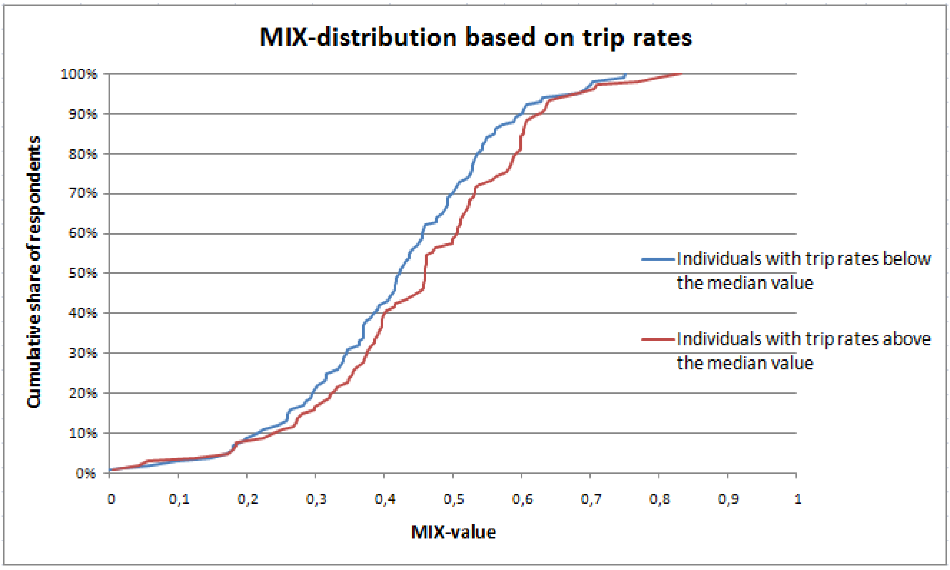

FIGURE 16: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, BASED ON TRIP PRODUCTION LEVELS

Figure 15 showed that the median value is 0.44 when two or more weeks of data are collected for the DMMP-panel. These distributions can be compared to the intrapersonal variability in mode use in the German context, as measured by Kuhnimhof (2009). The data used in the German case is the German Mobility Panel data. In this panel respondents report their travel behaviour seven days for three consecutive years. As a result three weeks of data is available for every respondent.

The median value in the German context is 0.34 (see Figure 9) while this was 0.44 in the Dutch context. Also, in the German case ten percent of the respondents showed no intrapersonal variability in mode use, while this is not the case in the Dutch context. One can say based on these values that the Dutch intrapersonal variability in mode use is higher compared to the Germans. This difference could either be the result of the method of data collection (more short trips registered because of smartphone registration in the DMMP-panel resulting in a different modal split) and the fact that the Dutch actually have more intrapersonal variability in mode use compared to the German population.

5.3.2. TRIP PURPOSE

This subsection analyses the effect of the trip purpose on the MIX-value distributions. As Table 4 showed, many trip purposes were used in the dataset. Since using this trip purpose classification would result in low trip rates per trip purpose, the trip purposes are reclassified. The following three categories are used:

1. Work & business
2. Shopping
3. Leisure

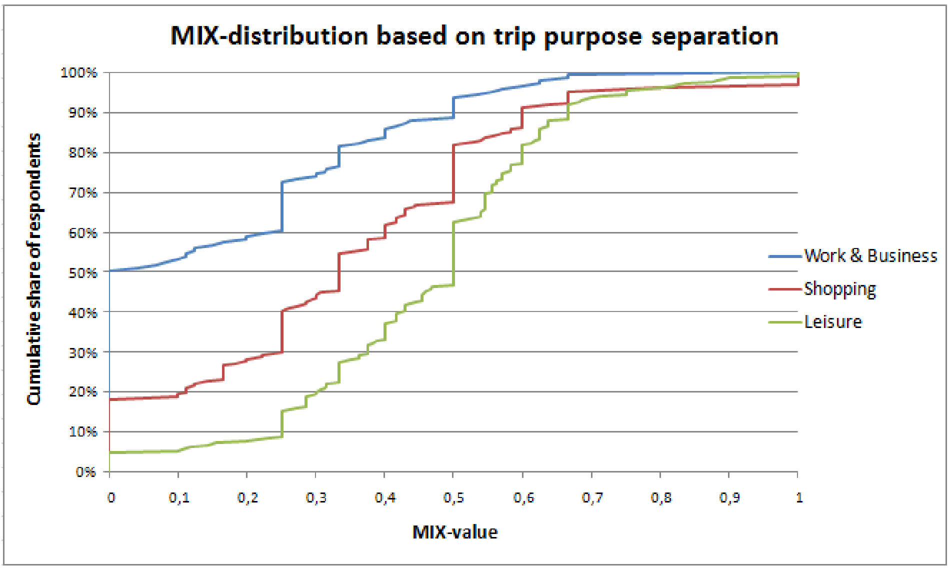

FIGURE 17: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, SEPARATION ON TRIP PURPOSE

Figure 17 shows the resulting MIX-distributions where significant differences between all three distributions are observed. The distributions are ordered as expected: the intrapersonal variability in mode choice is the largest for leisure trips and the least for work & business trips.

The figure shows that 50% of the individuals used only one mode alternative for their registered work & business trips (Ø=0). From these individuals, 75% only used the car as driver for work & business trips. These results are as expected: the car is predominantly used for work & business trips and the intrapersonal variability in mode choice is low.

5.3.3. EFFECT OF TRIP DISTANCE AND TRIP DURATION ON THE MIX-DISTRIBUTION

The effect of trip distance and trips duration on the intrapersonal variability in mode choice are analysed in this subsection. Figure 18 shows the MIX-distributions based on the following three distance classes:

1. Trips below 5 kilometres
2. Trips between 5 and 15 kilometres
3. Trips over 15 kilometres

This classification was used since walking is a feasible mode to five kilometres and cycling is a feasible mode to fifteen kilometres. The car and public transport are mainly used for trips over fifteen kilometres. Since the most mode alternatives are feasible for short distance trips it is expected that the intrapersonal variability in mode choice is the largest for short distance trips.

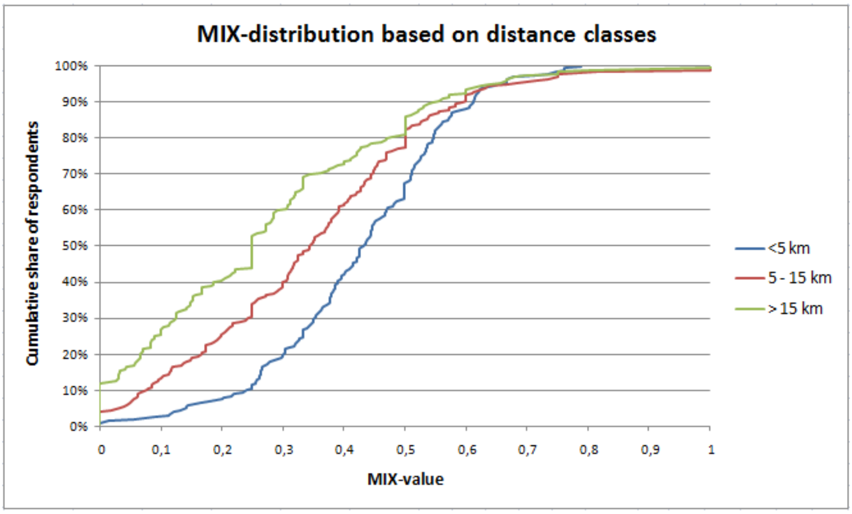

FIGURE 18: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, SEPARATION BY DISTANCE CLASSES

Figure 18 shows significant differences between the MIX-value distributions (K-S test: D=0.136; p<.05 for all comparisons). The shorter trips are, the higher the intrapersonal variability in mode choice of the respondents (<5 km: Ø=0.43; 5 to 15 km: Ø=0.35; >15 km: Ø=0.25).

The effect of trip duration on the MIX-distributions is shown in Figure 19. The trips of the respondents were split the following three trip duration classes:

1. 0 to 20 minutes
2. 20 to 60 minutes
3. Trips over 60 minutes

This classification was used to distinguish trips with a short, medium and long duration. The figure shows significant differences between trips with a short duration and trips with medium or long duration (K-S test: D=0.136; p<.05). A difference is distinguished between medium and long trips as well. However, this difference is not significant. Additionally, the figure shows no intrapersonal variability in mode use for ten percent of the individuals for trips with trip duration over 60 minutes, which means that only one mode is used for these trips and no intrapersonal variability in mode choice behaviour is observed.

The MIX-distributions are ordered unexpectedly. Trips with a short distance are expected to have short trip durations as well. Therefore, the order of the MIX-distributions is expected to be the same as in Figure 18. However, according to Figure 19 the intrapersonal variability in mode use is lower for trips with a short trip duration compared to the MIX-distributions for longer trip durations.

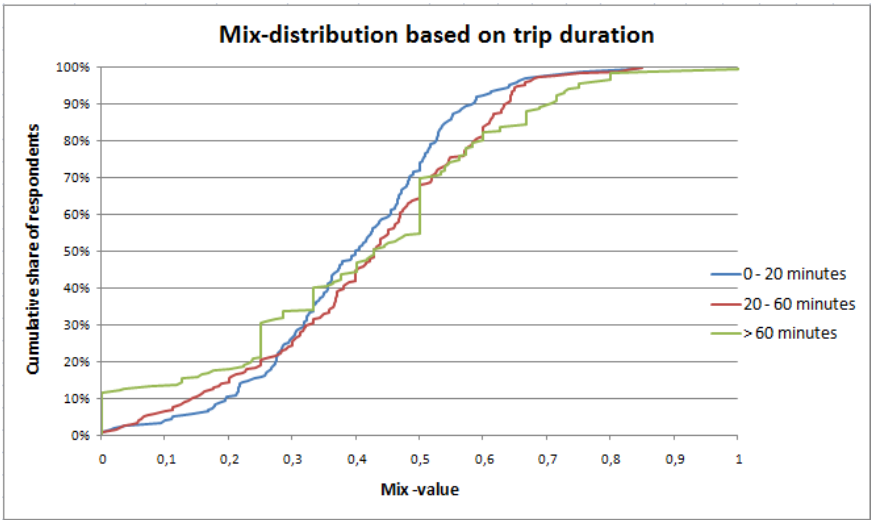

FIGURE 19: CUMULATIVE DISTRIBUTION FUNCTION OF MIX, SAPARATION BY TRIP DURATION

5.3.4. Effect of rain on the MIX-value distribution

The MIX-value was calculated for both rain and non-rain trips for all individuals in the sample to analyse the effect of rain on the intrapersonal variability in mode choice. Figure 20 shows the resulting MIX-value distributions, where the intrapersonal variability in mode use is higher for trips without rain (rain: Ø=0.44; no rain: Ø=0.39 for no rain). The effect of rain on the MIX-value distributions is significant (K-S test: D=0,136, p<.05). Based on the figure it can be concluded that the intrapersonal variability in mode choice decreases when it starts to rain.

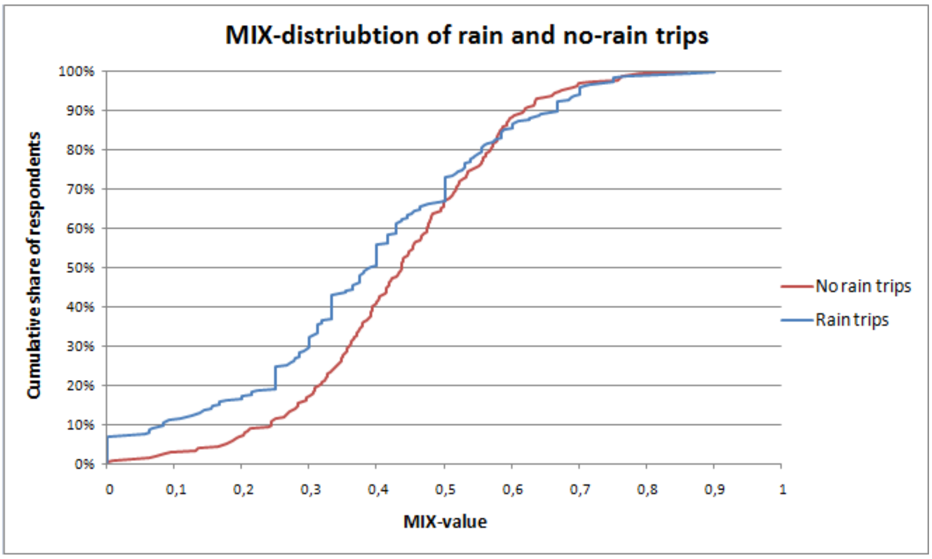

FIGURE 20: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON RAND AND NO-RAIN TRIPS

5.3.5. MIX-value distribution for socioeconomic characteristics

The previous subsections analysed the MIX-value distributions based on trip characteristics. This subsection analyses the effect of various socioeconomic characteristics on the MIX- distribution. Additional to the MIX-distributions, the relative car shares of the respondents are calculated. Calculation of the relative share of car trips provides insight in the difference in car use between the respondent categories. Three bins were created based on an individuals' MIX value to create representative group sizes. The same bin values were used throughout the entire subsection. The following bins were created:

1. MIX-values of 0 to 0,4
2. MIX-values of 0,4 to 0,55
3. MIX-values of 0,55 to 1

For these bins, both the share of car trips in that bin and the share of individuals in that bin were calculated. The relative car shares are calculated for each bin using the following equation: $RCS = \frac{share\ of\ car\ trips\ in\ bin}{share\ of\ individuals\ in\ bin}$

Urbanisation rate

The MIX-value distribution based on the urbanisation rates is shown in Figure 21. The following urbanization rates were used for this analysis:

1. Urban - Area address density over 2500 per km^2
2. Slightly urban - Area address density between 1000 and 2500 per km^2
3. Rural - Area address density below 1000 per km^2

One would expect that in urban areas more mode alternative are available and therefore compete with each other, as discussed in section 2.2.4. As a result, the intrapersonal variability is expected to be higher when the urban density increases. Looking at the figure, the MIX-value distributions are ordered as expected with the highest urbanization rate showing the highest intrapersonal variability in mode use (Urban: Ø=0.48; slightly urban: Ø=0.46; rural: Ø=0.42). However, the differences between the distributions are not significant.

The relative car shares show large differences. These are ordered as expected as well, with a low car share for urban areas. Since a low competition exists with the private car in rural areas, the car shares are relatively the highest for these regions.

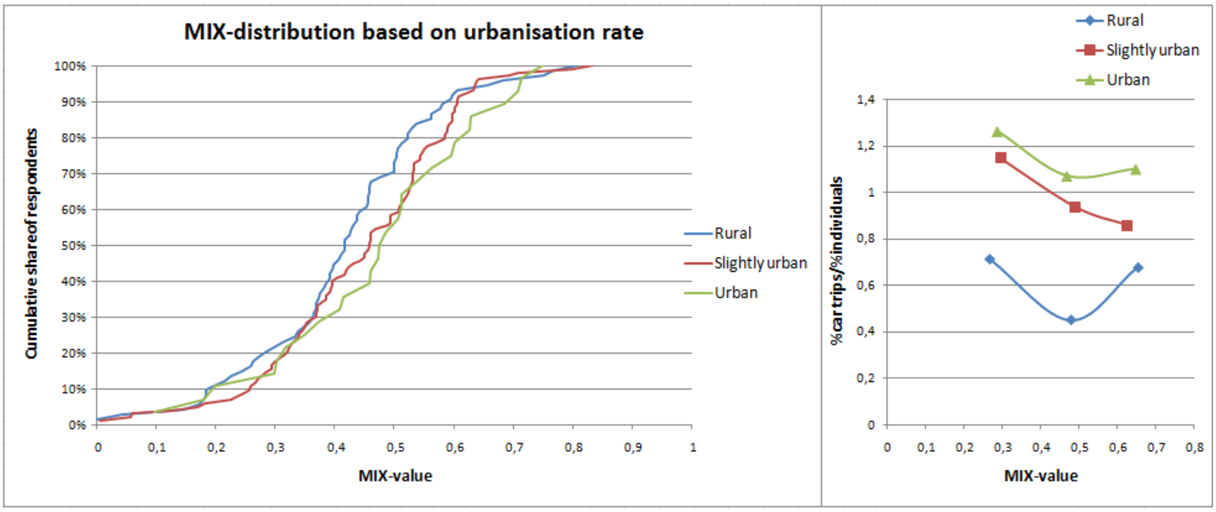

FIGURE 21: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON URBANISATION RATE

GENDER

Figure 40 observed a difference in the number of modes used between both genders. Since women used more modes than men, it is expected that the intrapersonal variability in mode use is higher for women, compared to men. Figure 22 shows the MIX-value distributions for males and females. As expected, the intrapersonal variability in mode use is higher for women, compared to men (male: Ø=0.41; female: Ø=0.48). The difference between the two distribution function is significant (K-S test: D=0.192, p<.05).

Additionally, the relative car shares are very similar with the male using the car a bit more often. Since males are associated with a higher car use compared to women this result is as expected.

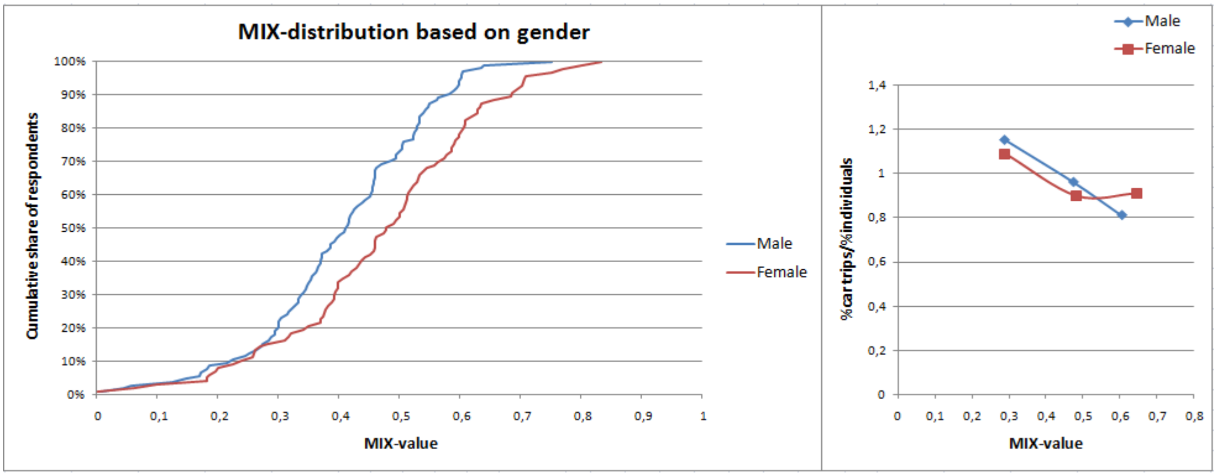

FIGURE 22: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON GENDER

PARTNER STATUS AND CHILDREN IN THE HOUSEHOLD

This subsection first describes the effect of living together with a partner on the MIX-distribution, followed by analysis of the effect children in the household have on the intrapersonal variability in mode choice behaviour.

