Technical University of Munich – Professor for Modeling Spatial Mobility

Prof. Dr.-Ing. Rolf Moeckel

Arcisstraße 21, 80333 München, www.msm.bgu.tum.de

Master’s Thesis

Development of a Microscopic Activity Sequence Modeler

Author:

Usman Ahmed

Supervision:

Prof. Dr.-Ing. Rolf Moeckel

Dr. Ana Tsui Moreno Chou

Dater of Submission: 29-12-2017

Abstract

To improve the travel demand models, activity-based modeling approach has now become more popular as compared to conventional trip-based modeling approach. In a trip-based model, trips are considered independent of each other. In reality however, trips are linked to each other and are interdependent. The sequence of trips to perform activities is termed as activity sequence or trip chaining and is an integral part of activity-based modeling approach.

This thesis aims to model trip chains at individual level. For this purpose, trip chains were generated from national household travel survey conducted in 2008 (MiD 2008). Trip chains were modeled in two steps. First, multiple correspondence analysis (MCA) was applied to study the relationship among trip chains and explanatory variables (socioeconomic variables). Secondly, a prediction module has been developed based on MCA model, which is able to predict trip chains for individuals. Model calibration has also been performed. The trip chain assigned to the person varies based on the probabilities calculated. Therefore, it represents the heterogeneity in trip making behaviour which is found in reality.

NOTE: The analysis has been updated after the oral defence and some figures have been removed due to possible copyright infringements. Munich, May 2018

Acknowledgements

I would like to express my sincere gratitude to my supervisor and mentor, Prof. Rolf Moeckel, for supervising my thesis. I deeply appreciate the guidance from my co-supervisor Dr. Ana Tsui Moreno Chou. Their knowledge and guidance resulted in completion of the work within stipulated time frame.

I would also like to extend my gratitude to Dr. Carlos Llorca Garcia and Hema Sharanya Rayaprolu for their suggestions and valuable comments. To Dr. Matthew Bediako Okrah and all my colleagues from the group Modeling Spatial Mobility for their kindness and support.

I am very grateful to my parents for their prayers and blessings and my wife Maria Usman for her constant support and motivation throughout my thesis.

Lastly, many thanks to my friends Eeshan Bhaduri, Mayookh Sikka, Uzair Sharif and Wajhi Naqvi for their help and encouragement.

Declaration

I hereby confirm that the presented thesis work has been done independently and using only the sources and resources as are listed. This thesis has not previously been submitted elsewhere for purposes of assessment.

Munich, December 29th, 2017

____

Usman Ahmed

Contents

Note - Appendix A.0, B.1.2 and B.1.3 are available in the cd only.

List of Figures

List of tables

Abbreviations

MiD	Mobilität in Deutschland
CA	Correspondence Analysis
MCA	Multiple Correspondence Analysis
PCA	Principal Component Analysis
RMSE	Root Mean Squared Error
TC	Trip Chain

1 Introduction

Travelling is an essential human need. People travel to fulfil their social, economic and other necessities of daily life. Travel demand has grown considerably in the past decades due to increase in the mobility needs of people. It affects the transportation system causing congestions on road network, overcrowded public transport and has severe environmental effects. To counter these problems, new infrastructure is built, or existing infrastructure is optimized. Such projects require heavy investments; therefore, policy makers are keen to know the travel behaviour of population in future. Urban travel forecasting models always played a crucial role in supporting infrastructure investment decisions.

Since 1960s, these models are based on classical four-step modeling approach. It is an aggregate approach based on zoning system of the study area. While trip based modeling approach has many advantages, it has limited application for modern policy issues. Gradually, the mobility patterns have become more complex which has motivated planners to build more sophisticated models. This results in the development of disaggregate modeling approach. In this approach, modeling scale is microscopic up to the level of households and even persons. It considers the travel decisions made by persons and the combined households. The underlying concept is that the demand to fulfil activities induces demand for travel, hence the term activity-based modeling is commonly used to reflect disaggregate modeling approach. A significant amount of research has been done in this domain and is still in experimental phase (Ort and Willumsen, 2011). The concept was developed under research environment in 1970s and improvements were being made in 1980s (Bowman and Ben-Akiva, 2001).

The current study focuses only on a part of the complete activity-based modeling framework and that is activity sequencing or trip chaining. Trip chains are conventionally modeled in a complex way which is described in more detail in chapter 2, while the current study presents a simplified modeling approach. In literature, the definition of trip chains or activity sequence varies. The current study defines a trip chain or an activity sequence as a sequence of activities performed during a day (e.g. home to work to shopping to home) regardless of the duration and location of activities. The German household travel survey (MiD) for the year 2008 was used to build the model. Besides using conventional approaches, a multivariate data analysis technique known as multiple correspondence analysis (MCA) is applied, which is probably the first application of this technique in trip chaining. The major goals of the study are:

1. To develop a model that can predict the activity sequence of an individual based on socioeconomic characteristics.
2. The model should be able to represent heterogeneity in travel pattern for the similar definition of individuals.

The rest of the report is structured as follows:

- Chapter 2 reviews the research and development in travel forecasting models.
- Chapter 3 briefly explains the literature gap found and the research approach adopted to fulfil that gap.
- Chapter 4 explains in detail, the process of generation of trip chains from the household travel survey.
- Chapter 5 describes the search for appropriate method and the conceptual framework of applied method.
- Chapter 6 lists the various aspects of model building, selection of variables, model estimation results, a framework for prediction of trip chains from the model and finally model calibration and results.
- Chapter 7 concludes the report with the summary of findings, limitations of study and recommendation for possible future work.

2 Literature Review

The two most common urban travel modeling approaches are trip-based and activity-based. This chapter describes the basic concepts of the two approaches and the reasons for the shift towards the activity-based approach. The research and development in the latter approach, specifically trip chaining process is also described.

2.1 Trip-based modeling approach

The most common paradigm for urban travel demand modeling is the classical four-step model. It is a trip-based model and was developed in the 1950s. The first application of the four-step model was in the Chicago Area Transportation study (Mcnally, 2007). Since the 1960s, it has been a general practice to use the classical four-step model, despite major advancements and improvements in modeling techniques (Ort and Willumsen, 2011). The travel demand is estimated in four steps which are listed as follows:

1. Trip Generation
2. Trip Distribution
3. Mode Choice
4. Assignment

The approach starts by dividing the area under consideration into zones which are normally termed as traffic analysis zones (TAZs) and the transportation network system of the study area is defined. The first step, i.e. trip generations, estimates the number of trips that are produced, and the number of trips attracted to each zone. The next step trip distribution, distributes the trips from each zone to all other zones. It provides an origin-destination matrix of trips. The third step mode choice, determines the travel mode used to perform trips. The final step of assignment, assigns the trips by the respective mode to the transportation network.

The sequence of steps defined above is commonly practiced. However, it is not the only sequence used. In the past, some studies have used mode choice before trip distribution as well (Ort and Willumsen, 2011).

As the classical four-step model is a trip-based model, the unit of analysis is the individual trip. For example, the trip purposes are home-based work, home-based other etc. For each trip purpose, a separate model is built, but for some analysis the last step i.e. assignment could be done simultaneously.

2.2 Changing paradigm from trip-based to activity-based modeling

Conventional four-step modeling method is still the most dominant travel demand modeling approach. However, the nature of the approach has some shortcomings as well.

Trip-based approach uses aggregate level of analysis by using zoning system. The study area is divided into smaller areas known as zones. Trips are generated and distributed among zones. In this way socioeconomic characteristics of individuals and households are represented in a very limited fashion. Therefore, the model is not able to represent shifts in long-term socioeconomic characteristics of population. Moreover, trip-based models are not able to predict the results of certain travel demand management policies such as congestion pricing (Pinjari et al., 2011). These models are not simulation-based models, so they are unable to simulate very fine details. They are not able to model individual behaviour of travellers.(Mladenovic and Trifunovic, 2014), (Pinjari et al., 2011).

In a trip-based model, the time-of-day modeling of trips are not performed or are performed in a limited way. Moreover, trips are considered individually with a limited number of trip purposes. The relationships between trip purposes are not defined. This way it shows that households plan trips independently, whereas in reality, households determine activities for whole day and plan trips accordingly. Therefore, the effect of trip chaining is missing and it fails to model combined travel time and cost of the whole trip (Mladenovic and Trifunovic, 2014). Furthermore, trip-based approach can model different modes for different trip purpose, while it is very likely that the trips are performed sequentially using the same mode. This behavioural aspect of human decision making is missing in trip-based modeling approach (Bhat and Koppelman, 1999).

The classical four-step model has been used for a long period of time. However, advancement in modeling techniques and a need to address more complex problems have paved way for a new and more sophisticated modeling system. It has also been made possible with the advancement in the computing power in the recent era. The activity-based modeling approach can handle the shortcomings of conventional trip-based models. Activity-based models are microscopic in nature, so they are much more detailed than the trip-based models. A detailed overview of activity-based modeling approach is presented in the subsequent sections.

2.3 Activity-based modeling

Activity-based modeling approach considers travel as a ‘derived demand’ to fulfil certain activities at different locations, in contrast to travel for the sole purpose of travelling. The activity-based modeling approach, considers tours or activity schedules instead of individual trips. It models the time, duration, location and mode of activities. It also keeps into account the individuals’ constraints over time and cost. Figure 2.1 shows trips, tours and daily activity schedules in a space time diagram. It shows how the trips are linked to each other to form tours. A tour is a trip chain that starts and ends at the same location. Tours combine to form day activity schedules (Ort and Willumsen, 2011). The day activity schedule is termed as activity sequence or trip chain for the whole day.

Figure removed due to possible copyright infringements

Figure 2.1 Trips, tours and activities (Ort and Willumsen, 2011)

In the activity-based approach, activities are generated, the assignment of destination, mode and finally route for every activity is determined. These are determined considering space-time constraints (Castiglione et al., 2015). These activities along with attributes are linked to form travel pattern of a person.

The activity-based modeling is a broad modeling system. It consists of various sub-models such as daily activity pattern model, destination choice models, time of day models, mode choice models etc. The objective of the current thesis is to model daily activity sequence of the population. Hence, it is a component of a big modeling framework.

One of the pioneering work on activity-based modeling was the study of the pattern of activity participation from Chapin in 1960s, but without considering spatial context of activities (Pinjari et al., 2011). Later, Hägerstraand, (1970) studied activity participation incorporating spatial and temporal aspects of activities. He introduced the system of constraints namely capability, coupling and authority constraints to pursue certain activities. According to him, individuals are bound by time and space which he calls as time-space prism. Jones (1977) tried to model activities with regard to their spatial and temporal context.

The activity-based approach gained much significance from the 1980s. Kitamura (1988) gave an overview of the developments on activity-based approaches in the 1980s. The method received extensive awareness. According to him, "*Progress has been made at a respectable rate with new emphasis on multi-day behaviour and dynamic aspects of activity and travel behaviour"* (Kitamura, 1988). He also pointed out the reasons for its limited application during that period. The same attention to activity-based modeling continued in the 1990s. C. R. Bhat and Koppelman (1999) gave an overview about activity-based modeling research in that era. They mentioned that due to new requirements posed at that time such as Clean Air Act Amendments (CAAAs) to which Transportation control measures(TCMs) was a crucial part, demanded in depth study of travel behaviour and became one of the reasons for the shift from trip-based approach to activity-based approach.

Due to microscopic level of analysis, activity-based models can represent shifts in long-term socioeconomic characteristics of population and are able to evaluate pricing scenarios. Travel behaviours of individuals are explicitly modeled. Microsimulation is yet another advantage of this approach. Activity schedules solve the problem of non-home-based trips. Combined travel time and mode of the trip can be modeled which provides better understanding of the travel behaviour. Detailed performance matrices can be produced which supports equity analysis. Moreover, activity-based approach can also produce trip-based measures. (Castiglione et al., 2015).

Till today substantial amount of work has been done on activity-based modeling. Some of the activity-based models include CEMDAP (Comprehensive Econometric Micro simulator for Activity-Travel Patterns), Sacramento Activity-based Model, FAMOS (Florida Activity Mobility Simulator),STARCHILD, SMASH, ALBATROSS (A Learning-BAsed TRansportation Oriented Simulation System etc. (Pinjari et al., 2011).

The activity-based approach also has some disadvantages. It requires detailed data including survey data such as household travel survey, employment and land-use data, synthetic population and network performance indicators. Model application also require efficient hardware and software (Castiglione et al., 2015). This modeling approach is progressing with the advancement in the field of behavioural sciences, but it will remain in research until better models are implemented first in research environment before application in the industry (Ort and Willumsen, 2011).

2.4 Activity sequencing

Activity sequencing or trip chaining is a part of a broader activity-based modeling system. Activity sequencing involves placing activities in a sequence to form a complete activity participation pattern of an individual. The process of activity participation could be quite complex as it contains a lot of behavioural constraints. (Heggie and Jones, 1978) defined four response domains for models of travel demand. These domains include independence, spatio-temporal linkages, inter-personal linkages and full interdependence. These domains affect the individual's choice of activity participation as well as trip making behaviour.

(Adler and Ben-Akiva, 1979) gave a theoretical and empirical model of trip chaining behaviour. They proposed a theory to model non-work travel behaviour of households. It consists of a utility-based theory to select a travel pattern among all available travel patterns. The travel pattern was described by the number and attribute of destinations, mode used and the number of tours. The alternative set of travel patterns used was obtained by the household activity survey. The utility was modelled as a function of scheduling convenience, travel times and costs, attributes of choice destination and the socioeconomic characteristics of households. Multinomial logit model was applied to estimate choice probabilities.

Later on, (Golob, 1986) used a nonlinear canonical correlation analysis to study the relationship between home-based trip chains and socioeconomic characteristics of the person. He identified 20 different chain types and used them as dependent variable, while independent variables used were income, age, gender, household lifecycle, working hours, household car ownership and residential location. Three categories of travellers were used namely workers, students and others. Spatial and temporal aspect of the activity was not modeled. He found out that gender, working hours and lifecycle were important explanatory variables for worker’s segment, while for students’ segment, age, lifecycle and income had the most explanatory effect and lastly for the others segment, lifecycle followed by age and income were more effective in explaining trip chaining behaviour.

(Eric I Pas, 1984) studied the effect of sociodemographic characteristics on trip chains or activity-travel patterns. They were defined by the number, type of stops along with the time of day the stop was being made. He applied cluster analysis to identify five travel-activity patterns. The five-step procedure is described in (Eric I Pas, 1983). According to the procedure, travel-activity patterns were described first, then a similarity index was derived for comparison of alternatives, followed by the transformation of the patterns to the Euclidean space, followed by the application of cluster analysis to group similar trip chains and finally different clusters were recognized. A representative pattern of each cluster group was selected based on the closest distance of the pattern from the cluster centroid. (Eric I Pas, 1984) then used contingency table analysis to examine the relationship between trip chains and explanatory variables. Linear logit model was applied to analyse contingency table. He found out that significant sociodemographic variables were marital status, gender, employment status, education level, income and residential density. Another study which recognized the interrelationship between car ownership of households and travel patterns of household members was done by C. R. Bhat & Koppelman (1993) in which they developed a conceptual framework for activity program generation. The allocation module was used to assign the activities to household members based on auto ownership module and household needs module. A study by Wen & Koppelman (2000) shows that employment and presence of children affects tour formation. The study modeled tour patterns for each adult household member in the second stage of the complete model. It intended to study the effects of maintenance stops and other stops in tour formation of workers and non-workers. The results showed that people tend to have more than one tour if other stops is included in their daily travel pattern.