Figure 23 shows the MIX-distributions for individuals living together with a partner and individuals living alone. Both distributions are quite similar (With partner: Ø=0.43; without partner: Ø=0.47) and differences are insignificant. Additionally, the relative share of car trips is far lower for individuals living without a partner in the household compared to individuals living with a partner. One explanation for this is that couples use the private car more often when travelling to activities together. This in general cheaper compared to public transport use.

Looking to the MIX-values at the individual level, the sample contained seven individuals with a different partner status in 2013 compared to 2014. The MIX-values of these individuals changed over time and were larger when the individuals were not living with a partner. A paired samples t-test was conducted to compare the MIX value for the situation with and without a partner. There was not a significant difference between the MIX-value scores for no partner (M=0.42, SD=0.18) and partner (M=0.34, SD=0.23) conditions; t(6)=1.55, p=0.05.

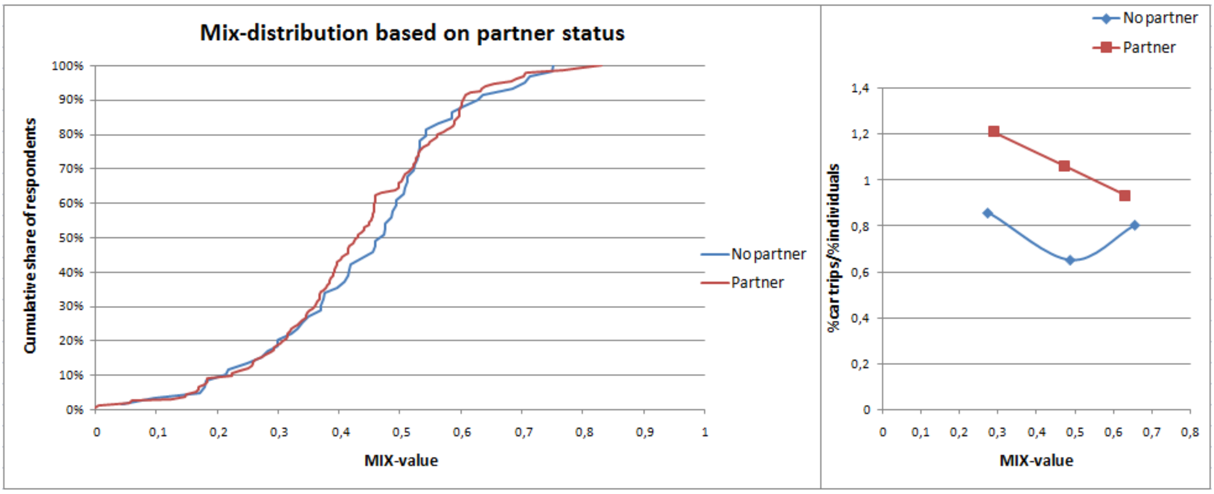

FIGURE 23: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON PARTNER STATUS

Having children influences the behavioural patterns of their parents, also the mobility pattern. Children are mostly dropped off and picked up by their parents from either school, from playing with their friends or from their sports trainings. Figure 24 shows the effect of children in the household on the MIX-distributions. No distinction was made in numbers of children in the household. Individuals either have children or they don't have children. One would expect that the intrapersonal variability in mode use of individuals with children is lower, because the private car is often used by individuals to transport their children. This expectation is met, but the differences between both distributions are insignificant (no child: Ø=0.46; with child: Ø=0.43).

The expectation stated earlier would also result in a higher car use for individuals with children in the household. This expectation is met in Figure 24, where the relative car shares are much higher for individuals with children in the household compared to individuals without a child in the household.

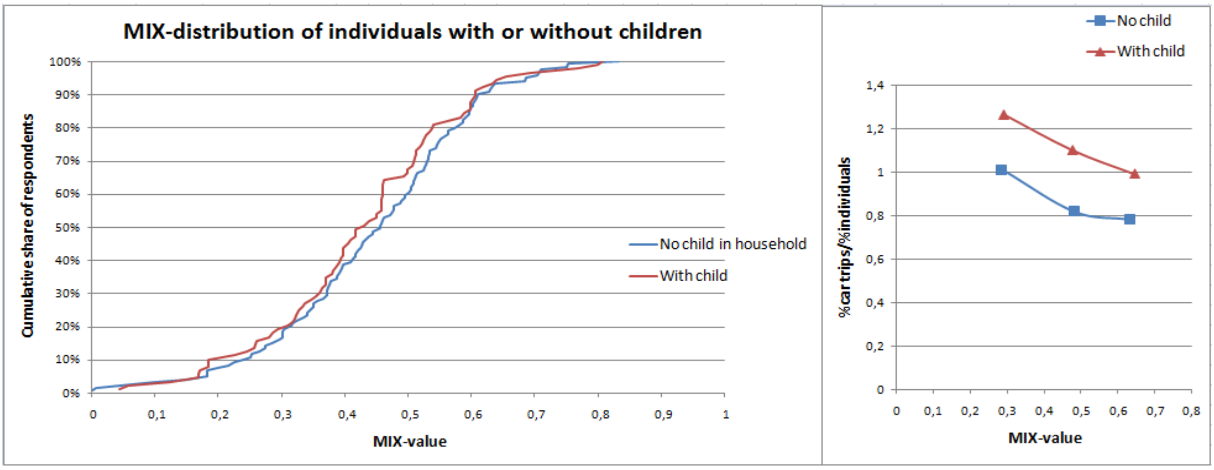

FIGURE 24: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON CHILDREN/NO CHILDREN IN THE HOUSEHOLD

AGE

Earlier analysis (see Figure 41) showed the effect of age on the modal split. This subsection analyses the effect of age on the intrapersonal variability in mode use. The following three age categories were used to create bins containing a sufficient amount of individuals.

1. Below 35 years old
2. Between 35 and 55 years old
3. Over 55 years old

Figure 25 shows the MIX-distributions based on the age categories. While the distributions show some differences, these are not significant. The median values are fairly close to each other as well (Ø1=0.47; Ø2=0.45; Ø3=0.43). Elder individuals show the largest intrapersonal variability in mode use. This is expected since this category contains retired individuals. When individuals retire, they change their mobility pattern. No work & business trips are made anymore resulting in a decrease of the car share, as observed in Figure 41. Because of this, a larger intrapersonal variability in mode use would be expected with a lower car share, which is not observed. This can be caused by the boundaries of the bins. Since the number of retired individuals in the sample was not sufficient for a decent MIX-distribution analysis, the boundary was lowered. Now, working individuals between the ages of 55 and 65 are included in the bin which might cancel the retirement effect out.

When the relative car shares are considered, the younger and older individuals show the same values. The middle age category uses the car relatively less than the other categories but the car share for the highest MIX-values is a lot higher. This is probably caused by the fact that these individuals work and use the car for work & business trips. The younger category contains students and the older category contains retired individuals, who use the car less often.

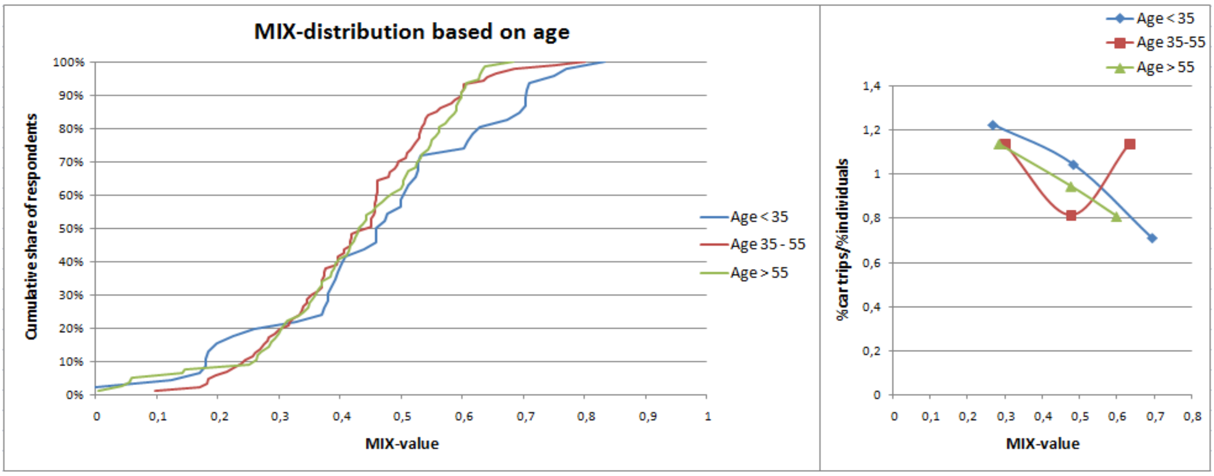

FIGURE 25: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON AGE

EDUCATION LEVEL AND INCOME

This subsection first describes the effect of the education level of the respondents on the MIX-distribution, followed by analysis of the effect of income on the MIX-distribution.

To analyse the effect of the education level of the respondents on the intrapersonal variability in mode choice behaviour, the individuals are split in a higher educated and poorly educated category. The following classification was used:

1. Higher educated – Individuals with a Bachelor or Master degree
2. Poorly educated – Individuals with a degree rated below Bachelor

Figure 26 shows the resulting MIX-distributions. The image shows small differences between both distributions when the MIX-value is below 0.4. The MIX-distributions are the same for higher MIX-values (higher educated: Ø=0.45; poorly educated: Ø=0.44). Additionally, the relative car shares for lower educated individuals are higher compared to the values for individuals with a higher education level. This means that higher educated individuals choose the car relatively less often compared to poorly educated individuals.

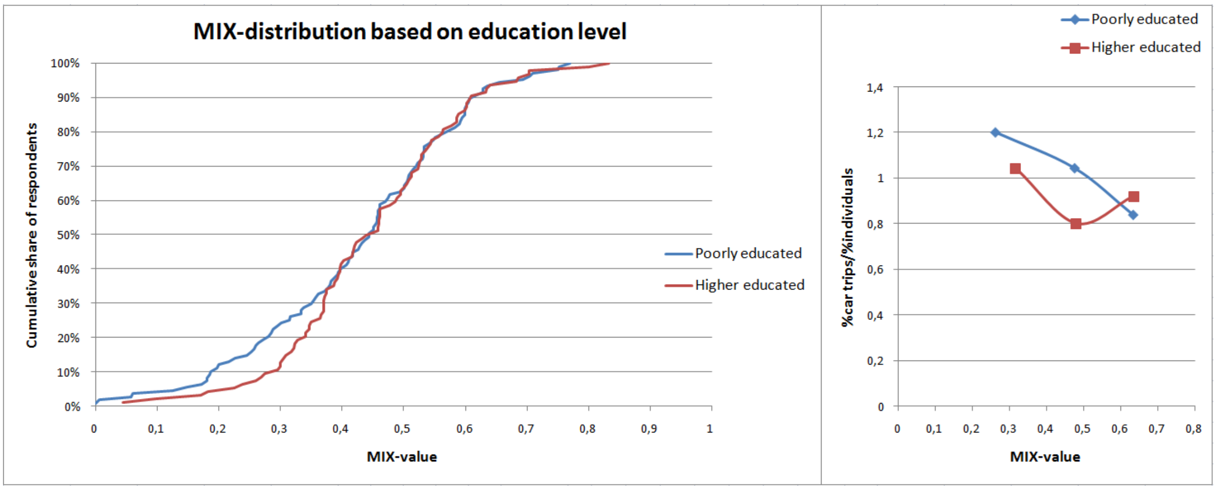

FIGURE 26: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON EDUCATION LEVEL

The effect of income on the MIX-distribution is analysed in Figure 27. The individuals were either classified as having an above modal income or a below modal income. It is observed that individuals with a higher income show higher intrapersonal variability values, but the difference is insignificant (Lower income: Ø=0.44; Higher income: Ø=0.44).

Additionally, the relative car share for individuals with a higher income is higher compared to individuals with a lower income. The effect shown in both figures can lead to the conclusion that individuals with a lower income have a lower mobility budget, causing them to choose more often for other modes than the private car to save costs, thus increasing the variability of this group.

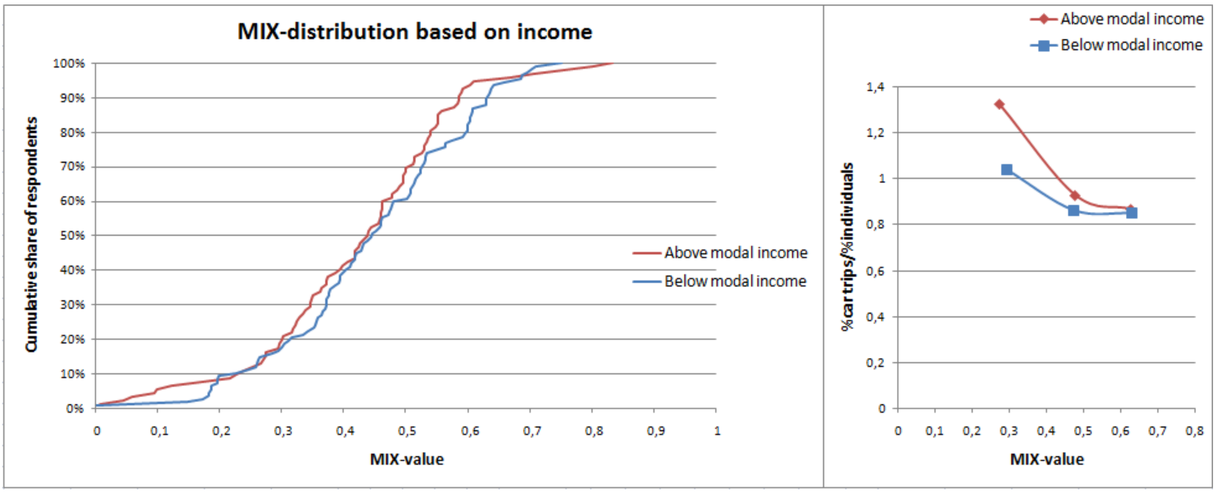

FIGURE 27: CUMULATIVE DISTRIBUTION FUNCTION OF MIX BASED ON INCOME

When both the education level and the income level are considered, one would expect to see that individuals with a lower income would show the same behaviour as poorly educated individuals. This is expected since a lower education is associated with a lower income. However, this effect is not shown by the MIX-distributions of both characteristics since only small differences were observed. When regarding the relative car shares the function of the higher educated car shares matches the function of the below modal income car shares. This is opposite to the expectation.

5.3.6. CONCLUSION INTRAPERSONAL MODE USE VARIABILITY ANALYSIS

This section analysed intrapersonal variability in mode use based on the method of Kuhnimhof (2009). Table 6 shows the median values of the MIX-distributions for all characteristics analysed in this section.

The analysis showed significant effects on the MIX-distributions when trip characteristics were considered. Scheiner (2010) showed the effect of trip distance on the modal split. Since less mode alternatives are feasible for trips with a higher distance, a lower intrapersonal variability is expected, which was observed as well in Figure 18. Following the trip distance, the trip duration showed to affect the intrapersonal variability in mode choice behaviour as well. The intrapersonal variability in mode use was significantly lower for short duration trips, while the opposite was expected. Additionally, individuals mainly use the same mode for work & business trips as observed in Figure 17. The figure suggests that, since the origin and destination are often the same for these trips, individuals tend to stick to their daily routine and that the daily mode choice for work & business trips is based on habit.

Also, the analysis showed that the effects of socioeconomic characteristics on the intrapersonal variability in mode choice are not significant. Only the effect of the gender on the intrapersonal variability in mode use was significant, which was explained by the fact that males drive more often to activities while women are more often a passengers.

Furthermore, no differences in the MIX-distributions were observed when more than two weeks of collected data was used. This leads to the conclusion that two weeks of data collection is a good time span to measure intrapersonal variability in mode use, since using data from an extended data collection time span did not affect the resulting MIX-distributions.

When the relative car shares were considered for the socioeconomic characteristics, the same phenomenon was observed. In general, the relative share of car trips decreases when the intrapersonal variability of the individuals increases.

Trip characteristics			**Socioeconomic characteristics**		
Purpose	Work & business	Ø=0	Urbanisation rate	Urban (1)	Ø=0,48
	Shopping	Ø=0,33		Slightly urban (2+3)	Ø=0,46
	Leisure	Ø=0,5		Rural (4+5)	Ø=0,42
Distance	0-5 km	Ø=0,43	Gender	Male	Ø=0,41
	5-15 km	Ø=0,35		Female	Ø=0,48
	>15 km	Ø=0,25	Partner status	Partner	Ø=0,43
Individual trip rates	Lower trip rates	Ø=0,42		No partner	Ø=0,47
	Higher trip rates	Ø=0,46	Age	<35 years old	Ø=0,47
Trip duration	0 - 20 min	Ø=0,40		35-55 years old	Ø=0,45
	20 - 60 min	Ø=0,43		>55 years old	Ø=0,43
	> 60 min	Ø=0,43	Education level	Higher educated	Ø=0,45
Other characteristics				Poorly educated	Ø=0,44
Rain	Rain	Ø=0,44	Income	Over modal income	Ø=0,44
	No rain	Ø=0,39		Below modal income	Ø=0,44
1 week data (1st)		Ø=0,41	Kids	No child in household	Ø=0,46
1 week data (5th)		Ø=0,40		Child in household	Ø=0,43
2 weeks data		Ø=0,44			
4 weeks data		Ø=0,44			
6 weeks data		Ø=0,44			
German mobility panel		Ø=0,34			

TABLE 6: RESULTS OF THE MIX-VALUE ANALYSIS

6. MODE CHOICE MODELLING AND MIX-SIMULATION

The previous chapter described the distribution of the MIX-variable based on observed values from the DMMP-dataset. Distributions of this variable representing the intrapersonal variability of the sample population were shown for both trip characteristics and socioeconomic characteristics. This chapter analyses whether the intrapersonal variability in mode choice behaviour can be described a mode choice model. The first section describes the construction of the mode choice model. The second section briefly describes the resulting model. Based on the constructed model the mode choice for the respondents is simulated. The MIX-value distributions resulting from the simulation are described in the third section.

6.1. CONSTRUCTION OF THE MODEL

Section 2.3 introduced the Multinomial Logit Model (MNL) as a discrete choice model used for mode choice modelling. This type of model tries to describe the mode choice process of a decision maker using utility functions for all mode alternatives containing characteristics which significantly explain mode choice variation.

Model estimations are performed using the software package BIOGEME (Bierlaire, 2003). A selection of the sample dataset was made for the model estimation. This selection contained all trips from the second data collection wave with the travel times for bicycle, car and public transport trips registered. Also, only trips were selected where the actual choice was one of these three modes. This selection was necessary since the travel times for these three mode alternatives were the only available characteristics describing the relative attractiveness of the three modes. Based on these criteria a total of 4286 trips were selected from 176 individuals.

This research already made assumptions regarding the mode availability for the bicycle and the private car. The availability of public transport is unknown for the selected trips. An assumption needs to be made since the model takes the availability of the modes into account in the estimation process. Since all selected trips have a travel time available for that trip in seconds, it is assumed that public transport is available for all trips.

Different models were estimated to create utility functions for the mode alternatives containing only significant parameters. Multiple characteristics were introduced in the utility functions iteratively to test their significance. Introduction of weather characteristics (rain, wind and temperature) and socioeconomic characteristics (gender, age, income, education level, kids and partner) did not result in significant parameters for the utility functions. So, while a significant effect on the MIX-distribution was observed for rain and gender in section 5.3, no significant parameters were estimated for these characteristics in the model.

Trip distance was significant for the bicycle and the car only. It became insignificant when distance was included in the utility function for public transport as well. The dummy variable for shopping trips was significant for the bicycle and the car as well, as were the dummy variables describing the urbanisation rate of the respondents' residence location

The model fit of the estimated models was assessed using the likelihood ratio index ρ^2. An increase of this value means an increase of the model fit compared to the previously estimated model. Following, using the likelihood ratio test it was assessed if the introduction of an extra estimated parameter significantly improved the model. The next section describes the resulting model.

6.2. MODELLING RESULTS

The iterative process of entering parameters in the utility functions of the three mode alternatives resulted in the following utility functions:

$$V_{Bike} = ASC_{Bike} + \beta_{TT} * TT_{bike} + \beta_d * d + \beta_{shop} * shopping + \beta_{Rural} * Rural + \beta_{SL\_urban} * SL\_Urban$$

$$V_{Car} = ASC_{car} + \beta_{TT} * TT_{Car} + \beta_d * d + \beta_{shop} * shopping + \beta_{Rural} * Rural + \beta_{SL\_urban} * SL\_Urban$$

$$V_{PT} = ASC_{PT} + \beta_{TT} * TT_{PT}$$

Where:

ASC	is the alternative specific constant for each mode;
β_{TT}	is the estimated parameter for the travel time, independent for every mode;
TT	represents the travel time in seconds for each mode;
β_d	is the estimated parameter for distance;
d	represents the trip distance;
β_{shop}	is the estimated parameter for the dummy variable shopping trips;
$shopping$	is a dummy variable with the value 1 if the trip purpose is shopping and the value is 0 when the trip purpose is other than shopping;
$Rural$	is a dummy variable with the value 1 if the respondent lives in a rural area;
β_{Rural}	is the estimated parameter for the dummy variable Rural;
SL_Urban	is a dummy variable with the value 1 if the respondent lives in a slightly urban area;
β_{SL_urban}	is the estimated parameter for the dummy variable SL_Urban.

The BIOGEME script used to estimate this model is shown in Appendix H. The model assessment statistics are shown in Table 7 where k represents the mode alternatives. Addition of the travel time and distance to the model resulted in a better model fit, represented by an increase in the ρ^2-values. Also, with the LR-values higher than the critical value the model performs significantly better with the added parameters.

Model	$\boldsymbol{\rho^2}$	**LL**	**LR**	**# estimated parameters**	**critical value (α=0,05)**
$V_k = ASC_k$	0,415	-2751		2	
$V_k = ASC_k + \beta_{TT} * TT_k$	0,455	-2563	-376	3	3,84
$V_k = ASC_k + \beta_{TT} * TT_k + \beta_d * d$	0,457	-2554	-18	4	3,84
$V_k = ASC_k + \beta_{TT} * TT_k + \beta_d * d + \beta_{shop} * shopping$	0,457	-2551	-6	5	3,84
$V_k = ASC_k + \beta_{TT} * TT_k + \beta_d * d + \beta_{shop} * shopping + \beta_{Rural} * Rural + \beta_{sl\_urban} * SL\_urban$	0,464	-2523	-56	7	5,99

TABLE 7: MODEL ASSESMENT STATISTICS

The estimation results of the resulting MNL-model are shown in Table 8, including the models' log-likelihood statistics. A negative parameter value means that the utility decreases when one extra unit of the characteristic is added. The small parameter values for the travel time result from the fact that the travel time was registered in seconds in the dataset. This means in this case the utility is lowered by a small amount when a trip would take one second longer. The only positive parameter value is the dummy variable for shopping trips. This means that the utilities of the alternatives car and bicycle increase when the trip purpose is shopping, compared to the utility for the same trip with an different trip purpose.

Parameter	Value	SD	T-test	P
ASC car	0			
ASC bicycle	-0,492	0,0505	-9,73	0,00
ASC PT	-1,62	0,219	-7,38	0,00
Distance (km)	-0,0451	0,00819	-5,15	0,00
Travel time (min)	-0,00112	0,0000675	-16,58	0,00
Shopping	0,773	0,385	2,01	0,04
Rural	2,59	0,444	5,84	0,00
SL_Urban	0,490	0,236	2,08	0,04
Log-likelihood	Initial	-4707,554		
	Final	-2522,605		
ρ^2		0,464		

TABLE 8: ESTIMATION RESULTS FOR MNL MODEL

6.3. SIMULATION RESULTS

The previous section showed the estimation results for the MNL model. Based on the estimated parameter values a mode choice simulation is performed using BIOGEME. The simulation is performed on the same data as the parameter values were estimated upon. As a result from the simulation an observed mode choice and a simulated mode choice is available for the trips in the modelling dataset.

Following the simulation, the MIX-value for the individuals is calculated for both the observed modes as well as the modelled modal choices. Since only three mode alternatives were distinguished in the model, the MIX-value for the individuals was calculated using three mode alternatives as well. Cumulative distribution functions are plotted based on the individual MIX-values using the same process as in the analysis in section 5.3. This section first describes the MIX-value distributions for all data. Secondly, the MIX-value distributions are described based on trip characteristics. Third, the MIX-value distributions are described based on socioeconomic characteristics. The significance of the difference between the cumulative distribution functions were tested using the K-S test as well.

6.3.1. MIX-VALUE DISTRIBUTIONS BASED ON ALL DATA

Figure 28 shows the observed and modelled MIX-distributions based on all trips in the dataset. The figure shows that the observed MIX-distribution is significantly different compared to the simulated MIX-distribution function (KS-test: D=0.146, p<.05). Also, there is a large difference in the median values (simulated Ø=0.33; observed Ø=0.13).

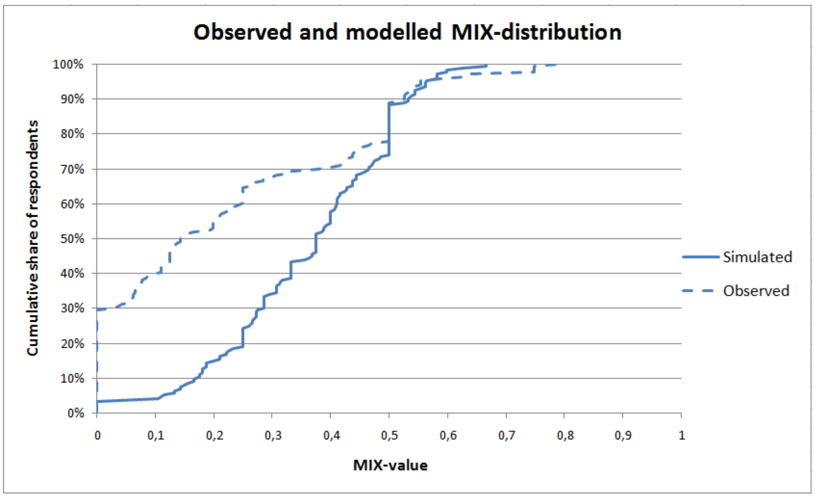

FIGURE 28: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON ALL TRIPS

6.3.2. MIX-VALUE DISTRIBUTIONS FOR SOCIOECONOMIC CHARACTERISTICS

Figure 29 shows the observed and simulated MIX-distributions, where three distance classes are distinguished. Less intrapersonal variability in mode choice behaviour is observed when the trip distance increases, as was already observed in Figure 18. The figure shows significant differences between the observed and simulated MIX-distributions for trips with a distance below fifteen kilometres. No significant difference was observed for trips over fifteen kilometres (KS-test: D=0.18, p<.05). Additionally, the differences between the simulated MIX-distributions are larger compared to the differences between the observed MIX-distributions. This means that apart from the systematic difference between the simulated and observed values, the model performs worse when trips are split according to their distance classes.

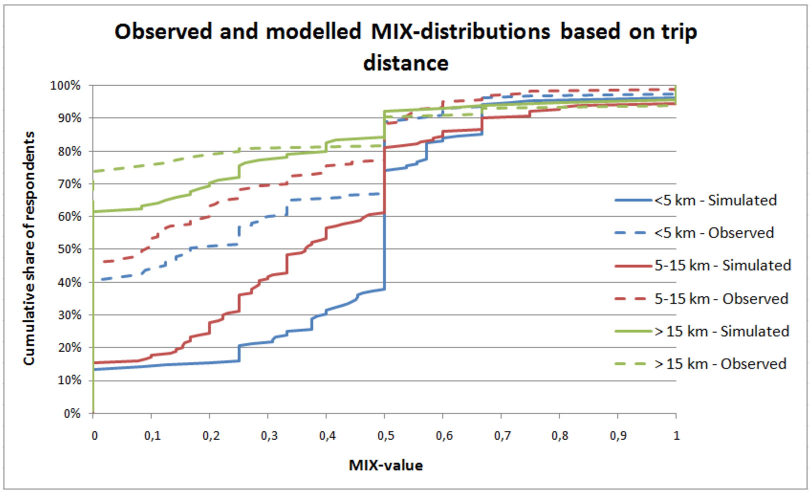

FIGURE 29: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON TRIP DISTANCE

The observed and simulated MIX-distributions with three distinguished trip purposes are shown in Figure 30. Again, the difference between the observed and the modelled distributions are significant. The simulated distributions are merely the same while the observed distributions show differences in the intrapersonal variability in mode choice behaviour.

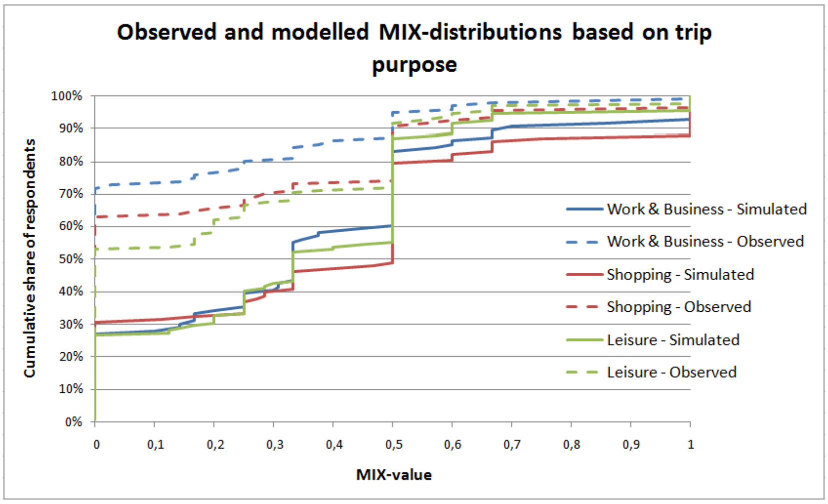

FIGURE 30: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON TRIP PURPOSE

6.3.3. MIX-VALUE DISTRIBUTIONS BASED ON SOCIOECONOMIC CHARACTERISTICS

Figure 22 showed that the intrapersonal variability in mode choice differs significantly between men and women; women showed more intrapersonal mode choice variability. Figure 31 shows the observed and simulated MIX-distributions for both genders. The figure shows a significant difference between the observed and simulated MIX-distributions.

Additionally, both the observed and the simulated MIX-distributions show merely the same values when comparing the genders. This is different compared to Figure X where women showed more intrapersonal variability in mode choice behaviour. The change can be explained by the calculation method. The previous chapter distinguished five mode alternatives in the calculation of the MIX-value. This section only uses three mode alternatives since only data about these three were available for the model. Since for the calculation method in this section the alternatives 'car passenger' and 'car driver' are merged, less difference in the intrapersonal variability is observed for women; Figure 39 showed only modal split differences between men and women for these two mode alternatives.

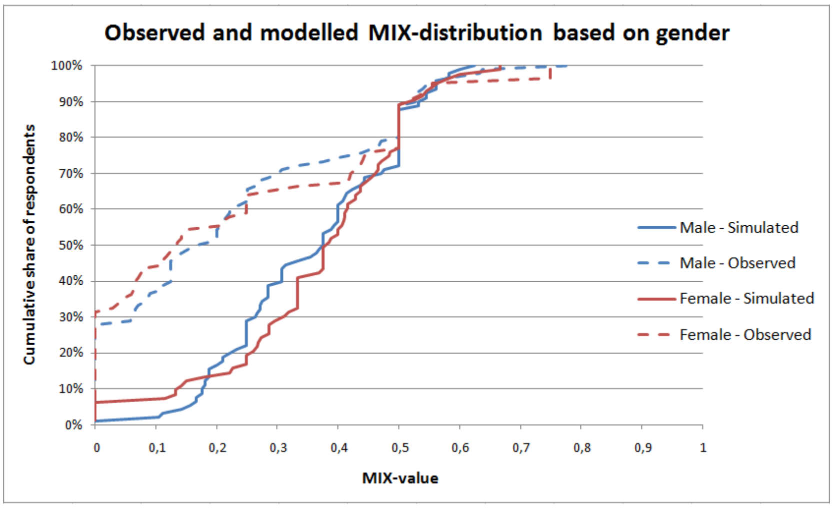

FIGURE 31: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON GENDER

The observed and simulated MIX-distributions based on partner status of the respondents is shown in Figure 32. Again, a significant difference is observed between the observed and simulated MIX-distributions. While the observed intrapersonal variability for individuals without a partner is higher for the observed values, this is the opposite for the simulated values. However, the differences are small.

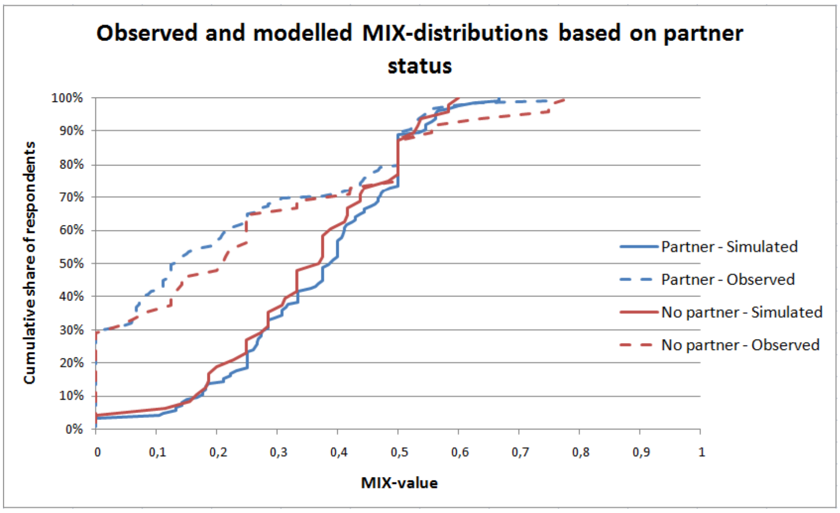

FIGURE 32: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON PARTNER STATUS

Figure 33 shows the observed and simulated MIX-distributions where a distinction is made between individuals with a child present in the household and individuals without a child present in the household. The observed and simulated MIX-distributions are significantly different for both identified groups. As observed in Figure 24, the group of individuals without children in the household shows more intrapersonal variability in their modal choice compared to the group with children in the household. However, the differences between the two are small.

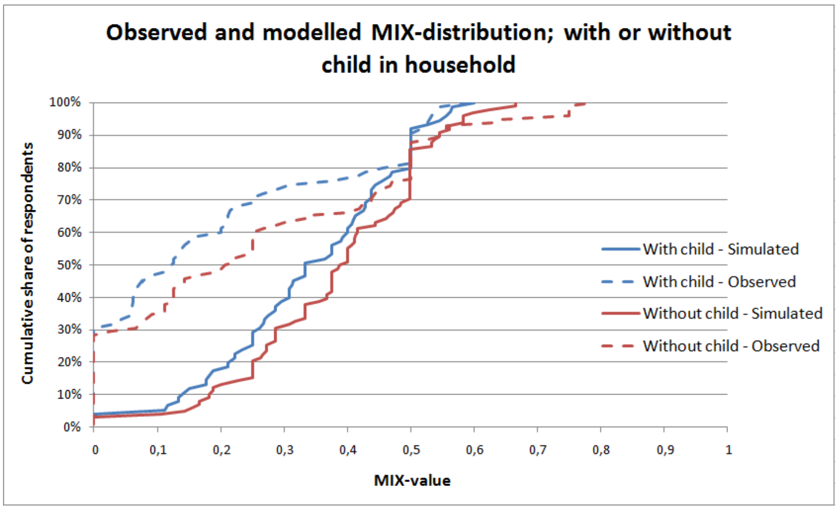

FIGURE 33: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON CHILD IN HOUSEHOLD

As can be observed in Figure 34, no differences in simulated MIX-distributions are present when comparing the lower income group of individuals with the higher income level group. Again, the differences between the observed and simulated distributions are significant.

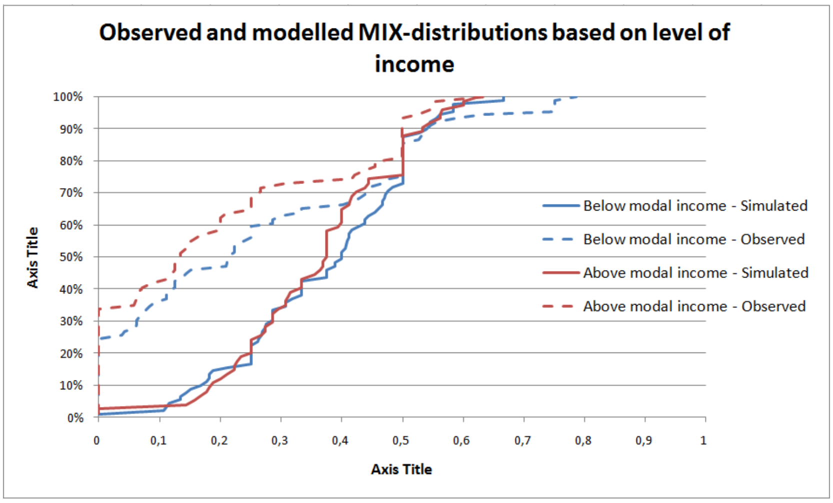

FIGURE 34: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON LEVEL OF INCOME

The observed and simulated MIX-distributions based on the urbanisation rate are shown in Figure 35. Again, the difference between the observed and the modelled distributions for all urbanisation rates are significant. Additionally, the intrapersonal variability in mode use of individuals living in urban areas is the highest for both the observed and the simulated mode choices. The differences between the simulated MIX-distributions are not significant, as was also the case for the observed MIX-distributions in Figure 21.