Strathman et al. (1994) applied binary logit model to estimate choice between alternative trip chains. Trip chains were divided into simple and complex chains. Work and non-work trip chains were modeled separately. The results depicted that household structure affects the trip chaining behaviour. Strathman and Dueker (1995) performed a study to examine trip chains from 1992 NPTS survey. Trip chains were first divided as work and non-work trip chains They were sub-divided into simple and complex chains which were further divided to define 7 trip chain types. Trip chains were then cross tabulated with socioeconomic characteristics including gender, income, household size, lifecycle and travel mode used to study their effect on travel behaviour. The analysis concluded with the need of more detailed activity-based surveys to understand travel behaviour in more detail.

Bowman (1998) developed an activity schedule model. The model contained two sub models namely pattern model which identifies activities and prioritize and place activities to form a tour. The second one being the conditional tour model then assigns timing, location and mode of activities. For the pattern model, a day activity pattern dimensions were defined as primary activity, primary tour structure, secondary tours, number and purpose and at-home maintenance activity participation. A multinomial logit model was applied to compute probabilities of patterns.

Golob (2000) suggested that the simple tours can be reduced by trip chaining. He developed a household trip generation model incorporating travel time of tour as time travelled to work, non-work and to return home time. The model showed the relationships of trip chains with demand for in-home and out-of-home activities. During the same time another study used micro simulation approach to develop travel patterns. Monte Carlo simulation was used to generate daily travel patterns. The study showed that it is possible to generate synthetic travel patterns through microsimulation (Kitamura et al., 2000).

Another study by (Bhat, 2001) modeled the activity pattern of the workers during the evening hours. Discrete choice systems (multinomial logit model) was used for modeling the decision to make first stop and activity type, while the duration of activities and travel time deviations were modeled as continuous choice systems. The research focused on the presence or absence of a first stop after work and the activity type of the stop for first model. The other two models were activity duration model and travel time deviation model. It was suggested to use joint model (the three models) to better model the activity participation behaviour.Gangrade et al. (2002) used nested logit model to model activity chains only for commuters. Out of home non-work activities were modeled with their temporal dimension. The model predicted the non-work out of home activities in relation to the work activity within three defined time periods namely ‘before work’, ‘at work’ and ‘after work’. Poisson and negative binomial regression models were used to predict activity frequency for each individual, while simple heuristic rule was applied to determine the sequence of activities.

Jiang et al. (2012)used the k-means clustering via principle component analysis (PCA) to cluster daily activity patterns of activities on weekdays and weekends. PCA was used to obtain eigen-activities. 21 eigen-activities were found for weekdays and 18 eigen-activities for weekends. These eigen-activities were then used to reconstruct the activity pattern. Afterwards k-means clustering was used to cluster the data, PCA was used to reduce the dimensions of the data. In this way a thorough understanding of the trip behaviour can be deducted.

Recently, a study by Ordóñez Medina (2015) focused on sequencing of activities, their start and end times, duration of activities, location of activities, mode used to perform those activities. For activity sequencing, the fixed and flexible activity approach was used, in which mandatory activities such as home and work were fixed and then prediction of flexible activities was performed. To model this an approach used was of "Activity agenda" i.e. a set of activities a person can perform, so it reduces the problem of all possible combinations of activities to restricted activities in an agenda. Separate models were created for activity location and activity type and then spatio-temporal network method was used to form activity chain. According to this method, a network is formed with nodes as location with time stamp, while links as activities or trips. Utility functions for links determine the links to be selected.

These were some of the studies in activity sequencing domain. The next section will describe the research gap found in literature along with the approach to fill the gap.

3 Thesis Contribution and Research Approach

Most of the studies described in section 2.4 showed that activity sequencing is performed in two ways. The first way is by distributing the activities as primary and secondary activities. The primary activities are fixed in time dimension and secondary activities are being placed within the available time bands. Second way is dividing the complete tour into primary and secondary tours and subsequently finding the probabilities to link tours. Time dimension and mode choice play an important role in most of the cases. In essence, the previous studies are more towards creating a combined activity sequence with activities, time and mode.

The current study is different in a way that it considers the complete activity/trip chain with only activities involved as observed in the household travel survey and then predicts the trip chain based on the person characteristics. The current study is not trying to create a trip chain.

Some of the studies that do model the trip chains from the survey data, either classify trip chains into simple and complex trip chains as in case the of (Strathman et al., 1994b), or consider very few trip chains without consideration of complete chain observed, as in the case of (Golob, 1986). It is often difficult to model all types of trip chains observed because the choice set could be quite large. This research work considers all and complete trip chains that are reported in the household travel survey as part of the analysis.

Some studies only model specific category of travellers such as commuters. The current study models all types of travellers reported. Additionally, the first application of multiple correspondence analysis (MCA) in activity sequencing is unique in itself. Therefore, this thesis makes contribution to the existing literature with respect to the points mentioned above.

The goal of the study is to develop a model that inputs a person with certain characteristics (such as age, income, gender) and to output the activity sequence/trip chain for a day. The term activity sequence or trip chain is defined as the complete sequence of activities a person performs during a day for e.g. home-work-shop-home. The terms activity sequence and trip chain are used interchangeably in this thesis. The approach involves the following steps:

1. Development of suitable activity sequence model
2. Prediction of trip chains for the individuals from the model

To achieve this goal, the following approach is used

1. Literature review of activity sequence modeling
2. Generation of trip chains from the data

3. Review of different modeling approaches and selection of a suitable approach
4. Specification of model variables
5. Estimation of model and analysis of results
6. Development of a prediction framework from the model
7. An example of model application

The first step is already covered in Chapter 2, while the rest are described in the subsequent chapters.

4 Generation of Trip Chains

Household travel surveys are important source of data for transportation modeling. These surveys contain information about separate trips of persons. For current research work, trip chains along with socioeconomic characteristics of individuals are required for modeling. The trips chains are not collected as separate entity; therefore, they must be generated from the data. The more correct trip chains generated, the more accurate the travel behaviour can be studied. The data contained inconsistencies and missing records. Therefore, some assumptions and simplifications were made. This chapter describes in detail the data characteristics and inconsistencies found in the data, the methodology adopted to generate trip chains and finally the results are discussed.

4.1 Household travel survey

Household travel surveys contain rich information about the personal travel behaviour. It has information about where a person travelled on a given day, which type of transportation mode was used, the time behaviour of trip making etc along with the socioeconomic characteristics such as age of the person, household size, income, car ownership, etc. This type of survey has a wide range of application in different fields of study such as transportation engineering, environmental studies, behavioural sciences and much more. Srinivasan and Santos (2011) summarizes the use and application of household travel surveys under the domain of transportation engineering from 2006 to 2011.

4.2 German household travel survey

The German household travel survey is also known as ‘Mobilität in Deutschland (MiD)’, meaning ‘Mobility in Germany’. For the rest of the thesis MiD is used to refer to German household travel survey. The history of household travel surveys in Germany started from mid-1970s with the objective of understanding the mobility behaviour of population. Bäumer (2017) presented a brief overview of the history of German national travel surveys. He also discussed about the approach and methodologies adopted for the MiD 2016. The latest survey has been done in 2017.

The primary data used for the model system of current study is MiD 2008. It was provided by the ‘Modeling Spatial Mobility’ research group. The MiD 2008 contains travel data reported in the form of individual trips performed by the household members. For every household, travel data of each member of the household is collected on the same day (Cyganski et al., 2013). Thus, some members in the household reported trip as “accompanying”, which implies that the person was with the other member of the household. It is useful for intra-household analysis. The trips chains are not collected as separate entity; thus, they must be generated from the data. The trips are reported for a day; therefore, trips chains also represent a day which can be assumed as average week day.

For each person, the variables used to describe the trip are ‘Origin of the first trip’, ‘Trip Purpose’ and ‘Destination of the trip’. The start time and end time of each trip are also reported. However, the spatial context of some trip purposes is missing. Each person is coded with the household id and person id, while each trip is coded with trip id. The survey covered 23089 households and 51333 persons.

Table 4.1 shows a sample from the MiD data. The column ‘Household ID’ shows the unique identification of the household the person belongs, ‘Person ID’ shows the person number within the household, ‘Trip ID’ indicates the number of trip the person performs, for instance ‘Trip ID 1’ signify the first trip of the person. The column ‘Origin of the first trip’ shows the start location of the first trip followed by the ‘Trip purpose’, which shows the subsequent activities a person performed and lastly the column ‘Destination of the trip’ shows the destinations of each trip.

Table 4.1 Sample data from MiD

Household ID	Person ID	Trip ID	Origin of the first trip	Trip purpose	Destination of the trip
200811	1	1	Home	Shopping	Other destination
200811	1	2	not recorded	Going home	Home
200811	1	3	not recorded	Shopping	Other destination
200811	1	4	not recorded	Going home	Home

For instance, the trip chain for the person from the Table 4.1 is “Home-Shop-Home-Shop-Home”. The methodology of generation of trip chains is described in section 4.4. Before generating trip chains, the data must be checked for any inconsistencies or errors to avoid implausible or illogical trip chains. Therefore, a thorough processing of MiD data was performed which is discussed in the following section.

4.3 Post processing of MiD data

4.3.1 Data inconsistencies

All the records of the survey data were checked for inconsistencies. A data exercise was carried out and trips were checked based on the starting and ending times. It was found that for some records, starting time of n-th trip was earlier than the staring time of (n-1)-th trip. This type of error could be the human error while converting the survey data from manual sheets to computerized data. The data was rearranged based on the trip start and end times, and new trip ids were generated. Other checks, such as household id and person id checks were also carried out to make sure that the person belongs to the household reported. By comparison with the household data, it was also found that for some households, trip diary of some members is missing. It could be because they did not make any trip on survey day. Other discrepancies found at certain steps are discussed while describing that step throughout the report.

4.3.2 Origin of the first trip

The first attribute of the trip reported is the 'Origin of the first trip'. MiD contains six categories for this attribute which are shown in Table 4.2 below.

Table 4.2 Frequency of trip origin

Sr.no	Origin of the first trip	Total records	Trip ID 1
1	Home	46356	45553
2	Workplace	208	207
3	Elsewhere within the locality	416	397
4	Somewhere outside	2145	2084
5	Answer denied	2	2
6	No input	1937	1880

The column 'Total records' in Table 4.2 shows the frequency of each category of the attribute in the dataset. While the column 'Trip ID 1' shows the frequency of each category as the first trip of the person. Note that it is important because this category should occur only for the first trip. But, in the dataset, it was found that these categories were also found with other than first trip. This is another discrepancy in the dataset. All the records with this type of discrepancy were simply removed.

The category 'Home' indicate that the person starts the first trip from home. Similarly, the category 'Workplace' shows that the person starts the first trip from his/her workplace. It could be possible as the person could be at work location when the survey time started, possibly if the person is working on night shifts. The category 'Elsewhere within the locality' indicates that the origin of the trip is somewhere with the city, while the category 'Somewhere outside' indicates that the first trip starts from somewhere outside the city. It could be possible as the person could be at some party the previous night and returning, or the person was at his/her relatives or friends place when the survey time starts and returning. The categories 'Answer denied' and 'No Input' show that there is no information about the start location of the trip. For current study, only the person records with first two categories, and if they appear on the first trip were selected. Rest all the persons were removed simply because for all other categories it was not sure about the location of the trip origin. Therefore, the trip chain can only begin from either home or work.

4.3.3 Trip purpose

The second attribute that explains the trip behaviour is the trip purpose. The current study considers seven activity types for modeling purpose. They are *'Home', 'Work', 'Education',*

’Shopping’, ‘Leisure’, ‘Accompanying’ and ‘Private errands’. They are selected because these are the most common type of activities used in literature.

MiD had three variables explaining the trip purpose as ‘Trip purpose’ containing 13 categories, ‘Detailed trip purpose’ containing 34 categories and ‘Main purpose of the trip’ containing 7 categories. The variable ‘Trip purpose’ was selected for the current study because the other variables contained less categories or were too detailed. Table 4.3 shows the categories along with the frequency of occurrence. The MiD contains trip based data therefore, the trips are reported. These trips are then converted into activities. For example, in Table 4.3 a trip purpose is reported as ‘Going home’. It shows that the respondent is going home to perform some activity at home. Therefore ‘Going home’ is modified to activity ‘Home’. The column ‘Activity’ in Table 4.3 shows the modification of trip chains to activities. Similarly, ‘Reaching workplace’ is modified to ‘Work’. The categories ‘Reaching training centre or school’, ‘To school or pre-school’ and ‘Day-care or kindergarten’ is combined to form category ‘Education’. The category ‘Shopping’ remains the same, while the category ‘Recreational purpose’ is named as ‘Leisure’. The categories ‘Private errands’ and ‘Other activity’ are combined and given the name as ‘Private errands’. Private errands contain activities such as walking a dog, visit to an event, private dealings and going to a religious place etc. The category ‘Dropping or picking up people’ and ‘Accompanying an adult’ are combined and given the name ‘Accompanying’. These are activities such as, someone is dropping his/her child to school or, one member of a household is travelling with the other member of the household. For example, if two members of the household are travelling together, the one member of household reports trips as e.g. home to shopping to leisure to home, the other member simply reports home to accompanying to accompanying to home.

Cyganski et al., (2013) defines a methodology of re-coding of trip purposes to produce coherent trip chains. According to the proposed methodology the category ‘Accompanying an adult’ is changed to other specific trip purposes. There is another variable in MiD which gives the information about the household members travelling together. The time of all the activities of the two members are matched and the trip purpose of the member with ‘Accompanying an adult’ categories are modified, as of the other member. In this research, this type of modification was not applied. The category ‘Return from the previous trip’ was recoded to the actual activity the person was returning from. Same was also proposed by (Cyganski et al., 2013). The category ‘Business/service’ indicates that the person is making a trip for private business or a service. This category was not included for the present study. It is a limitation of the current model as it is unable to predict business/service trips. This activity was removed from the final trip chain. For example, the original trip chain home-business-leisure-home was change to home-leisure-home. The persons who reported trip type as ‘Answer denied’ or ‘Unaware’ were simply removed from the dataset. After all these reductions, seven categories were finalized for the generation of trip chains. The column ‘Symbol’ in Table 4.3 shows the symbols of each of the finalized category as used in trip chain. Figure 4.1 shows the bar chart of the finalized categories. Note that this bar chart does not contains activities from either ‘Origin of the first trip’ variable or ‘Destination of the trip', which also contributes to the final trip chain. As it can be seen from Figure 4.1 that the activity 'Home' occurs the most, while the next in line has almost half the frequency as of 'Home'. From this it could be inferred that people prefer to perform multiple tours instead of a single tour. The activities 'Leisure', 'Shop' and 'Private errands' are the next three after 'Home', which is understandable as people perform these activities multiple times in a day. Activities 'Work' and 'Education' have low numbers which is justifiable, as people tend to perform one or two activities of this type a day. The category 'Accompanying' is at the second last place probably because it is a combination of two activities, as described above and Table 4.3 shows that 'Dropping or picking up people' contributes much more as compared to the other one, and normally this activity is also performed once or twice a day.