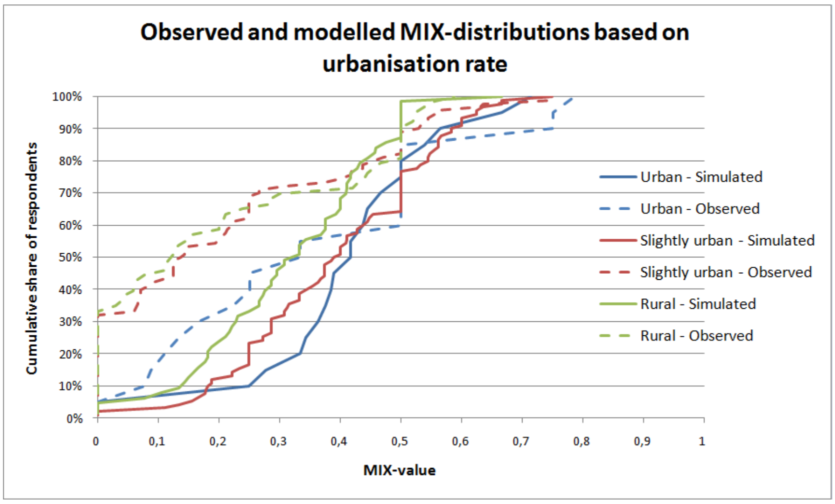

FIGURE 35: OBSERVED AND MODELLED MIX-DISTRIBUTIONS BASED ON URBANISATION RATE

6.4. Conclusion and evaluation

This chapter described the construction of a mode choice model. Modelling was done using a MNL model in BIOGEME. A mode choice simulation was performed using the constructed model, based on which the intrapersonal variability in mode use was calculated for both the observed and the simulated modal choices, using the same method as used in section 5.3.

As also observed in section 5.3, socioeconomic characteristics for the simulated MIX-distributions have no significant effect on the intrapersonal variability. Only the trip distance had a significant effect on the simulated MIX-distribution. Trip purpose did not have a significant effect on the simulated MIX-distributions, while this was the case in section 5.3.

The simulated MIX-distributions were ordered in the same way as the observed MIX-distributions. This shows that modelling using the utility theory captures the effects of characteristics on the intrapersonal variability in mode use. However, the MIX-distributions showed significant differences between the observed and simulated MIX-values for all the characteristics described in section 6.3. A higher level of intrapersonal variability is expected according to the simulation, but is not observed. This systematic difference shows that intrapersonal variability in mode choice is difficult to model using the utility theory. Altogether this leads to the conclusion that both the economist view using the utility theory, as the theory of planned behaviour using habit, play a role in the mode choice process of individuals.

7. Conclusions

The previous chapters described the research performed for this thesis. This chapter reports the conclusions of this research by providing answer to the research questions.

1. *What are important attributes that influence the modal choice according to the literature?*

Many characteristics that significantly influence the modal choice were found in the literature varying from individual characteristics, psychological characteristics and trip characteristics to external characteristics as weather influences and network disruptions. The main elements influencing the modal choice can be divided in four categories; individual characteristics, mode specific characteristics, trip-specific characteristics and external characteristics. In addition to these groups, household interaction influences the modal choice as well. Figure 36 shows the interrelationships between the characteristics identified in the literature. The coloured characteristics were available in the research dataset used for this thesis.

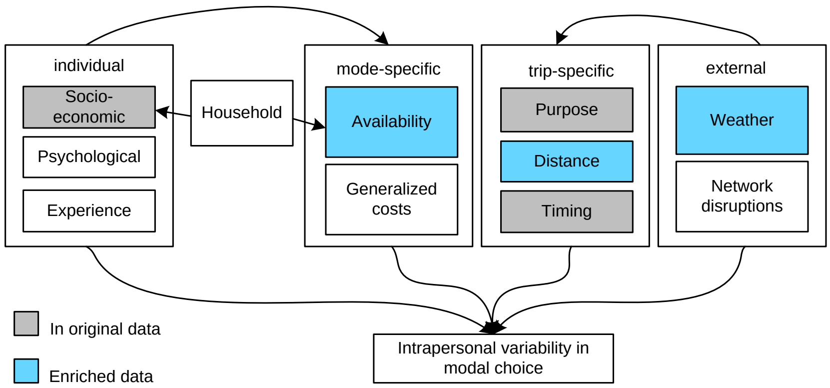

FIGURE 36: CONCEPTUAL MODEL ADAPTED FOR DATA AVAILABILITY

2. *What are the available methods to measure intrapersonal variability in mode choice behaviour?*

The theory showed two methods to measure intrapersonal variability in mode use. Buehler & Hamre (2013) expressed intrapersonal variability as a share of individuals being multimodal after several observation intervals. An extended data collection time interval resulted in higher shares of multimodality. Individuals were categorized as multimodal after registering a second mode used in the data collection period. Using this method on the available data would result in all individuals to be multimodal.

Kuhnimhof (2009) expressed intrapersonal variability with a Mode Variation Index (MIX), representing a relative deviation from a state of maximum variability. A MIX-value can be calculated for every individual, based on the observed frequencies. By using these values to create a cumulative distribution function, the intrapersonal variability in mode use of a population can be determined and compared to other samples. Since this method is more capable to handle different data collection intervals in the Dutch context, as described in section 5.3, this method was used in the research to express intrapersonal variability in mode use behaviour.

3. *What characteristics explain intrapersonal variability in mode choice behaviour?*

The research first analysed the effect of the data collection time span on the intrapersonal variability in mode use of the respondents. Collecting two weeks of data is regarded to be sufficient in order to capture the intrapersonal variability of a sample population based on this study. Research showed that collection of more data did not result in different results.

Secondly, the intrapersonal variability in mode choice behaviour of the panel respondents was analysed for both trip characteristics as socioeconomic characteristics using the method of Kuhnimhof (2009). The research showed that socioeconomic characteristics, apart from the gender, do not have a significant effect on the intrapersonal variability in mode use. The intrapersonal variability in mode use of females is higher compared to the intrapersonal variability in mode use of men.

A significant effect of multiple trip-specific characteristics was found on the intrapersonal variability in mode use. Rain, trip distance and trip purpose all showed to significantly affect the intrapersonal variability in mode use, while the trip travel time showed no significant effect.

The median values of the MIX-distributions were used to compare the intrapersonal variability of the characteristics. Table 9 shows the resulting median values for the analysed characteristics. Next to the socioeconomic- and trip characteristics, the intrapersonal variability in mode use of the Dutch was compared with the results of Kuhnimhof (2009) in the German context. It was concluded that, since all median values of the analysed data collection time spans are higher, Dutch individuals show a higher intrapersonal variability in mode use compared to German individuals.

Trip characteristics		
Purpose	Work & business	Ø=0
	Shopping	Ø=0,33
	Leisure	Ø=0,5
Distance	0-5 km	Ø=0,43
	5-15 km	Ø=0,35
	>15 km	Ø=0,25
Individual trip rates	Lower trip rates	Ø=0,42
	Higher trip rates	Ø=0,46
Trip duration	0 - 20 min	Ø=0,40
	20 - 60 min	Ø=0,43
	> 60 min	Ø=0,43
Other characteristics		
Rain	Rain	Ø=0,44
	No rain	Ø=0,39
1 week data (1st)		Ø=0,41
1 week data (5th)		Ø=0,40
2 weeks data		Ø=0,44
4 weeks data		Ø=0,44
6 weeks data		Ø=0,44
German mobility panel		Ø=0,34

Socioeconomic characteristics		
Urbanisation rate	Urban (1)	Ø=0,48
	Slightly urban (2+3)	Ø=0,46
	Rural (4+5)	Ø=0,42
Gender	Male	Ø=0,41
	Female	Ø=0,48
Partner status	Partner	Ø=0,43
	No partner	Ø=0,47
Age	<35 years old	Ø=0,47
	35-55 years old	Ø=0,45
	>55 years old	Ø=0,43
Education level	Higher educated	Ø=0,45
	Poorly educated	Ø=0,44
Income	Over modal income	Ø=0,44
	Below modal income	Ø=0,44
Kids	No child in household	Ø=0,46
	Child in household	Ø=0,43

TABLE 9: MEDIAN VALUE OF THE ANALYSED MIX-DISTRIBUTIONS

4. *Can the observed intrapersonal variability in mode choice be approximated using a mode choice model?*

Chapter 6 described the construction of a mode choice model, based on the trips in the dataset containing travel time in seconds for the bicycle, car and public transport. The model used was a MNL model for which the travel time, trip distance, purpose and urbanisation rate of the residence location were significant parameters. The ρ^2-value of 0,464 of the final model showed a good model fit.

The estimated model was used to simulate the modal choice. Using the simulated mode choices, simulated MIX-distributions were plotted for important characteristics. These simulated MIX-distributions were ordered in the same way as the observed MIX-distributions. This shows that modelling using the utility theory captures the effects of characteristics on the intrapersonal variability in mode use. However, the MIX-distributions showed significant differences between the observed and simulated MIX-distributions for all analysed characteristics. A higher level of intrapersonal variability is expected according to the simulation, but is not observed. This systematic difference shows that intrapersonal variability in mode choice is difficult to model using the utility theory. Altogether this leads to the conclusion that both opposed views from the literature, the economist view using the utility theory, and the theory of planned behaviour using habit, play a role in the mode choice process of individuals.

8. DISCUSSION AND FURTHER RESEARCH

The previous chapters described the performed research for this thesis. Reflection on the processes used in this research is required; this chapter deliberates on that observation and contains three sections. The first section deliberates on the processes used to construct the dataset used for the analysis. The second section places some remarks by the used analysis methods and the results of this thesis and the third section provides some directions for further research.

8.1. DATA PROCESSING AND SELECTION

The data collected in the DMMP-project contained a lot of missing values. Large effort was put in the process to create a consistent database containing multiple essential variables that were not registered during the data collection period. First, missing postcodes were inserted using an algorithm to transform the manually entered descriptions of the origin and destination locations. These were used to find the matching postcode in a list containing all postcodes of the Netherlands. Large variations existed in how respondents filled in their origin and destination locations, even variations existed when a respondent registered a certain location multiple times. This is probably the effect of providing one field for entry of the location. Addition of fields in a way that street, town and postcode can be filled in separately can result in less variations in location entries and therefore a higher success rate in the addition of missing postcodes.

Trip distance, location type data and KNMI weather data were inserted and not measured during the data collection period. The accuracy of the inserted distances is questionable since the distances for walking and bicycle trips were based on average speed estimates. Trip distances for other modes were inserted based on a matrix containing distances between all postcodes on a postcode-4 level. This overestimated the trip distance for short trips since the radius of the postcode-4 area was used as distance when the trip was actually shorter.

After the insertion of the postcodes, respondents providing data with a proper quality were selected for further research. These respondents were also selected based on the amount of days they filled in the daily questions where possible influences on their daily travel patterns were asked. When individuals failed to fill in these questions for multiple days, respondents were not selected for the research. During the research the daily questions were not used to analyse the influence on the travel behaviour by the respondents. By deleting the data from respondents with a few days of filled in daily questions it is possible that some respondents were not included in the research while they might have provided proper quality data.

Respondent selection was based on the shares of erroneous trips, which also included trips without a complete pair of postcodes for both the origin and the destination. Some alterations were made in the algorithm used to find missing postcodes after the respondent selection. Using the new algorithm resulted in a higher rate of success for the addition of missing postcodes. When the final algorithm was used on the initial data, possibly more respondents would have been selected for analysis.

Some assumptions about the mode availability for individuals were made during this research. It was assumed that a bicycle is available for all individuals in the Netherlands. Also, it was assumed that a private car is available when at least one trip was registered with car driver as the chosen mode. The share of individuals with a car available was the same as the car availability in OViN and therefore it was assumed to be a valid assumption. Additionally it was assumed in the modelling dataset that public transport was available as a feasible mode for all trips. However, these three assumptions can be untrue for multiple individuals. Since the parameters of the MNL model are estimated based on the available mode alternatives, it is possible that faults in the identified mode availability influenced the estimation results.

8.2. Analysis and results

This research used the method of Kuhnimhof (2009) to measure intrapersonal variability in mode choice behaviour. The method was able to handle multiple time intervals of data and resulted in an intrapersonal variability value for every individual. Apart from the variability MIX-value, no information about what modes were actually used by the respondents is available when using this method of analysis. An individual showing no intrapersonal variability at all can both be a car driver or a bicyclist. By including the shares for car and non-car trips some information is added to the analysis in this research.

Chapter 6 showed that it is difficult to capture intrapersonal variability in mode choice behaviour in a model. Since the order of the observed and simulated MIX-distributions were the same it was concluded that both the economists view using the utility theory, and the theory of planned behaviour using habit, at least partly explain intrapersonal variability in mode choice behaviour. The conclusions of this analysis match the conclusion of Thøgersen (2006), who showed that an effect of past mode choice behaviour on current mode choice behaviour exists. Individuals do not consider all available modes in their choice set again when a repetitive trip is made. Therefore, individuals act out of habit and do not always choose the mode with the highest utility.

Sabir et al. (2008) stated that in rural areas the private car is preferred over other modes, but that the effects of the degree of urbanisation on mode choices are small. This research confirmed in Figure 50 that the car is preferred in rural areas. Additional to the work of Sabir et al. (2008), this thesis analysed the intrapersonal variability in mode use based on urbanisation rate. It was shown that an increase of the urbanisation rate increases the intrapersonal variability in mode use as well, but the effects on the MIX-distributions were small.

8.3. Directions for further research

No data of public transport level of service was available for this research, which lead to the assumption that public transport was available for all trips in section 6.1, an assumption that is most likely not entirely correct. For some regions the level of service of the public transport system is very low, for instance large access and egress walking distances to the public transport system are present, and public transport is considered not feasible for trips made in those regions. To make better predictions using the model, the level of service of the public transport system can be taken into account. When the level of service of the public transport system falls below a set threshold, public transport is considered as an infeasible mode and the probabilities of choosing the other available alternatives will increase.

Furthermore, in future research one can research the intrapersonal variability in mode use of the respondents over time since new waves of data will be collected. Appendix G already showed the differences when two, four and six weeks of data are considered. With the third wave of data collected in the DMMP panel, the individual MIX-values can be plotted for eight and ten weeks as well. This gives insight in the change on the MIX-value when extra data is added. Additionally, it is possible to create the same type of figure as Figure 64, but only calculate the MIX-value for every period of two weeks rather than adding two weeks every time. This will show the stability (or instability) of an individuals' mode use variability. When significant differences in MIX-values occur for individuals, it is possible to analyse why the behaviour of these individuals altered in future research.

The research was at first only based on the first wave of collected data. During the research the second wave became available and was then included in the dataset to have more data per person available for the research. The third wave of data will be collected in 2015, resulting in even more available data for every individual. The availability of more data means that individuals (might have) repeated some of their trips multiple times. While the analysis performed in this research using the method of Kuhnimhof (2009) was based on intrapersonal variability in mode use on an individual level, it is possible to research intrapersonal variability in mode use on a trip level. Research questions as: “Do individuals always use the same mode for the same trip? And if not, what characteristics do these individuals and trips have?” can then be answered. The available individual MIX-value is then an indicator for individuals who are more prone to use multiple modes of transportation on a trip level.

9. References

Aarts, H., Verplanken, B., & Van Knippenberg, A. (1998). Predicting behavior from actions in the past: Repeated decision making or a matter of habit? *Journal of Applied Social …*, *28*(15), 1355–1374. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1559-1816.1998.tb01681.x/full

Azjen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Bamberg, S., Ajzen, I., & Schmidt, P. (2003). *Past behavior and reasoned action*. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0

Bergström, a, & Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A: Policy and Practice*, *37*(8), 649–666. doi:10.1016/S0965-8564(03)00012-0

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. In *Proceedings of the 3rd Swiss Transportation Research Conference*. Ascona: EPFL.

Bohte, W. (2010). *Residential self-selection and travel: The relationship between travel-related attitudes, built environment characteristics and travel behaviour*. *The rela*. TU Delft.

Buehler, R., & Hamre, A. (2013). *Trends and Determinants of Multimodal Travel in the USA*. Retrieved from http://trid.trb.org/view.aspx?id=1264713

Chikaraishi, M., Fujiwara, A., Zhang, J., & Axhausen, K. W. (2011). Identifying variations and co-variations in discrete choice models. *Transportation*, *38*(6), 993–1016. doi:10.1007/s11116-010-9317-6

Clifton, K., & Handy, S. (2001). Qualitative methods in travel behaviour research. Kruger National Park, South Africa. Retrieved from http://www.researchgate.net/publication/228811651_Qualitative_methods_in_travel_behaviour_research/file/60b7d52c1baafa6ff0.pdf

Diana, M., & Mokhtarian, P. L. (2009). Grouping travelers on the basis of their different car and transit levels of use. *Transportation*, *36*(4), 455–467. doi:10.1007/s11116-009-9207-y

Exel, & Rietveld. (2010). Perceptions of public transport travel time and their effect on choice-sets among car drivers. *Journal of Transport and Land Use*, *2*(3-4), 75–86. Retrieved from http://repub.eur.nl/pub/23335/

Field, A. (2009). *Discovering statistics using SPSS* (Third Edit.).

Gärling, T., & Axhausen, K. (2003). Introduction: Habitual travel choice. *Transportation*, *30*, 1–11. Retrieved from http://link.springer.com/article/10.1023/A:1021230223001

Kenyon, S., & Lyons, G. (2003). The value of integrated multimodal traveller information and its potential contribution to modal change. *Transportation Research Part F: Traffic Psychology …*, *6*, 1–21. doi:10.1016/S1369

Kitamura, R. (1990). Panel analysis in transportation planning: An overview. *Transportation Research Part A: General*, *24*(6), 401–415. Retrieved from http://www.sciencedirect.com/science/article/pii/0191260790900322

Klöckner, C. a., & Friedrichsmeier, T. (2011). A multi-level approach to travel mode choice – How person characteristics and situation specific aspects determine car use in a student sample. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(4), 261–277. doi:10.1016/j.trf.2011.01.006

Kroesen, M. (2014). Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. *Transportation Research Part A: Policy and Practice*, *65*, 56–67. Retrieved from http://www.sciencedirect.com/science/article/pii/S0965856414000974

Kuhnimhof, T. (2009). *The Implications of a Longitudinal Perspective on Travel for Modelling, Forecasting and Policy Making*. Retrieved from http://www.eurocities-datta.eu/documents/tobias.pdf

Lin, P., Wu, B., & Watada, J. (2010). Kolmogorov-Smirnov two sample test with continuous fuzzy data. *Integrated Uncertainty Management and …*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-11960-6_17

Louviere, J. J., Henscher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: analysis and application*. Cambridge University Press.

Minesnot, G., & Takano, S. (2005). Modeling the Relationship between Seasonal Constraints and Modal Choice Characteristics of High School Students in Snowy Regions. *Journal of the Eastern Asia Society for Transporation Studies*, *6*, 1844–1857.

Ministerie van infrastructuur en Milieu, M. van infrastructuur en M. (2011). Resultaten Mobiliteitsprojecten. *Den Haag*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Resultaten+mobiliteitsprojecten#0

Mokhtarian, P., Salomon, I., & Redmond, L. (2001). Understanding the demand for travel: It's not purely'derived'. *Innovation: The European …*, *14*(4), 355–380. doi:10.1080/13511610120106

Olde Kalter, M., Hoogendoorn-lanser, S., & Geurs, K. (2014). De eerste resultaten van het Mobiliteitspanel Nederland. In *Colloquium Vervoersplanologisch Speurwerk*.

Ortúzar, J. de D., & Willumsen, L. G. (2001). *Modelling Transport* (Third Edit.). John Wiley & Sons.

Pas, E. I. (1987). Intrapersonal Variability and model goodness-of-fit. *Transportation Research Part A: General*, *21*(6), 431–438.