Table 4.3 MiD trip purpose and activities

Sr.no	Trip Purpose	Activity	Symbol	Freq.
1	Going home	Home	H	69324
2	Reaching workplace	Work	W	14625
3	Reaching training centre or school	Education	E	3455
4	To school or pre-school	Education	E	2613
5	Day-care or kindergarten	Education	E	941
6	Shopping	Shop	S	23276
7	Recreational purpose	Leisure	L	38329
8	Private errands	Private errands	PE	14665
9	Other activity	Private errands	PE	1365
10	Dropping or picking up people	Accompanying	AC	7966
11	Accompanying an adult	Accompanying	AC	2910
12	Return from the previous trip	recoded	-	3243
13	Business/service	ignored	-	10548
14	Answer denied	ignored	-	13
15	Unaware	ignored	-	17

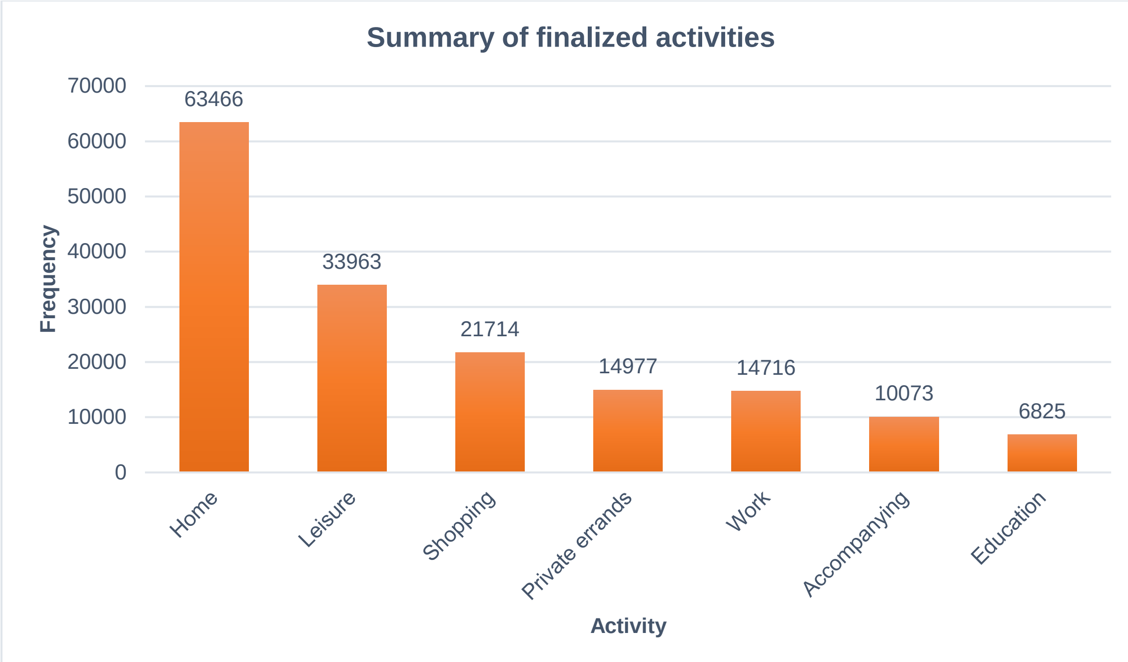


Figure 4.1 Summary of finalized activities

4.3.4 Destination of the trip

The last attribute defining activity sequence is the destination of the trip. It defines the destination of the trip purpose or in other words, the destination of the activity. Table 4.4 below shows the categories for this attribute as reported in MiD along with the frequency of each category.

Table 4.4 Frequency of trip destination

Sr.no	Destination of the trip	Freq.
1	Home	29776
2	Workplace	14175
3	Other destination within the city or the area	50579
4	Other destination outside the city or the area	26365
5	Round trip	8414
7	Answer denied	2
8	Unaware	23
9	No input	154

The category 'Other destination within the city or the area' and 'Other destination outside the city or the area' is mostly used to describe the destination of the activities like shopping, leisure, private errands etc. For the destinations 'Answer denied', 'Unaware' and 'No input', the corresponding activities were used for generation of trip chains. The categories 'Home' and 'Workplace' mostly corresponded to activities 'Home' and 'Workplace' in trip purpose variable. But in some records, it was also found that the two does match. For example, the activity shows 'Accompanying' and the corresponding destination is 'Home'. It is possible as may be a household member is accompanying another household member to home. For destination as ‘Workplace’, for some records it corresponds to activities such as education, leisure, shop etc. After thorough checking and cross validation of some records individually, it seemed to be the reporting error by some individuals. For e.g., when records for corresponding activity as education was checked, it was found by other attributes such as age, employment activity, etc. that the person is a student and reported activity as education but destination as workplace. It could be possible as the respondent may have confused occupation (student in this case) with work category. Similar checks were undertaken for other mismatches of activities and destination as work. Therefore, for destination as ‘Workplace’, corresponding activities were considered for generation of trip chains. Similar checks were performed for destination as ‘Home’ and the results supported destination category as compared to corresponding activity category. Consequently, if the destinations for ‘Home’ did not match with the activity category, then the activity category was replaced by destination category as shown in Figure 4.2. The destination ‘Round trip’ indicates that the destination is the origin of that individual trip. It was re-coded as the origin of that individual trip from the trip purpose/activities as shown in Figure 4.2**.**

Trip ID	Origin of the first trip	Trip Purpose	Destination of the trip	Origin of the first trip	Trip Purpose (modified)	Destination of the trip (modified)
1	H	AC	other destination	H	AC	other destination
2	not recorded	AC	other destination	not recorded	AC	other destination
3	not recorded	AC	H	not recorded	H	H
4	not recorded	AC	round trip	not recorded	AC	H
Trip chain				**Trip chain**		
H - AC - AC - AC - AC -AC				H - AC - AC - H - AC - H		
H = Home AC = Accompanying						

Figure 4.2 Recoding of trip purpose and destination of the trip

It can be seen from Figure 4.2 that trip chain after modification is more logical as compared to trip chain before modification.

4.4 Formation of trip chains

4.4.1 Procedure

After completion of data filtering and re-coding of attributes, trip chains were generated. The process of trip chain generation is as follows. For each person, origin of the first trip was selected first, followed by incrementally adding activities performed and finally the destination of trip. However, destination of the trip is not added for all trips as discussed in previous section. Figure 4.2 also depict this process. Another complex example is shown in the Figure 4.3 to understand the process better.

Trip ID	Origin of the first trip	Trip Purpose	Destination of the trip
1	H	S	Ro
2	not recorded	L	other destination
3	not recorded	Re	other destination

Generated Trip Chain
H - S - H - L - H

Legend
H = Home S = Shop L = Leisure Re = Return from previous trip
Ro = Round trip

Figure 4.3 Trip chaining example

The trip chain was generated as starting from trip origin i.e. ‘Home’ to trip purpose i.e. ‘Shop’ and then destination, which says ‘Round trip’, so the next activity to be added is ‘Home’ again, followed by ‘Leisure’ and the last activity is ‘Return from previous trip’, so in this case it returns not to ‘Shop’ but to trip destination which was ‘Round trip’, which in turn was ‘Home’. So, the final trip chain generated is ‘home-shop-home-leisure-home’. Note that in this example, only destination of first trip is used in trip chaining process. This whole procedure was coded in R.

4.4.2 Post processing of trip chains

After the generation of trip chains for each member of a household, the next step was to remove certain activities from the trip chain. The first activity to be removed was business trips. As discussed earlier in section 4.3.3 that business trips are not modelled in this study, which is also a limitation of the model. The persons with activities were not removed, but only the activity was removed. This activity (i.e. Business/service) was not removed before the generation of trip chains, because of ‘Round trip’ and ‘Return to previous trip’ category, as trip can be returned to business trip. Another problem encountered was occurrence of consecutive activity of ‘Home’, ‘Work’ and ‘Education’ in the trip chains. They were also reduced to a single activity, because either of the activity type can have a single destination and it does not make sense that a person is making a trip from home to home. It was also occurred after removing business trips (e.g. H-S-Bu-H-Bu-H which after removing Bu becomes H-S-H-H). Figure 4.4 shows example of activity reduction from trip chains.

Sr.no	Observed Trip Chain	Modified Trip Chain
1	H - W - Bu - Bu - H - L - H	H - W - H - L - H
2	H - E - H - PE - S - H - H	H - E - H - PE - S - H
3	H - W - W - L - W - H	H - L - W - H
4	H - E - E - PE - E - PE - H	H - E - PE - E - PE - H
5	H - L - H - L - H - Bu - H	H - L - H - L - H

Legend
H = Home W = Work E = Education L = Leisure S = Shop
PE = Private errands AC = Accompanying Bu = businuess/service

Figure 4.4 Trip chain post processing

4.4.3 Results

A total of 5683 unique trip chains were generated from 45746 individuals. Figure 4.5 shows top 20 trip chains based on frequency. A simple work commute has the highest frequency of occurrence which was expected. The second in line is the simple leisure trip followed by a shop trip. It shows that most people prefer to perform shopping and leisure explicitly in a day rather than combining it with other tours. It is important to note here that trip chain 10 does not end at home, which is the case for many other trip chains not shown in Figure 4.5. Out of 5683 trip chains 4547 trip chains have the last destination as home, while the rest of the chains have other activities as destinations. It is possible as the respondent may not have returned home as the survey period ends. Therefore, the trip chains not ending at home were kept. It is also to be reminded that all the trip chains only start from either home or work location. Another activity sequence of only ‘Home’ was observed with a frequency of 249, not shown in Figure 4.5. It was mainly due to the result of removing business trips and this trip chain was kept for the modeling purposes. For this case, it was assumed that the person did not make any trip on that day and he/she rather stayed at home. Rest, all the trip chains are self-explanatory. A complete list of all observed trip chains along with their frequency is attached in Appendix A.0.

Sr no.	Trip Chain	Freq.
1	H - W - H	4467
2	H - L - H	4324
3	H - S - H	2360
4	H - E - H	2091
5	H - PE - H	1524
6	H - E - H - L - H	1240
7	H - S - H - L - H	1101
8	H - L - H - L - H	1084
9	H - W - H - L - H	1041
10	H - L	1029
11	H - W - H - S - H	675
12	H - AC - H	425
13	H - S - H - S - H	394
14	H - L - L - H	385
15	H - PE - H - L - H	382
16	H - W - S - H	351
17	H - S - H - PE - H	349
18	H - W - H - PE - H	304
19	H - S - S - H	260
20	H - W - H - W - H	251

Legend
H = Home   W = Work   E = Education   L = Leisure   S = Shop
PE = Private errands   AC = Accompanying

Figure 4.5 Trip chains observed (top 20)

5 Modeling of Trip Chains

Modeling trip chains is a complex process. The purpose of this research is to model trip chains based on household and person attributes. To achieve this goal, it is important to study the relationship between trip chains and individuals. For this purpose, different statistical methods were studied to find out which method best serves the purpose. This chapter gives an overview of the different approaches studied and the reasons for not implementing them. It provides future researchers an overview about limitations and suggestions of different approaches with transportation modeling view point. Finally, the last section describes in detail the modeling methodology adopted for current study.

5.1 Review of modeling approaches

5.1.1 Cluster analysis

Cluster analysis is a multivariate technique, used to form clusters from multivariate data objects. The data is divided into various groups (clusters) based on similarity and differences among data objects. The data points in a cluster are homogenous among themselves and different among other clusters. Clusters are formed by using different proximity measures which determines how data points are similar or dissimilar to each other. There are also different algorithms for clustering, the most commonly used being hierarchical clustering and k-means clustering algorithms. It is used in various disciplines such as marketing, psychology, medicine, sociology, linguistics and biology ("Cluster Analysis," 2007).

In transportation modeling, various studies have been done using cluster analysis. For instance, E I Pas (1984) used clustering algorithm to form groups of similar daily travel-activity patterns based on a similarity matrix derived by the author. Jiang et al. (2012) used k-means clustering via principal component analysis to cluster individuals based on dissimilarity of their daily activities. A similar attempt was made during this study to cluster trip chains into various groups based on the household and individual attributes using k-means clustering algorithm. The first problem encountered was the presence of categorical variables which makes it difficult for clustering algorithm (Jiang et al., 2012b). Different studies propose different methods to tackle this problem (Ralambondrainy, 1995), (Huang, 1998). Each method has its own pros and cons. Other possible way was to develop a kind of similarity matrix as was developed by Eric I. Pas (1983), which could be used as a proximity measure. But, this kind of similarity matrix is to be based on trip chains, which somehow neglects the attributes of person making the trip.

Another issue in applying clustering algorithm was how to form clusters. Either the clusters should be formed based on attributes or some other proximity measure or to form cluster for each trip chain. The former one suggests, that each cluster contains set of trip chains, which limits the whole data set into few chains. For each cluster there could be one representative chain. Therefore, while predicting, a cluster could be predicted but it was further difficult to predict the trip chain from the chosen cluster. The later approach could suggest forming around 5000 clusters representing each trip chain. Each cluster would have individuals that perform that trip chain. Formation of large number of clusters does not serve the purpose of current study as the relationship of persons with trip chains was not modeled.

Trivedi, Pardos, and Heffernan (2015) explain how prediction can be made from cluster analysis. They suggested building a separate prediction model for each of the cluster and then form a combined prediction model. But the problem remains the same as each cluster contains a set of trip chains, and one trip chain is to be selected, which is somehow similar to having all trip chains and selecting one trip chain.

The reasons identified above correspond to the specific purpose of the study. They also correspond to the specific definition of trip chains. Further research can be done based on specific requirements of the study.

5.1.2 Other methods

Apart from cluster analysis, other methods were also studied. The most common method used is the logit model. It includes multinomial logit model and nested logit model. The methodology behind using logit models is to form a trip chain by including each activity step by step. Such studies include (Gangrade et al., 2002) and (Bhat, 2001) among many other. For the current study, the whole trip chain must be used, which makes logit models a bit difficult to use because of the large number of choice alternatives.

5.2 Multiple Correspondence Analysis (MCA)

Correspondence analysis (CA) is a multivariate data analysis tool. It is used to study the relationship between rows and columns categories of a survey data (Härdle and Simar, 2015). Rows and columns can represent a variable with categories of interest. It is also used as data visualization technique (Härdle and Simar, 2015). One of the pioneer works on development of correspondence analysis was done by Hirschfeld in the first half of the 20th century (Härdle and Simar, 2015).