Pas, E. I., & Koppelman, F. S. (1986). An examination of the determinants of day-to-day variability in individuals ' urban travel behavior. *Transportation*, *200*, 183–200.

Pas, E. I., & Sundar, S. (1995). Intrapersonal variability in daily urban travel behavior : Some additional evidence. *Transportation*, *22*, 135–150.

Planbureau voor de Leefomgeving. (2003). Bestand locatietypen per postcode 4.

Sabir, M., Koetse, M., & Rietveld, P. (2008). The impact of weather conditions on mode choice: empirical evidence for the Netherlands. *Department of Spatial Economics, …*, *2100*(Stern 2006). Retrieved from http://www.webmeets.com/files/papers/EAERE/2009/1021/Sabir.pdf

Savelberg, F., & Korteweg, J. A. (2011). *Slim benutten : bereikbaarheidsmaatregelen op een rij*. Den Haag.

Scheiner, J. (2010). Social inequalities in travel behaviour: trip distances in the context of residential self-selection and lifestyles. *Journal of Transport Geography*, *18*(6), 679–690. doi:10.1016/j.jtrangeo.2009.09.002

Scheiner, J., & Holz-Rau, C. (2013). A comprehensive study of life course, cohort, and period effects on changes in travel mode use. *Transportation Research Part A: Policy and Practice*, *47*, 167–181. doi:10.1016/j.tra.2012.10.019

Schlich, R. (2001). Analysing intrapersonal variability of travel behaviour using the sequence alignment method. In *Paper presented at the European Transport Conference*. Cambridge.

Schönfelder, S., & Ax. (2010). *Urban Rhythms and Travel Behaviour. Spatioal and Temporal Phenomena of Daily Travel*. Farnhem/Burlington: Ashgate Publishing.

Sheth, J. N. (1976). A Psychological Model of Travel Mode Selection. *Advances in Consumer Research*, *3*, 425–430.

Stopher, P. R., & Zhang, Y. (2011). The repetitiveness of daily travel. Paper presented at the Transportation Research Board Annual Meeting, Washington, January 2011.

Thøgersen, J. (2006). Understanding repetitive travel mode choices in a stable context: A panel study approach. *Transportation Research Part A: Policy and Practice*, *40*(8), 621–638. doi:10.1016/j.tra.2005.11.004

Thomas, T., Bijlsma, M., & Geurs, K. (2013). Hoe mobiel zijn we nu eigenlijk? Eerste inzichten uit het Mobiele mobiliteitspanel. Paper presented at the CVS, Rotterdam. Retrieved from http://www.cvs-congres.nl/cvspdfdocs_2013/cvs13_060.pdf

Train, K. (2002). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Van de Kruijs, T. (2013). *Leisure facilities in railway station areas*. University of Twente.

Van Exel, N. J. a., & Rietveld, P. (2009). When strike comes to town…anticipated and actual behavioural reactions to a one-day, pre-announced, complete rail strike in the Netherlands. *Transportation Research Part A: Policy and Practice*, *43*(5), 526–535. doi:10.1016/j.tra.2009.01.003

Van Wee, B., & Dijst, M. (2002). *Verkeer en Vervoer in hoofdlijnen*. Uitgeverij Coutinho.

Whalen, K. E., Páez, A., & Carrasco, J. a. (2013). Mode choice of university students commuting to school and the role of active travel. *Journal of Transport Geography*, *31*, 132–142. doi:10.1016/j.jtrangeo.2013.06.008

Wikipedia. (n.d.). Kolmogorov-Smirnov test. Retrieved December 03, 2014, from http://en.wikipedia.org/wiki/Kolmogorov–Smirnov_test

Wolf, J. (2000). *Using GPS Data Loggers To Replace Travel Diaries In the Collection of Travel Data*.

Zhu, S., & Levinson, D. (2012). Disruptions to transportation networks: a review. *Network Reliability in Practice*, 5–20. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-0947-2_2

APPENDICES

Appendix A – Calculating the MIX-variable

As already described in Chapter 1, intrapersonal variability in mode use has been an underlying issue in the literature compared to other influences on the modal split. From the literature, Kuhnimhof (2009) provides a method to measure the intrapersonal variability of respondents in a panel research. Kuhnimhof based his analysis on data from the German Mobility Panel (MOP). The measure of intrapersonal variability can be described as a value for the relative deviation from a state of maximum variation. The method to calculate this measure is described below and follows several calculation steps.

First, the classified mode categories are ordered per respondent in a decreasing manner. The mode used the most times is placed first and the least used as last. These ordered values represent the observed frequencies in mode use. With these ordered frequencies the following calculations are performed.

Second, frequencies for mode use in a state of maximum variability are modeled using the following equation:

$$mF_j = integer\left(\frac{N_T}{N_M}\right) + H_j$$

Where N_T represents the number of observed trips per respondent and N_M represents the number of classified modes for this research. The term H_j is introduced to add non-assigned trips to the modes. Following, the number of choices deviating from the maximum variation was calculated using:

$$SD = 0{,}5 \sum_j abs\left(rF_j - mF_j\right)$$

Where rF_j represents the observed frequency of the mode used. The number of relevant choices made by the respondent (a measure comparable to degrees of freedom) was calculated by subtracting the largest modeled frequency from the number of registered trips:

$$NRC = N_T - mF_1$$

The Mode Variation Index (MIX) is then calculated by subtracting the fraction resulting from dividing the number of choices deviating from the maximum variation by the number of relevant choices

$$MIX = 1 - \frac{SD}{NRC}$$

A sample calculation of this MIX-variable is provided in Table 10. As shown in the, a MIX-value is calculated for every individual in a research. With a large sample of individuals a cumulative distribution function can be plotted. An example of a cumulative distribution function based on data from the German mobility panel is shown in Figure 9. Here, the distribution intercepts with the Y-axis at the 0,1 point. This means that ten percent of all individuals show no intrapersonal variability in their mode use at all and only used one mode of transportation during the data collection period.

Modes by use		**Respondent 1**				**Respondent 2**			
	j	Modes by mode use frequency	rFj	mFj	abs(mFj-rFj)	Modes by mode use frequency	rFj	mFj	abs(mFj-rFj)
Most used mode	1	Auto	84	38	46	Fiets + te voet	81	47	34
2nd most used mode	2	Fiets + te voet	29	38	9	Auto	41	46	5
3rd most used mode	3	OV + overige	1	38	37	OV + overige	17	46	29
Number of trips			114				139		
SD					46				34
NRC					76				92
MIX					**0,395**				**0,630**

TABLE 10: SAMPLE CALCULATION MIX VARIABLE

Appendix B – Appending KNMI data

Mobidot added weather conditions to the original registered trips during the data collection period. Trips verified by the respondents did not contain any weather conditions at all, so in the process of verifying trips the weather conditions were lost (a new record was created for every alteration and verification in the process). Even so, the weather conditions which were added to the original trips had no clear classifications and the number of available characteristics was very limited (only temperature, rain and sky condition). Based on these arguments the choice was made to add weather data as gathered by the Royal Dutch Weather Institution (KNMI) to the dataset. This appendix describes the process of doing so. Data provided by the KNMI is high quality data with a clear classification and is therefore even more useful than the original weather data as collected by Mobidot.

Weather data can be added based on the postcode since this was registered for the trips and it is a location characteristic. To add the weather data from the nearest weather station, it was needed to determine the nearest weather station for every postcode in the Netherlands. Following the assumption that respondents make a modal choice based on the weather condition at the origin location rather than the destination location, data from the hour of departure from the nearest weather station was appended to the trips.

Several process steps were needed to append the weather data to the trips. First, a map was created containing all weather stations of the KNMI currently collecting weather data in the Netherlands. This map was created using open source GIS software called Quantum GIS and freely available weather station coordinates. As a background the map of the Netherlands was used provided by OpenStreetMaps.

Following, the shortest distance to each of the weather stations needed to be determined. This was done by creating Thiessen polygon layer containing Thiessen polygons for all weather stations, as shown by Figure 37. A map layer containing the centroids of all postcodes in the Netherlands on a postcode-4 level was then added to the map (Figure 38). By overlaying the data, the ID number of the weather station in the corresponding Thiessen polygon was added to the postcode. Based on the ID number of the weather station, the origin postcode and the hour of departure, it was possible to append the KNMI data to the trips.

Not all verified trips contained origin postcodes. For the trips not containing the postcode of the trip origin, no weather data was added at all and the cells were left as blank. Because no location data could be transferred from CentERdata it was not possible to add the corresponding weather stations to these trips when GPS coordinates for these locations were available in the dataset.

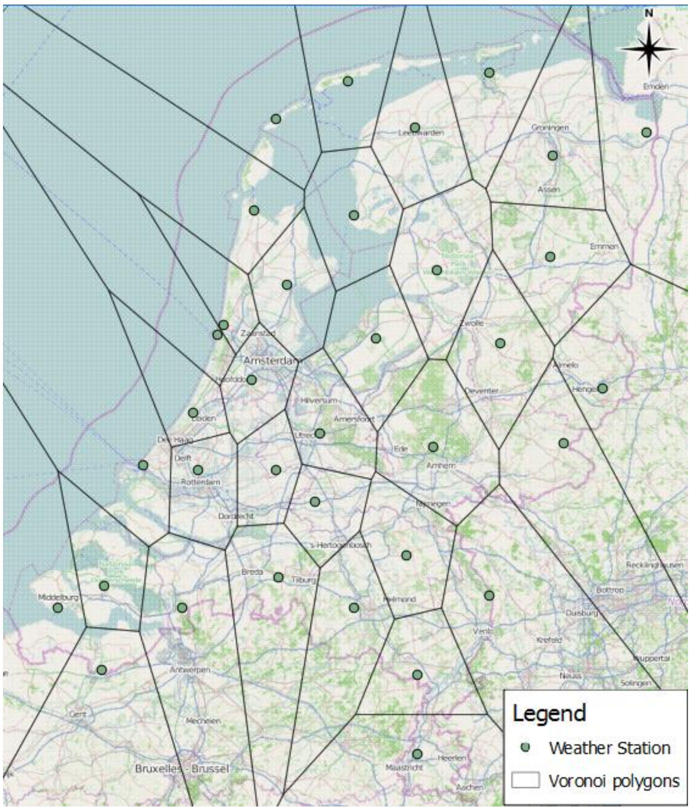

FIGURE 37: DISTRIBUTION OF WEATHER STATIONS IN THE NETHERLANDS

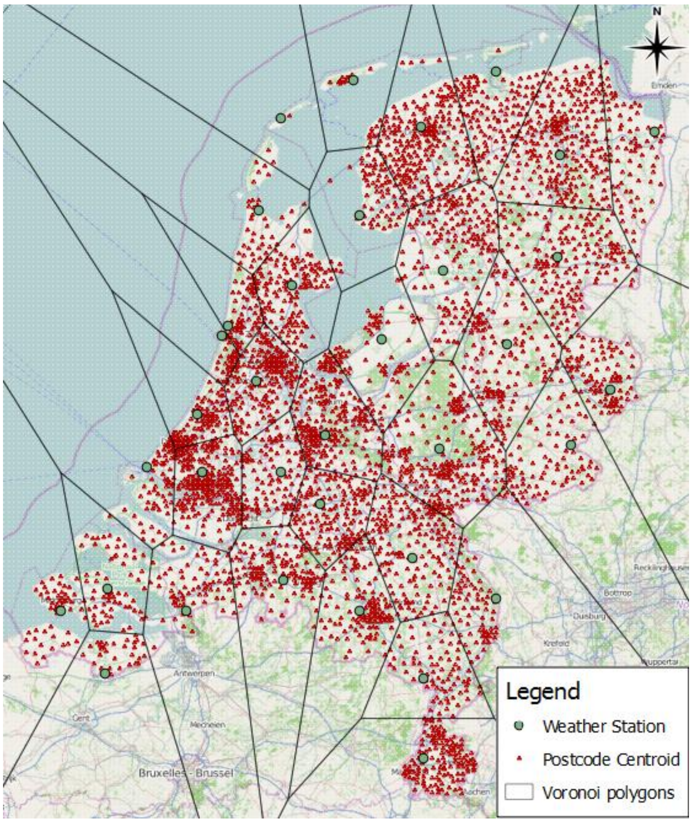

FIGURE 38: MAP SHOWING THE DISTRIBUTION OF POSTCODES IN THE NETHERLANDS

APPENDIX C – AVAILABLE CHARACTERISTICS FOR THIS RESEARCH

This appendix shows the available characteristics for this research based on the results of the data collection and processing as described in Chapter 3. The characteristics are divided in four categories; trip characteristics, weather characteristics, location characteristics and socioeconomic characteristics of the respondents.

Trip characteristics

- Trip purpose
- Trip duration
- Trip length/distance
- Trip departure and arrival time
- Trip departure and arrival postcodes
- Departure and arrival location (street, town)
- Travel mode used

Weather characteristics

- Wind strength
- Hourly precipitation
- Precipitation
- Temperature
- Sky condition
- Amount of sunshine at the hour of the trip

Location characteristics

- Distance to nearest highway
- Distance to nearest railway station
- Distance to nearest intercity railway station
- Distance to nearest metro/streetcar/bus station
- Location type according to the ABCR classification

Socioeconomic characteristics

- Age
- Sex
- Income
- Household income
- Household composition
- Occupation
- Level of education
- Urbanisation rate residence location
- Housing type (rented home or bought)
- Phone type

APPENDIX D – DISTRIBUTIONS OF VARIABLES INFLUENCING THE MODAL SPLIT

This appendix explores the effects of a multitude of characteristics on the respondents' modal split. The analysis is based on the trips and respondents resulting from the processes performed in Chapter 3. The characteristics are grouped according to the categories identified in the conceptual model for this research. First, the effects of socioeconomic characteristics on the modal split are explored followed by the influence of trip specific characteristics. This appendix finishes by an analysis of the effects of external characteristics on the modal split. The significant influence of these variables has been summarized in Table 5 in section 5.1 of this research.

SOCIOECONOMIC CHARACTERISTICS

This subsection explores the effects of socioeconomic characteristics on the modal split. First, gender and age are explored followed by characteristics which resemble the respondents' lives.

Figure 39 shows the differences in modal split between men and women. The figure shows that the share of male car drivers is higher compared to the share of female car drivers, while the share of female passengers is higher compared to the share of male passengers. Also, the share of walking is higher for men, while the share of bicycle use is higher for women.

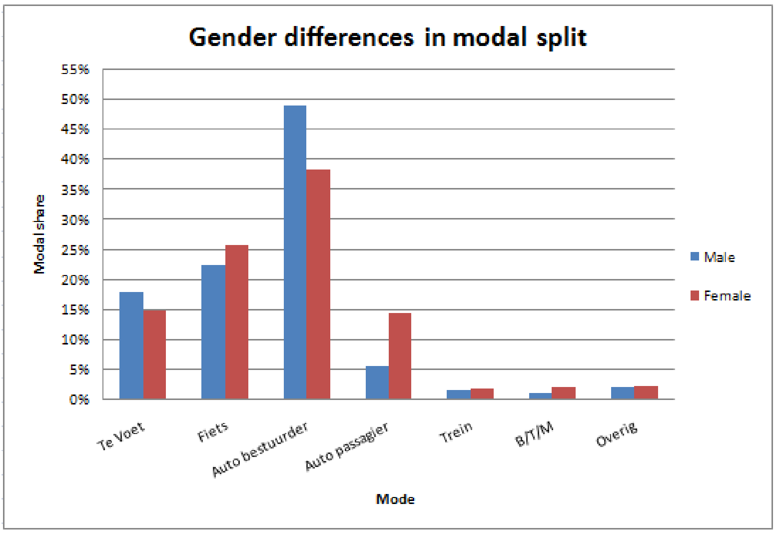

FIGURE 39: EFFECT OF GENDER ON THE MODAL SPLIT

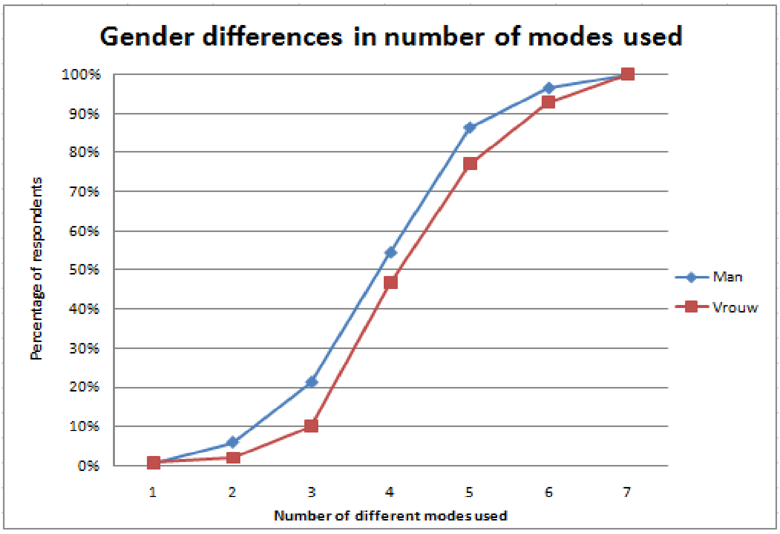

FIGURE 40: GENDER DIFFERENCES IN NUMBER OF MODES USED

Additional to the modal splits for both men and women, the amount of modes used are interesting since a larger number of modes used might be an indication of a larger intrapersonal variability in mode use. Figure 40 shows this in a cumulative manner. For instance, the dot on the 10% line and three used modes means that 90% of the female respondents have used more than three modes during the data collection period. The figure shows that women tend to use more different modes of transport compared to men, since their cumulative graph lies continuously beneath the line representing males. This difference can be explained by the fact that men use the families' private car more often (also backed up by the modal split in Figure 39) and women need to use other modes instead for their trips.

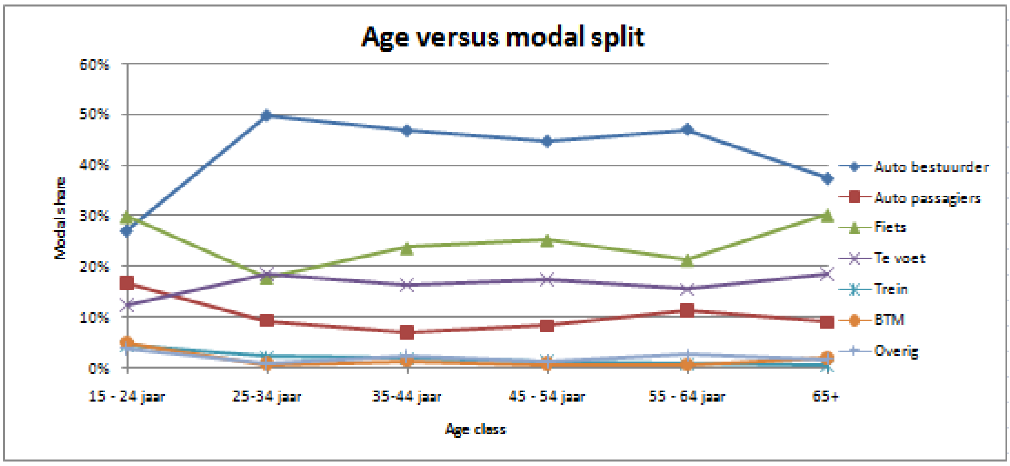

FIGURE 41: EFFECT OF AGE ON THE MODAL SPLIT

Figure 41 shows the modal split for different age categories. The use of the car increases between the first and the second age category. This is explained by the fact that the youngest age category contains students (with OV-cards) and individuals without a private car. When students graduate, the car share increases, while the shares of the other modes decrease. The share of the car decreases and the shares of the bicycle and walking increase after retirement. This is explained by the fact that the car is used a lot for work and business trips, which are not made any more after retirement.

The following characteristics in this analysis describe the respondents' lives. First, Figure 42 considers the household composition type of the respondent, where the number of respondents in that category is placed between brackets in the legend. The figure shows that the amount of persons in a household affects the modal split. An increase in the number of household members results in an increase in the share of car driver trips. The share of bicycle use and walking is lower for these respondents. A household comprising of only one single person tends to use the car relatively less often compared to households comprising of more than one member.

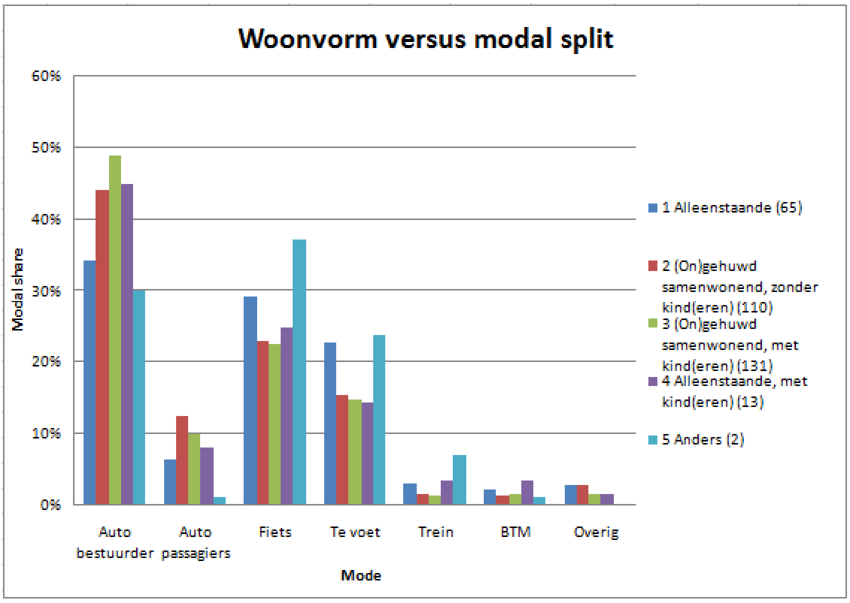

FIGURE 42: MODAL SPLIT PER HOUSEHOLD COMPOSITION TYPE

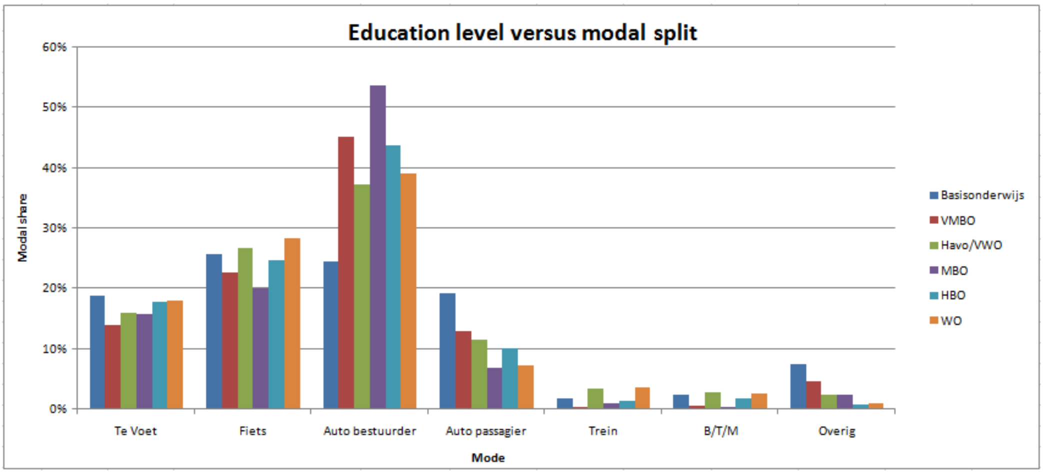

FIGURE 43: EFFECT OF EDUCATION LEVEL ON THE MODAL SPLIT

The effect of the highest completed education level on the modal split is analysed based on Figure 43. The figure shows that the education level affects the modal split. Considering completed follow-up education after secondary school, a decrease of the car driver share is noticeable and an increase of the bicycle share. Also, respondents who have completed a university degree tend to use both B/T/M and the train relatively more often in comparison to lower educated respondents.

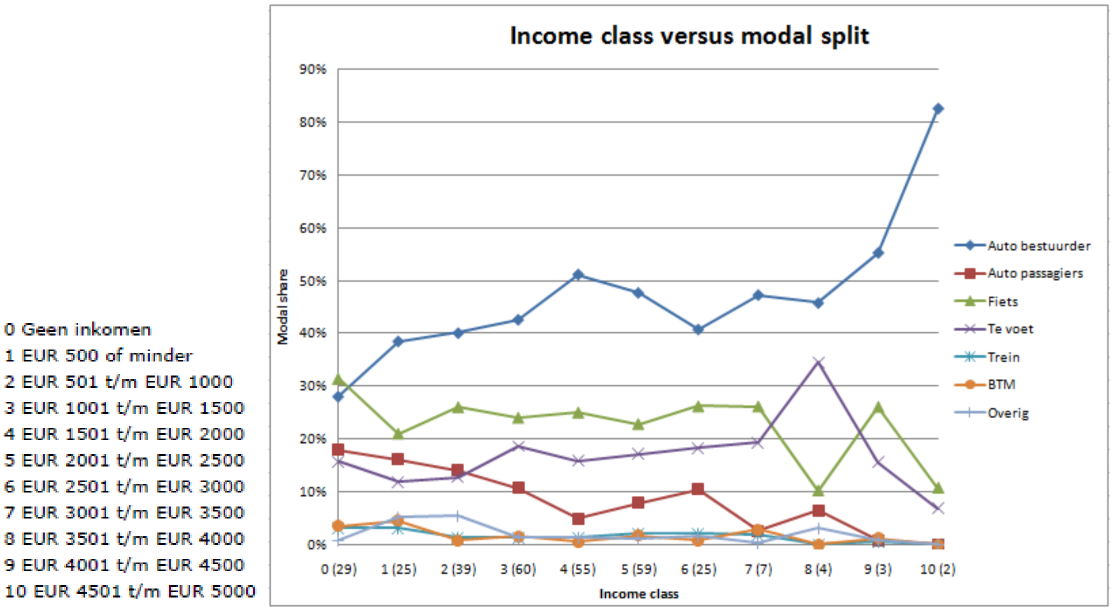

FIGURE 44: EFFECT OF RESPONDENTS' INCOME CLASS ON THE MODAL SPLIT

As stated in the theoretical framework, income influences the modal split. This effect is analysed in two different ways in this research. First, Figure 44 shows the effect of the income class of the respondent on the modal split. The values between brackets are the number of selected respondents in that income class. The figure shows that when the income increases, the share of car passenger decreases and shifts to car driver. An explanation for this phenomenon can be that respondents with a higher income are able to purchase a second car in the household and have therefore always a car at their disposal, whereas households with only one car are influenced by household interactions. Since the data for this research is collected on a person level instead of a household level and the number of cars in a household is unknown, these interactions cannot be analysed.

Following the income class of the individual respondent, the effect of the household income on the modal split is shown in Figure 45. Again, the numbers behind brackets on the X-axis represent the number of respondents in that household income class. As can be seen from the figure, the car river share gradually increases with an increasing household income. The effect is less clear compared to the increase of modal share in the previous figure. For car passengers the decrease in modal share is less evident compared to the previous figure. The effects shown in this figure matches the conclusion drawn by Sabir et al. (2008) that individuals tend to drive a car instead of using a bicycle or going on foot when the household income increases.

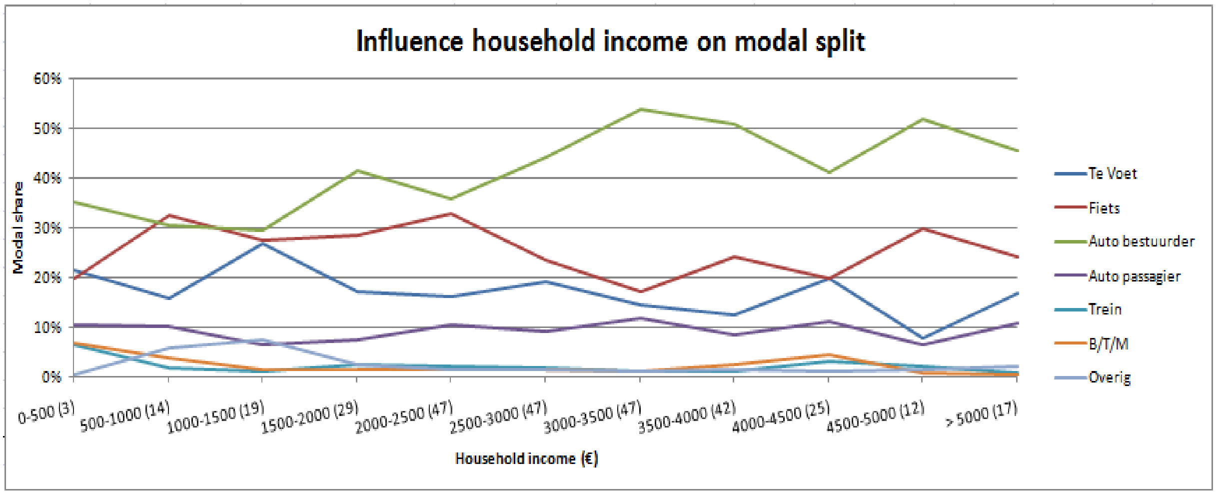

FIGURE 45: EFFECT OF HOUSEHOLD INCOME ON MODAL SPLIT

Another variable influencing the modal split is the household composition. To analyse its influence the number of children in the household was chosen as a parameter of household size. Figure 46 shows the distribution of the modal split when the household size increases and shows that car driver increases while the rest of the modes show a decrease in use. The extreme influence of five and six children in the household is caused by the fact that there is only one observation in this category.

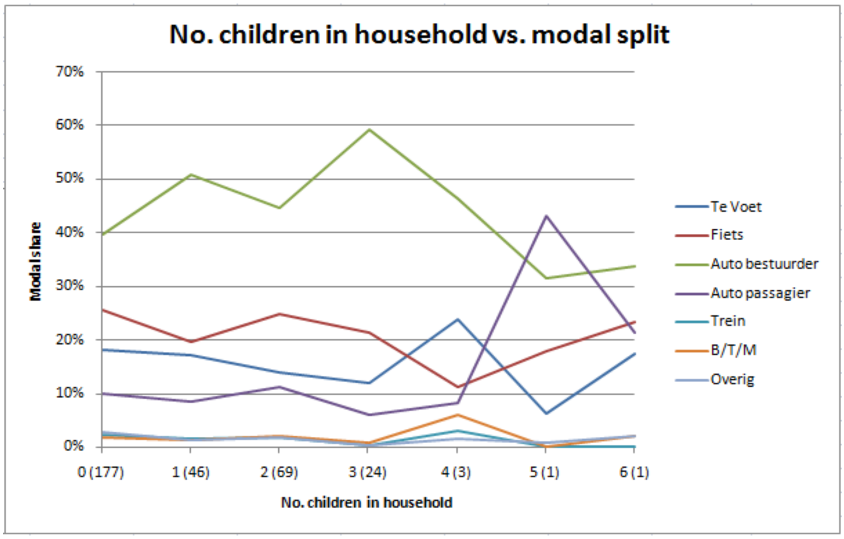

FIGURE 46: NUMBER OF CHILDREN IN THE HOUSEHOLD VS. MODAL SPLIT

Additional to the characteristics analysed above, the effect of the type of phone of the respondent is analysed. While in advance it was expected not to have any influence, Figure 47 shows that the modal split differs between the different phone users. Iphone users tend to drive the private car more often in comparison to android users.

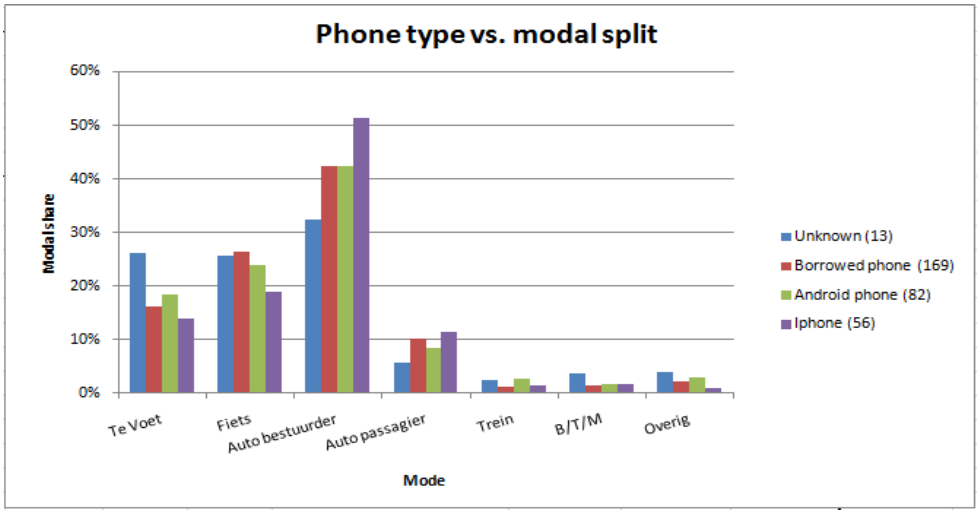

FIGURE 47: PHONE TYPE VS. MODAL SPLIT

TRIP SPECIFIC CHARACTERISTICS

This subsection explores the influence of trip specific characteristics on the modal split. First, Figure 48 shows the difference in mode use for home based and non-home based trips and shows that the share of bicycle use is higher for home based trips and that the share of car drivers is higher for non-home based trips. The differences in modal split are not high, but significant. This figure was only based on the data collected in 2013 since determining the home location of the respondents was regarded as too time consuming to perform for 2014 as well.

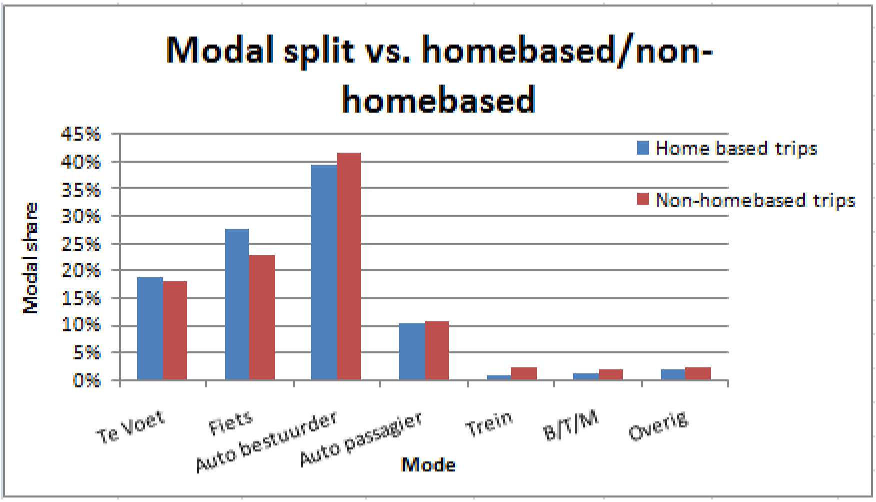

FIGURE 48: MODEL SPLIT FOR HOMEBASED/NON-HOME BASED TRIPS

Figure 49 shows the effect of the location type of the origin location on the modal split. Since the destination location showed similar modal shares only the origin is displayed here. The figure shows that, the worse accessible a location is, the more trips are made from (and to) this location by using a private car. Additionally, the share of the train decreases when the accessibility decreases.

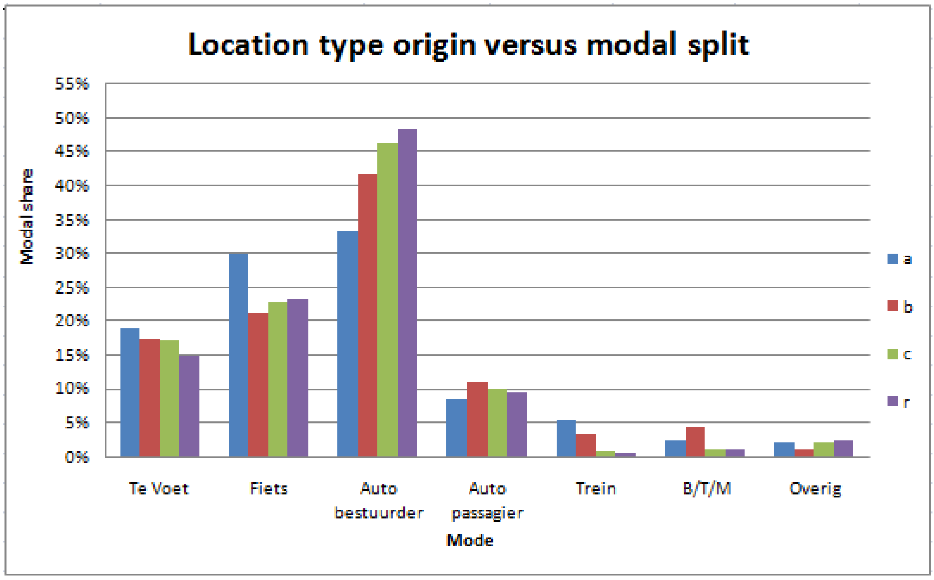

FIGURE 49: EFFECT OF ORIGIN LOCATION TYPE ON THE MODAL SPLIT

The degree of urbanisation of the respondents' residence location is another location-based characteristic available in the dataset. This characteristic is used since the city size of the respondents' residence location was unavailable for this research. Figure 50 shows the effect of the degree of urbanization on the modal split. The value of 1 is represents very urban locations and the value 5 represents rural areas. The share of bicyclists decreases and the share of car drivers increases when the urbanisation rate decreases. These effects match the argumentation made earlier on in the theoretical framework on the notion that the urbanization rate influences the modal split.