Multiple correspondence analysis (MCA) is an extension, or in other words an application of correspondence analysis when there are more than two categorical variables. MCA is mostly used to analyse surveys where individuals respond to questions with specific answers. Typically, rows represent respondents and columns represents variables. For MCA, each question is considered as a variable and each answer is considered as a category of that variable (Husson et al., 2017). Its application is not limited to surveys but is also applied in other fields. It has applications in environmental science, archaeology and others ((Greenacre and Blasius, 2006),(Husson et al., 2017)).

MCA procedure is similar to the principal component analysis (PCA)*, while the latter one is used for continuous data. MCA is generally used to perform three types of analysis from survey data which are (Husson et al., 2017, pp. 128–131).

1. To identify group of individuals with similar behaviour
2. To identify the associations between variable categories
3. To identify relationship between individuals and categories

MCA is performed on indicator matrix where each row represents an individual and each column is a variable category. Each element in a matrix can be represented by x_{ik}, which represents an individual 'i' selecting category 'k'. The element x_{ik} can only have value 0 or 1 based on the absence or presence of category 'k' for individual 'i'. The table has the dimension of $I\ X\ K$. (Husson et al., 2017, p. 130).

5.2.1 Relationship between Individual

Relationship between individuals is measured as the distance between individuals. Each individual has a set of points in K dimensions, where each dimension is a variable category. In MCA, each category has a weight which is proportional to the number of individuals the category contains, while each individual has the same weight which is equal to the inverse of the total number of individuals (Husson et al., 2017, p. 130). The distance between the individual i and individual i' can then be calculated by using the equation described below:

Equation 5.1

$$d_{i,i'}^2 = C\ \sum_{k=1}^{K} \frac{(x_{ik} - x_{i'k})^2}{I_k} \quad \text{(Husson et al., 2017, p. 130)}$$

Where;

x_{ik} = element for individual i with category k
I_k = number of individuals carrying category k
C = constant = I/J
I = total number of individuals
J = total number of variables

*PCA is a statistical technique used to reduce the dimensions of data by creating principal components

5.2.2 Relationship between categories

Relationship between categories is also based upon the distance between categories. Each category in the contingency matrix has points in 'I' dimensions where each dimension corresponds to an individual. Each category has a weight which is proportional to the number of individual the category contains, while each individual has the same weight equal to the inverse of the total number of individuals (Husson et al., 2017, pp. 130–131). By counting the number of individuals in each category, the distance between two categories say k and k' can be calculated as:

Equation 5.2

$$d^2_{k,k'} = C' \sum_{i=1}^{I} \left(\frac{x_{ik}}{I_k} - \frac{x_{ik'}}{I_{k'}}\right) \text{ (Husson et al., 2017, p. 131)}$$

Where;

x_{ik} = element for individual i with category k
I_k = number of individuals carrying category k
${I_k}'$ = number of individuals carrying category k'
C' = constant = I

5.2.3 Supplementary Elements

Another important aspect of MCA is the use of supplementary points. It can be individuals or variable categories (Husson et al., 2017, p. 142). Supplementary variables are additional variables and are not involved in the formation of dimensions*. They are projected on the existing model and helps to explain the results. Transition formulas are used to display supplementary points (Greenacre and Blasius, 2006, p. 31).

5.2.4 Dimensionality

Similar to the principal component method, dimensions are formed in MCA. Each dimension represents a certain amount of variance in the dataset. Each dimension can also be interpreted as a synthetic variable (Husson et al., 2017, p. 15), which is a formed by contribution of each variable category. Some dimension may explain variability of a certain category more as compared to other category. The method finds the principal dimensions which explain the variability between individuals and associations among variable categories. The dimensions are orthogonal to each other. Thus, instead of K dimensions for individuals and I dimensions for categories, the reduced number of dimension i.e. principal dimensions are formed. They also help to plot the data in two-dimensional space.

*Refers to principle dimensions explained in sec 5.2.4

6 Modeling Framework

This chapter demonstrates the modeling framework based on multiple correspondence analysis (MCA), which was described conceptually in section 5.2. Section 6.1 discuss about the microscopic level of analysis used for modeling. Section 6.2 describes in detail model parameters and design for MCA. Section 6.3 explains model estimation using *FactoMineR* package and estimation results. The prediction framework is described in section 6.4. The model calibration process is described in section 6.5. The last section of the chapter shows the application of the model with the help of an example.

6.1 Household versus person level analysis

For this study, person-level is used for the modeling purposes. Therefore, that trip chains for individual persons are modelled as compared to a household. There is a debate in the literature as to which analysis unit is better than the other. For instance, some studies support person-level of analysis while some support household-level of analysis. There are also some studies which are not conclusive. Downes et al. (1978) compared household and person-level trip generation models. They found out that both the models performed similarly, however, the person-level model was statistically better than the other one. Supernak et al. (1982) developed a person-category trip generation model. They concluded that person-category model is better than household-category model as it has better behavioural background. Contrastingly, a study by Badoe & Chen (2004) suggested household-level of analysis as the predictions of trips made are more accurate than person-level of analysis. A recent study by Pokhrel (2017) was not conclusive and he stated that he did not find any concrete evidence to support any of the analysis unit. As per the research done, there was no consensus found on analysis unit in the literature.

This study modeled trip chains at person-level because the methodology used for model building can provide better results with person-level as compared to using household as analysis unit. However, a shortcoming of using person-level is the limitation to study intra-household relationships properly. For instance, it is difficult to explain accompanying trips - a household member reported trip chain as home *to accompanying to home*, it is difficult to say which household member the respondent was accompanying because it is possible that when predicting trip chains for households, one member gets assigned trip chain from one household while the other member gets assigned trip from another household in the model. Therefore, joint trips of household members are lost or ignored. On the other hand, individual-level can provide a greater heterogeneity in trip chaining behaviour. A sequential procedure must be adopted while assigning individual trip chains to household. For example, after the assignment of trip chain to one member of a household, other members should be assigned trip chains accordingly, to avoid assigning all members with accompanying trips.

6.2 Model specifications

6.2.1 Trip chains

To study the relationship between trip chains and socioeconomic variables, trip chains were generated from MiD 2008 as described in section 4.4. A total of 5683 unique trip chains were generated from 45746 individuals. Records with missing value of any of the variables (section 6.2.2) were deleted. The individuals reduced to 39080 while the trip chains reduced to 5136. Most of the trip chain types were rare types with very low number of survey records. For MCA, it is important to have sufficient sample size for each category to obtain correct relationships. The one approach to overcome this problem is by grouping categories. (Husson et al., 2017, p. 150) For current study, trip chains with 30 or more records were considered for modeling purposes based on two reasons.

First, for household type definition in trip generation, it is state of practice to use 30 survey records as a threshold for each household type to be considered (Moeckel et al., 2015). 112 trip chains were found with survey records of 30 and more. It is important to note that these 112 trip chains represent 72% of the total number of records as shown in Table 6.1. It can also be seen from Table 6.1 that criteria of 20 and above would add around 60 more trip chains while the dataset representation would only increase by 4%. The criteria of 10 and above would add 193 trip chains with dataset representation of 8% more.

Table 6.1 Trip chain statistics

Records	Unique trip chains	Frequency/Individual	Share (%)
Total	5136	39080	100
30 and above	112	28172	72
20 and above	171	29553	76
10 and above	305	31379	80

Representation of more than 70% of the dataset with reasonable number of trip chains is the second reason to select trip chains with survey records of 30 and more. These 112 trip chains were generated by 28172 individuals. For simplicity, each trip chain is coded as TC (1-112), where TC1 is the most frequent trip chain while TC112 is the least frequent trip chain. All 112 trip chains with codes are listed in Appendix A.1.