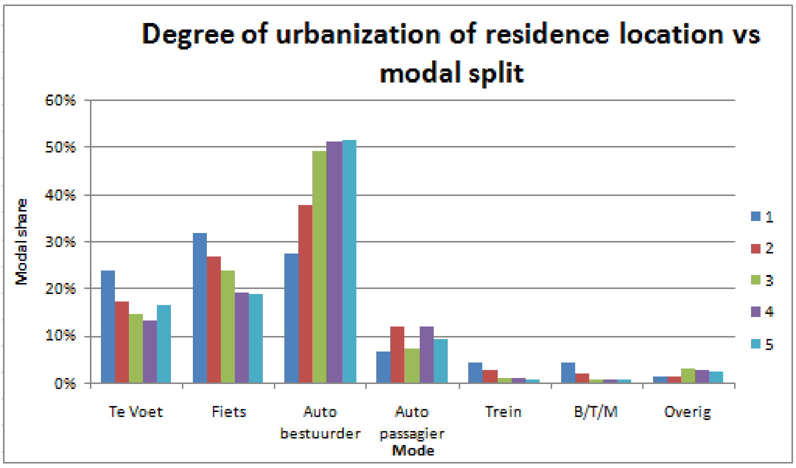

FIGURE 50: EFFECT OF URBANISATION RATE OF RESPONDENTS' RESIDENT LOCATION ON THE MODAL SPLIT

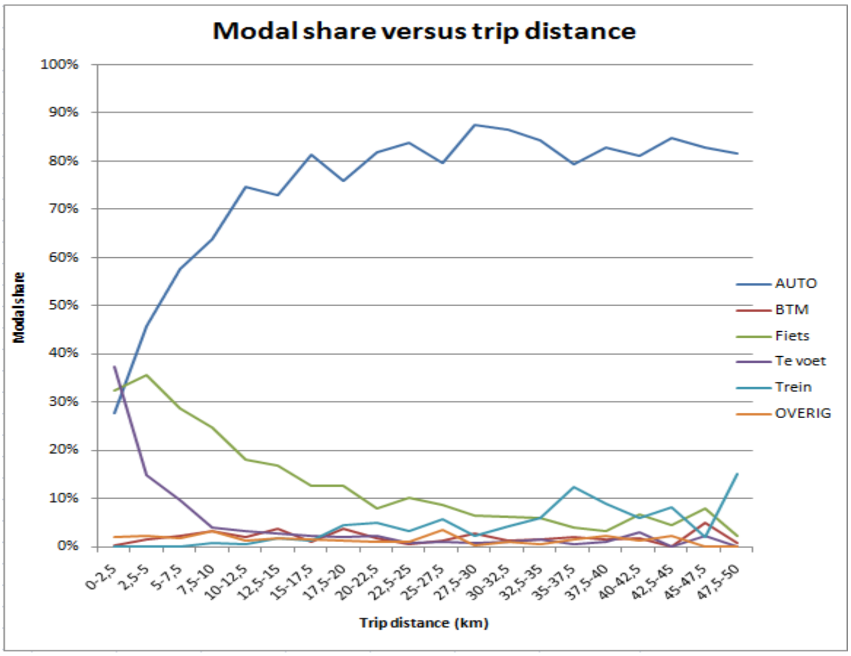

FIGURE 51: MODAL SPLIT AS FUNCTION OF THE TRIP DISTANCE

Following characteristics of the urban surrounding of the trip, Figure 51 shows the effect of trip distance on the modal split. As can be seen from this figure the share of pedestrians drops rapidly with an increase in trip distance. The share of bicycle trips decreases at a slower rate and both are compensated for by the share of the car. The share of the train starts to increase from fifteen kilometres onwards which means it is not an option for short distance trips. Most train trips are observed with a distance between 30 and 45 kilometres.

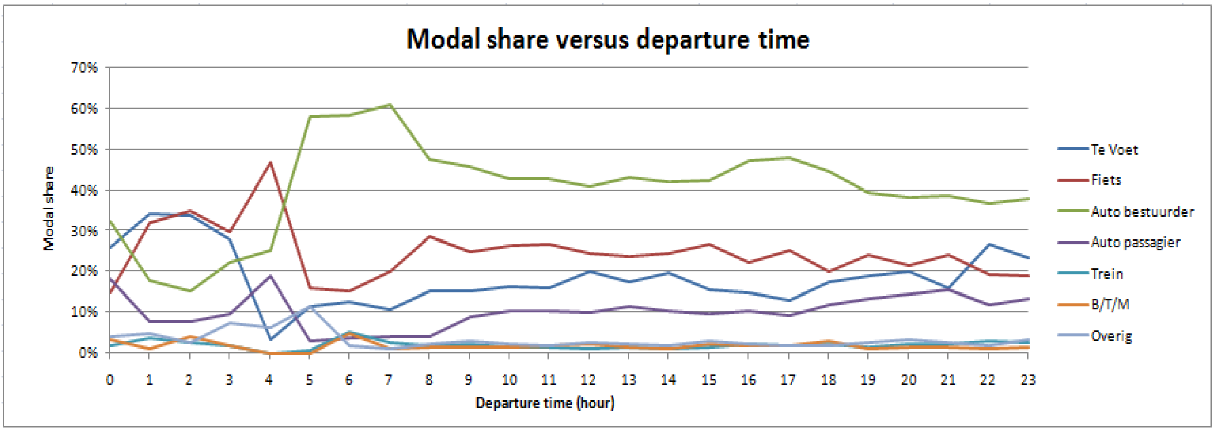

FIGURE 52: MODAL SPLIT PER DEPARTURE TIME OF THE DAY

Another trip specific characteristic is the trip departure time. Figure 52 shows the modal split observed for every hour of departure. During the morning peak hours the share of car drivers is dominant and the share of car passengers and bicyclists is low. The modal shares are fairly stable during the rest of the day for all the modes. During the night a large variability is observable. This variability is partly caused by a low number of registered trips. A few trips account for a large share for their modes here.

The timing of the trips also affects the modal split. This is shown in two different ways in the following figures. First, the difference in modal split between work days and the weekend are shown in Figure 53. This figure shows that during the weekend the share of cyclists and car drivers decreases and more car passenger trips and trips made on foot are registered. This is explained by the fact that during the weekend more people tend to make trips together (carpool) and therefore act as a passenger instead of a lone car driver to work during the week. Also, the weeks' grocery shopping is done most often during the weekend by car instead of the small groceries for which a small bag on a bicycle is sufficient.

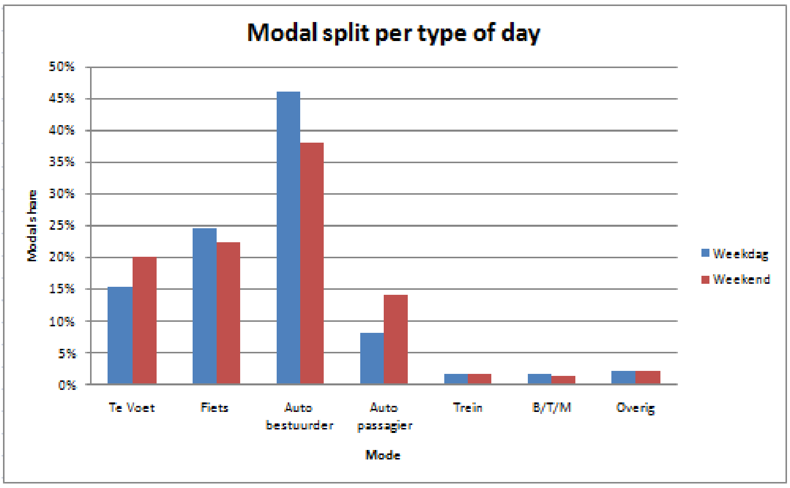

FIGURE 53: WEEKDAY/WEEKEND DAY ON MODAL SPLIT

Secondly, Figure 54 shows the modal split of the registered trips from the DMMP-panel divided per day of the week. The figure shows that the share of car drivers drops at a fairly large rate on Sundays while the shares of all other modes increase. Also, the figure shows that the share of car passengers rises during the week with a peak on Sunday.

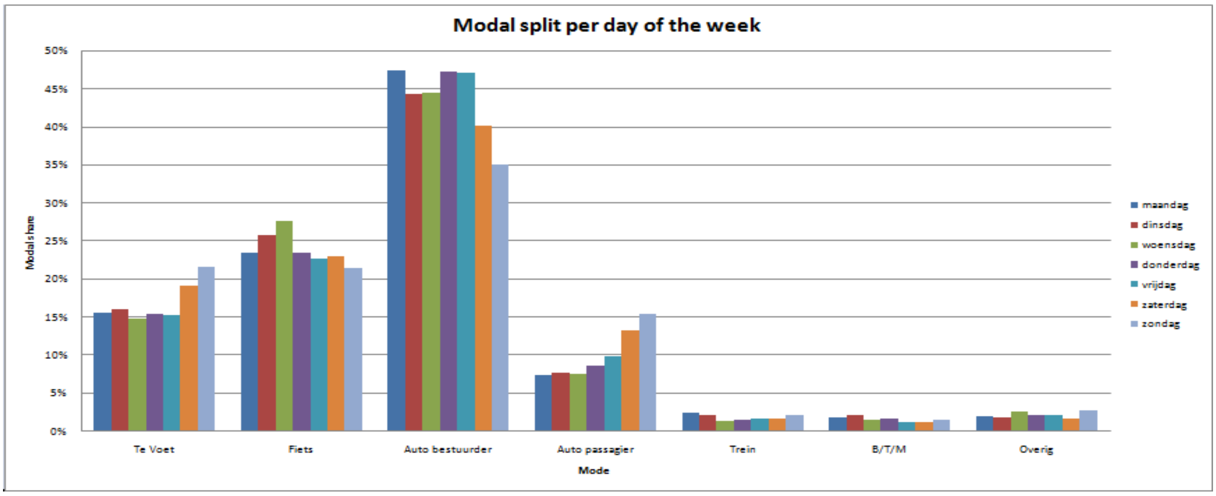

FIGURE 54: DAY OF THE WEEK VS. MODAL SPLIT

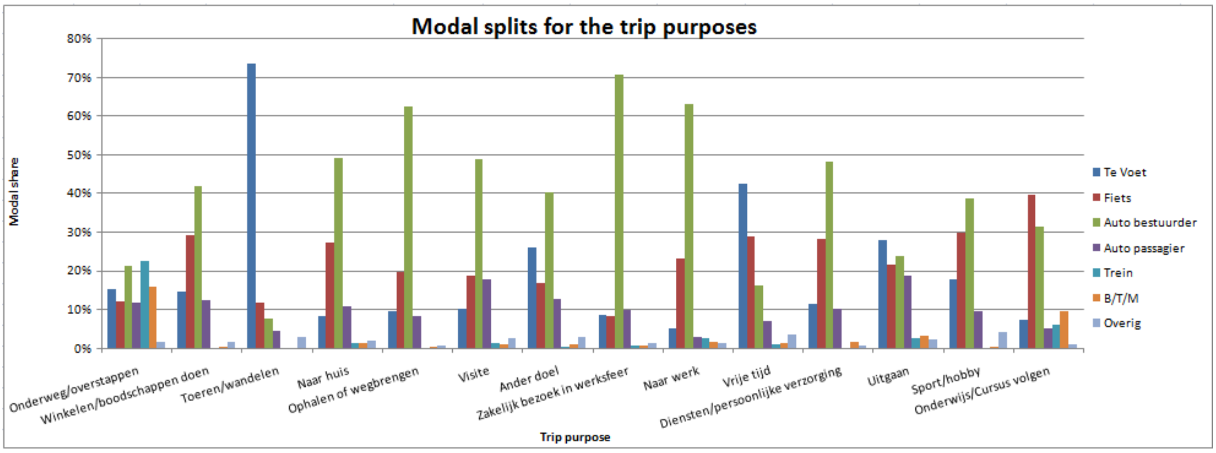

FIGURE 55: TRIP PURPOSE VS. MODAL SPLIT

The last trip characteristic analysed here is the difference in modal split per trip purpose. Figure 55 shows the modal split for all the trip purposes in the DMMP-data set. A large difference in mode use between the trip purposes exists. The car is dominant for most purposes (especially for work-, business- and pickup and return trips) but the share differs greatly. Walking is dominant for touring/hiking and 'spare time' which is logical since walking is the purpose of that trip. Educational trips are mostly done by bike but also have a high share of train and B/T/M compared to the other trip purposes, probably caused by students with a students' OV-chip card.

EXTERNAL CHARACTERISTICS

WEATHER CHARACTERISTICS

The previous sections of this appendix described the influence of personal- and trip characteristics on the modal split. This section explores the effects of weather characteristics on the modal split. The most important characteristics as identified by the theoretical framework are analysed here. Other characteristics were available from the KNMI dataset but were left out of the analysis. For instance, occurrence of snow was one of the available variables, but since there was no registration of any snowfall during both data collection periods the effect of snow cannot be analysed.

The first analysed characteristic is the effect of the amount of sunshine during the hour the trip departed on the modal split. Figure 56 shows that the effect is rather limited. When the amount of sun increases, the share of bicycle trips increases and the car driver share decreases. The shares of the other modes are not influenced by the amount of sun.

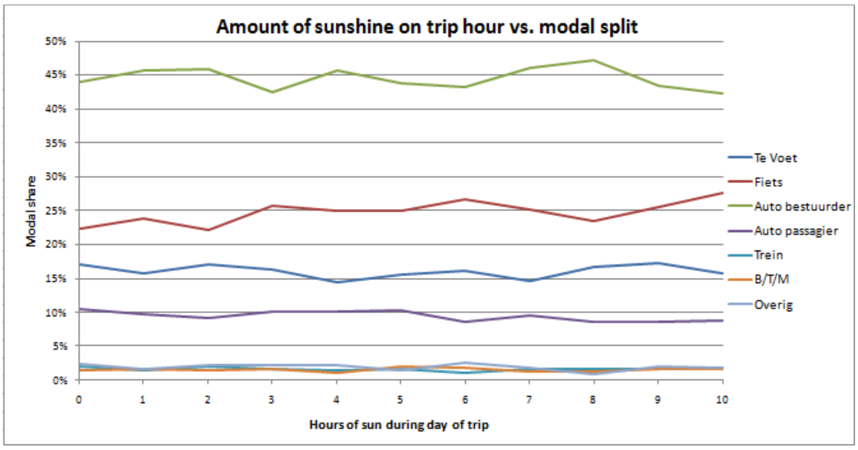

FIGURE 56: INFLUENCE OF HOURS OF SUN ON TRIP DAY ON THE MODAL SPLIT

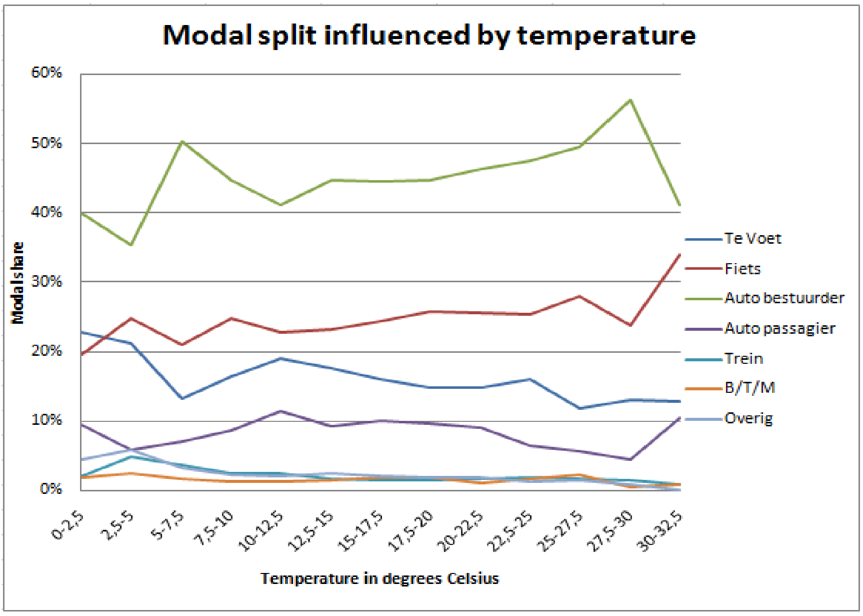

FIGURE 57: INFLUENCE OF TEMPERATURE ON MODAL SPLIT

The second weather characteristic analysed is the influence of temperature on the modal split. Figure 57 shows that the variability for both the lowest and highest temperature is large. This is due to the fact that there were hardly any trips registered for these temperatures and one extra trip could therefore increase the modal share quite strongly. Looking at the temperatures between the 7.5-10 degrees and 25-27.5 degree category, an increase in modal share for both cycling and car driver and a decrease for walking and car passengers is observed. An explanation for these values is that the willingness to cycle increases with an increasing temperature. The results match the results from the research performed by Sabir et al. (2008) as described in the theoretical framework.

Figure 58 shows how the sky condition can affect the modal split. The value 0 means a clear sky without clouds and a value of 8 means that the sky is fully covered with clouds. The KNMI divided the values in-between in eights. While the influence is rather limited, an increase of car drivers with increasing sky coverage while the share of the bicycle decreases is observed.

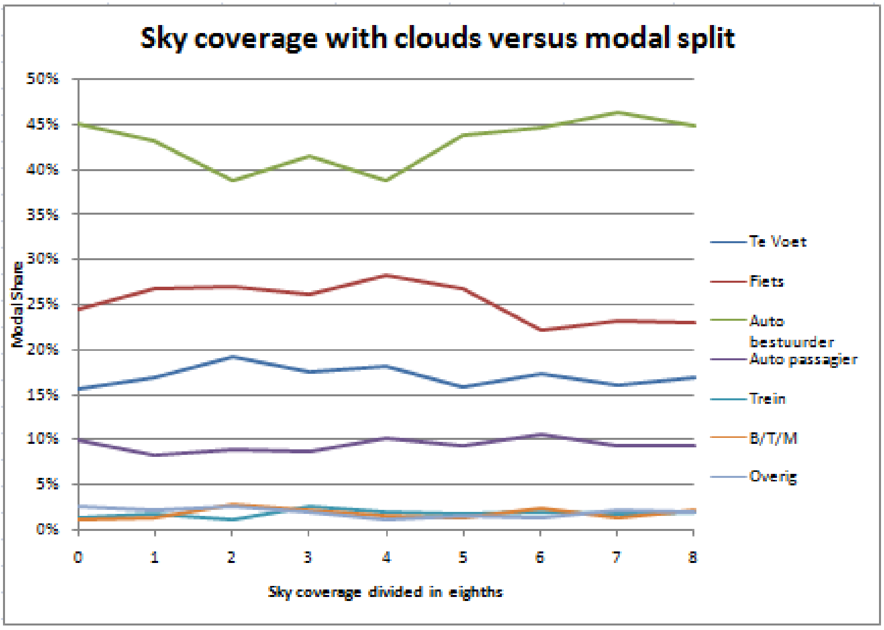

FIGURE 58: EFFECT OF SKY CONDITION ON THE MODAL SPLIT

The effect of rain is analysed in two different ways. First, Figure 59 shows the difference between rain registered and no rain registered at the hour of departure. Secondly, Figure 60 includes the amount of rainfall during the hour of the trip departure. Figure 59 shows that the share of car drivers increases with 6% at the cost of cyclists and a small share of pedestrians, when it starts to rain. The shares for the other modes are comparable. The effect shown in Figure 59 matches the results from the research of Sabir et al. (2008) as described in the theoretical framework.

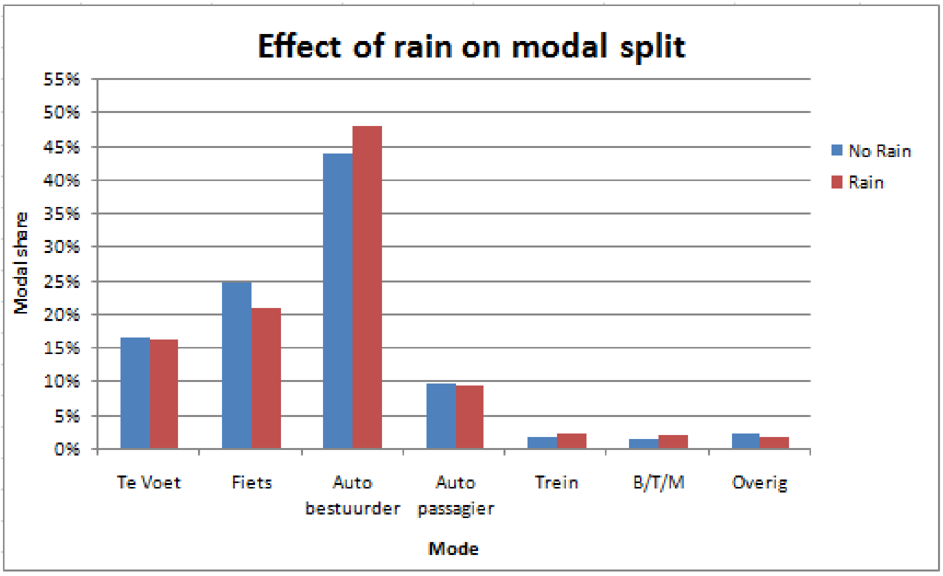

FIGURE 59: EFFECT OF RAIN/NO RAIN ON THE MODAL SPLIT

Figure 60 shows the effect of the amount of hourly rain on the modal split. Since the numbers of registrations of more severe rains were very low, these values were left out of the analysis since no sound conclusions can be drawn from low registrations regarding the modal split. Collecting more longitudinal data can overcome this problem in future research, when more heavy rain situations are registered. The figure shows that an increase in hourly precipitation does not lead to a decrease in the modal share of bicyclists. This contradicts with the research of Sabir et al. (2008) as described in the theoretical framework.

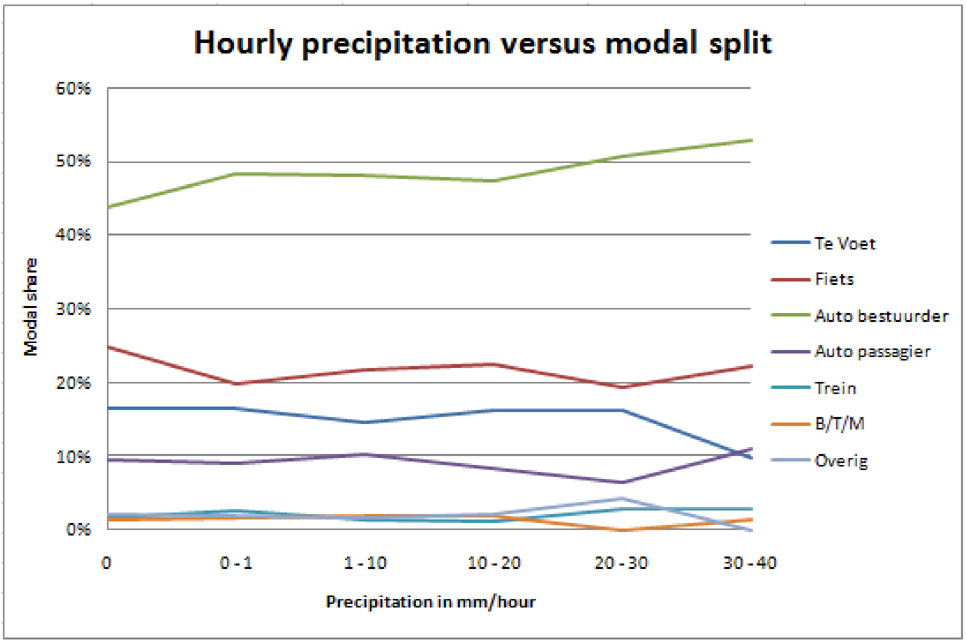

FIGURE 60: EFFECT OF THE AMOUNT OF HOURLY RAIN ON THE MODAL SPLIT

The last weather characteristic analysed in this research is the effect of wind speed on the modal split. Figure 61 shows fairly stable shares for the modes with an increase of wind speed. The share of bicyclists decreases a little, which is compensated for by car passengers. This is different in comparison to the effects stated in the theoretical framework where both car driver and walking increased while bicycle use decreased.

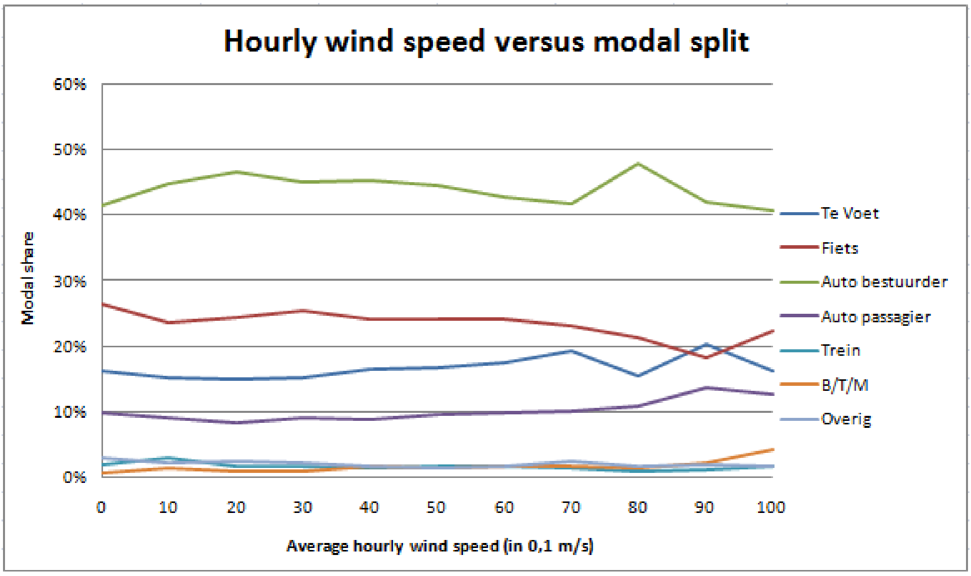

FIGURE 61: EFFECT OF WIND SPEED ON THE MODAL SPLIT

Appendix E – Dominant mode use

This appendix describes the dominant mode use analysis mentioned in Section 5.2. First, the modal split was determined for all individuals in the dataset. There was no specific time span for which the modal split was determined (as was the case for Buehler & Hamre (2013)), all trips in the dataset were used. The mode with the largest modal share was determined to be the dominant mode with the accompanied 'dominancy' (the share of trips with the mode which was used the most). Figure 62 shows the distribution of the dominant modes where the X-axis represents the share of trips with the dominant mode. As observed, the car is the most common dominant mode used.

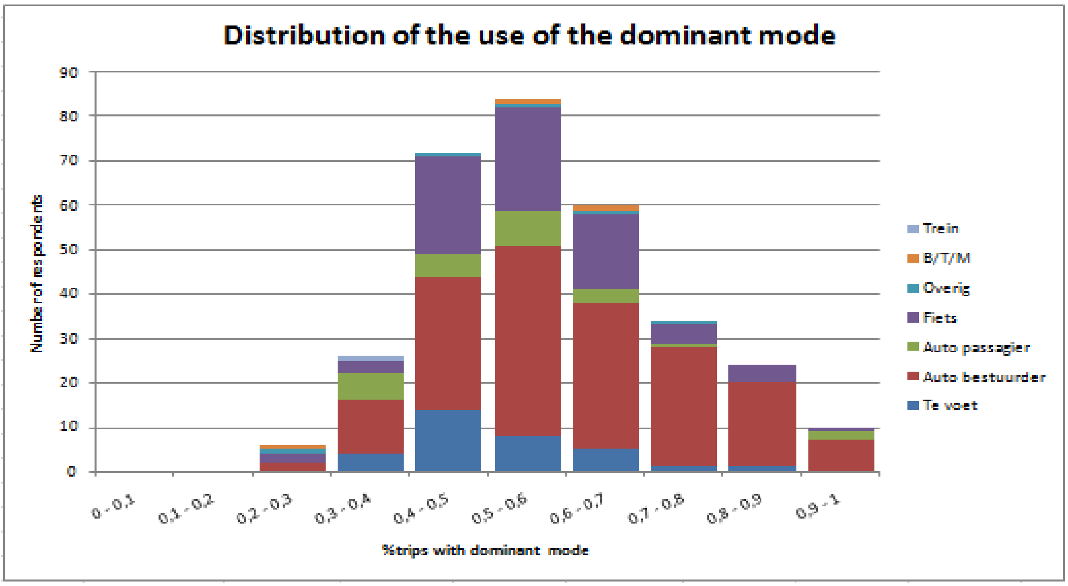

FIGURE 62: DISTRIBUTION OF THE DOMINANT MODES USED BASED ON NUMBER OF TRIPS

The analysis described above was solely based on the number of trips registered by the respondents. Since trip distance is an important influential characteristic for modal choice this same analysis was performed based on the distance travelled. For all modes the amount of kilometres travelled was calculated followed by the distance-by-mode shares. The mode with the largest distance share was determined to be the dominant mode with respect to trip distance. Figure 63 shows the distribution of dominant modes based on travelled distance.

When comparing both figures one can observe a large difference. On average the shares of the dominant mode are far larger compared to the trip level distribution (median value 56,1% versus 69,6%). Also, the train is used more often as dominant mode when the trip distance is considered. This means that the train is not used most often, but when it is used the trip distances are large. Additionally, the shares of the car show a large increase. This means that while the share of trips is lower, the trips account for a large share of the distance travelled by the respondents.

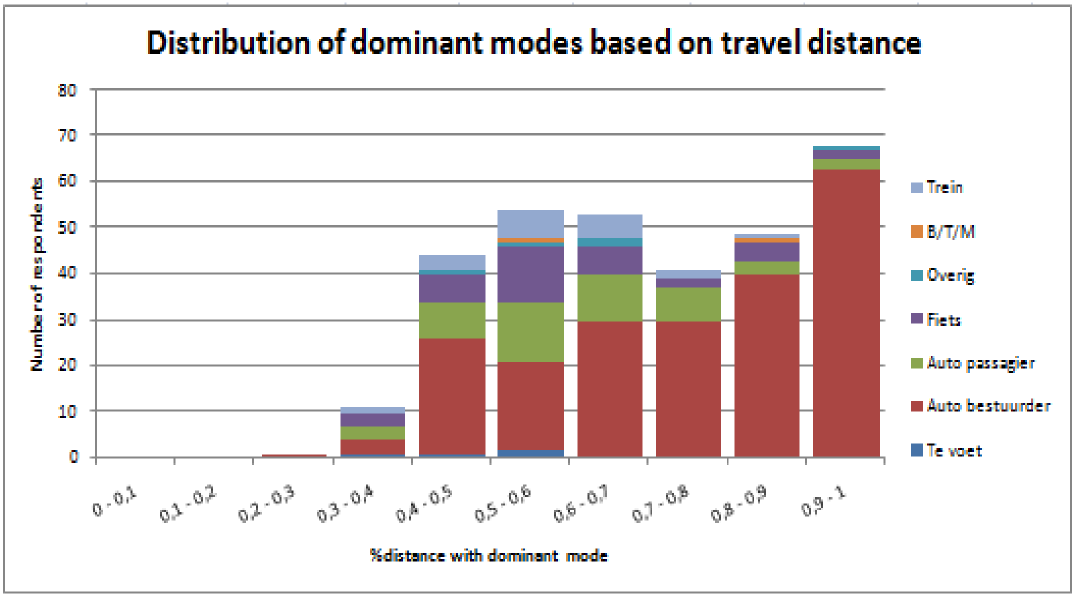

FIGURE 63: DISTRIBUTION OF THE DOMINANT MODES USED BASED ON TRAVELLED DISTANCE

APPENDIX F – KOLMOGOROV-SMIRNOV TEST

The Kolmogorov-Smirnov test (K-S test) is a nonparametric test of the equality of continuous, one-dimensional probability distributions. The two-sample K-S test is one of the most useful methods for comparing two samples since it is sensitive to differences in both the shape and the location of empirical cumulative distribution functions of two samples (Lin, Wu, & Watada, 2010). It tests whether two independent samples have been drawn from the same population. The statistic focuses on the largest of the observed deviations between the empirical cumulative distribution functions of two samples. Rejection of the null hypothesis means that the two samples are from different populations. The null hypothesis is rejected at level α if

$$D_{n,n'} > c(\alpha)\sqrt{\frac{n+n'}{nn'}}$$

With $D_{n,n'}$ is the maximum vertical deviation of the two cumulative distributions, n and n' representing the sample sizes of both samples and $c(\alpha)$a constant with its value based on the level of significance α. A confidence level of 95% was chosen for this research resulting in $c(\alpha) = 1.36$ (Wikipedia, n.d.).

APPENDIX G – INDIVIDUAL MIX-VALUES

This appendix shows the MIX-values for individuals using different data collection time spans. Between two and four weeks of data used, some differences exist (see Figure 64). These differences are explained by the fact that the first two weeks of data were collected in 2013 and the second two weeks in 2014. Individual circumstances changed which had an impact on the MIX-value. For instance, an elderly cyclist from 2013 registered moped trips instead of cycling trips in 2014. Because of this the cycling category and OV&other-category received merely the same frequencies after the second wave, resulting in a higher MIX-value. Overall, the intrapersonal variability in mode choice behaviour remained more or less the same.

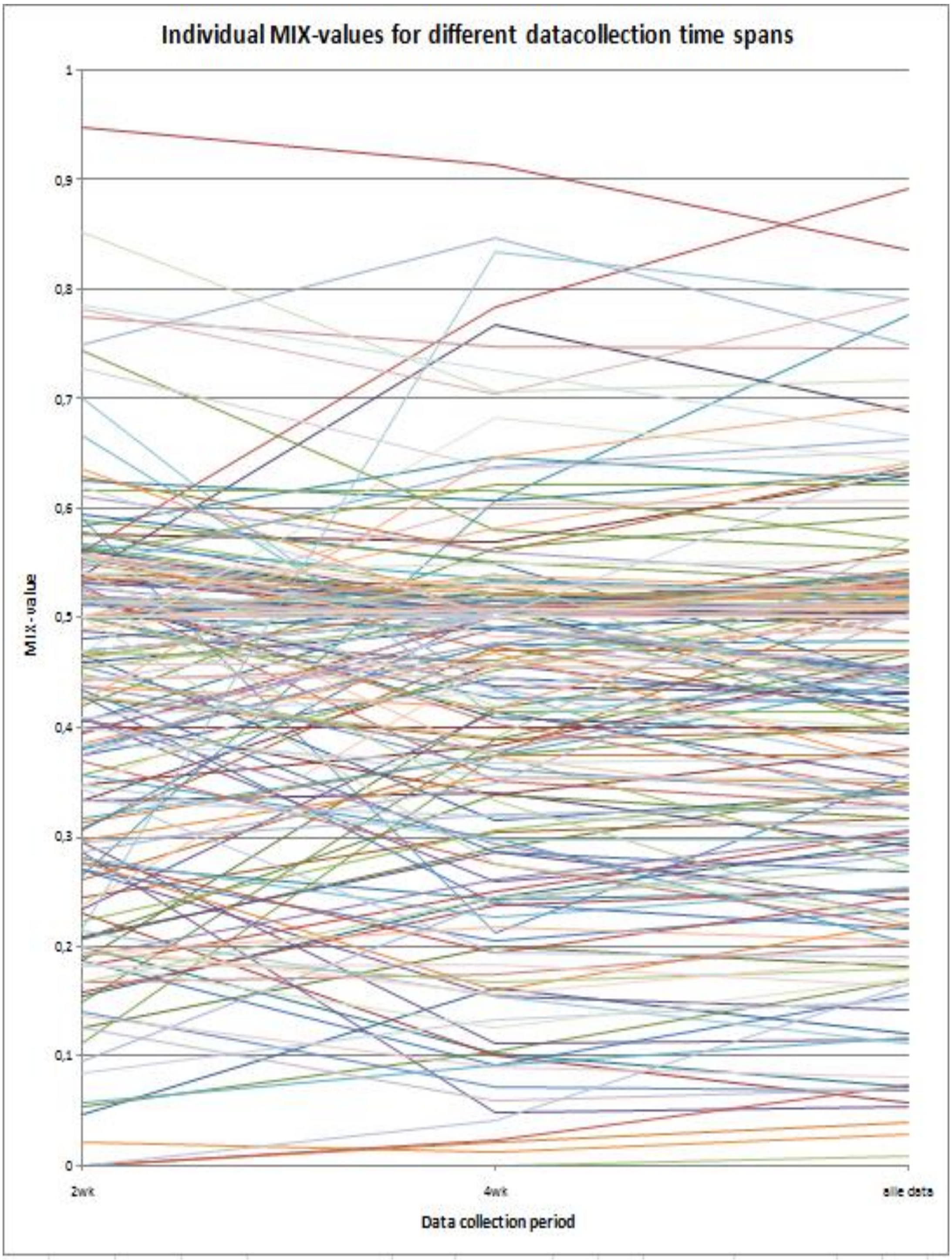

FIGURE 64: INDIVIDUAL MIX-VALUES FOR DIFFERENT DATA COLLECTION TIME SPANS

APPENDIX H – BIOGEME SCRIPT

This appendix shows the script of the final model used to estimate the significant parameters of the model. No code was included to account for the panel effect of the data. BIOGEME can account for this panel effect by specifying observations belonging to individuals. The inclusion of the panel effect resulted in computational problems for BIOGEME and the program kept on crashing after only a few iterations time after time. To get estimated parameters the choice was made not to specify the panel effect in the script.

```
// MNL with three alternatives
// Bike|Car|PT

[ModelDescription]
// MNL based on selected trips
// Travel time and distance used
// Shopping significant as trip purpose
// Rural significant for bike and car
// Slightly rural significant for bike and car

[Choice]
CHOICE

[Beta]
//Name  Value   LowerBound      UpperBound      Status (0=variable, 1=fixed)
ASC_Car 0       -1000   1000    1
ASC_OV 0        -1000   1000    0
ASC_Bike 0      -1000   1000    0
B_Time 0        -1000   1000    0
B_Dist 0        -1000   1000    0
B_Shopping 0    -1000   1000    0
B_Rural 0       -1000   1000    0
B_SL_Urban 0    -1000   1000    0

[Utilities]
//ID    NameAvail       Expression
1       A1Fiets AVFiets ASC_Bike * one + B_Time * Fietsreistijd + B_Dist * Distance
                                + B_Shopping * Shopping + B_Rural * Rural
                                + B_SL_Urban * SL_Urban
2       A2Auto  AVAuto  ASC_Car * one + B_Time * Autoreistijd + B_Dist * Distance
                                + B_Shopping * Shopping + B_Rural * Rural
                                + B_SL_Urban * SL_Urban
3       A3OV    AVOV    ASC_OV * one + B_Time * Ovreistijd

[Expressions]
one = 1

[Exclude]
Distance = -999

[Model]
$MNL
```