6.2.2 Socioeconomic variables

Two types of socioeconomic variables were considered for modeling purpose. First type is the one that explains characteristics of household and second type being the one that explains characteristics of individual. Table 6.2 shows variable types, categories for each variable along with number of observations.

Table 6.2 Variable and categories considered for the model

Variable type	Variable	Categories for analysis	Category description	Number of records
Household	Household size (number)	HHS1	1	2065
		HHS2	2	9794
		HHS3	3+	16313
	Household Auto ownership (number)	AO_0	0	2175
		AO_1	1	12858
		AO_2	2	10349
		AO_3+	3+	2790
	Income (€)	IG_1	0 to < 2000	7863
		IG_2	2000 to <3000	9177
		IG_3	3000 to < 5000	8719
		IG_4	> 5000	2473
Person	Gender	G_M	Male	14381
		G_F	Female	13791
	Age (years)	AG_1	0 to 22	7215
		AG_2	23 to 47	7498
		AG_3	48 to 64	7943
		AG_4	65 and above	5516
	Driving License	DL_Y	Yes	20079
		DL_N	No	8093
	Occupation	O1	Worker	13059
		O2	Student	5084
		O3	Other	10029

The variables were considered based on the state-of-the-art practice in transportation modeling that affects the trip chaining behaviour. These are similar to the variables used in studies described in chapter 2. As per the author's knowledge, there is no method in MCA to detect which variable is statistically significant. However, contribution of each variable to the construction of MCA dimensions can give an idea of the most important categories which explains variability in the dataset. The variables listed in Table 6.2 are discussed below in detail.

Household size and auto ownership

Household size and auto ownership are the household characteristics. The household size categories were described as single, couple and multi-member household. Single member household can have complete different travel pattern than other two categories. As a single member has to perform all the necessary tasks required in a household such as work, shopping etc. Two-member household could be a couple and a single parent with a child household, and the travel behaviour could be different from other two categories. Households with more than two members were combined as the travel behaviour with three or more members is almost similar. For example, a household with a couple and three children is assumed to have similar travel behaviour to a household with a couple and four or five children. Figure 6.1 shows the household size categories with frequency.

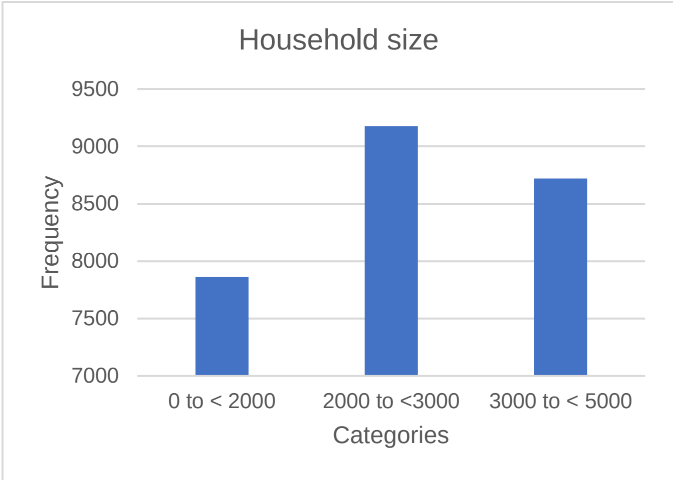


Figure 6.1 Household size categories

Auto ownership has four categories. Category 0 i.e. no car is to distinguish between whether a household owns a car or not. Auto ownership of 1 car and 2 cars are more common and were considered as separate categories. If a household owns three or more cars, it was combined to form category 3+. It is due to the fact that it is rare for a household to have more than 3 cars, and there were less number of records for more than 3 cars, so they were simply combined to form one category i.e. 3+. Figure 6.2 gives and overview of auto ownership categories with number of records for each category.

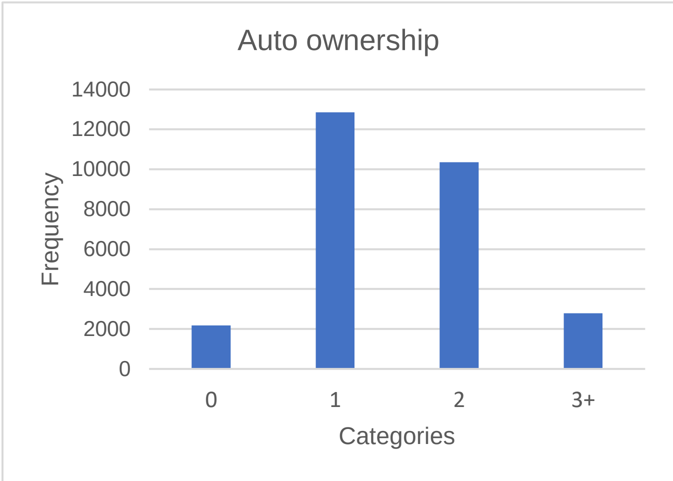


Figure 6.2 Household auto ownership categories

Household income

Household income is another important variable that affects travel behaviour. The income categories were defined based on number of records in each category. MiD 2008 defined 15 income categories. These categories were combined as to have similar number of records in each category. The last category with income of more than 5000 euros has very few records as compared to the other categories. It shows that such high incomes are generally rare in Germany. Figure 6.3 shows the summary of income categories used in the current study.

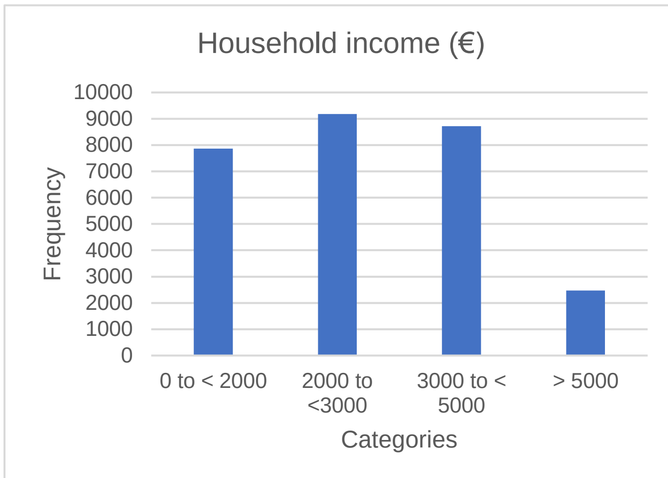


Figure 6.3 Household income categories

Gender and driving license

Gender and driving license can have only two possibilities. For gender one could be either male or female. Figure 6.4 shows that the dataset captures almost equally both the groups with male population slightly higher than female population. For driving license, persons holding a driving license are approximately double than the one who does not possess as shown in Figure 6.5 below.

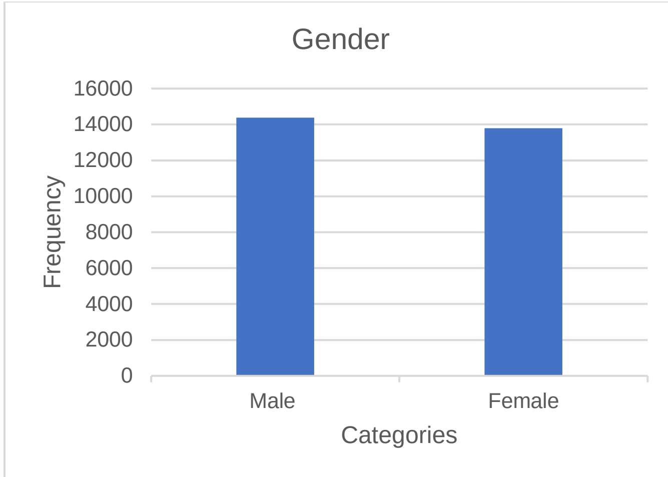


Figure 6.4 Gender categories

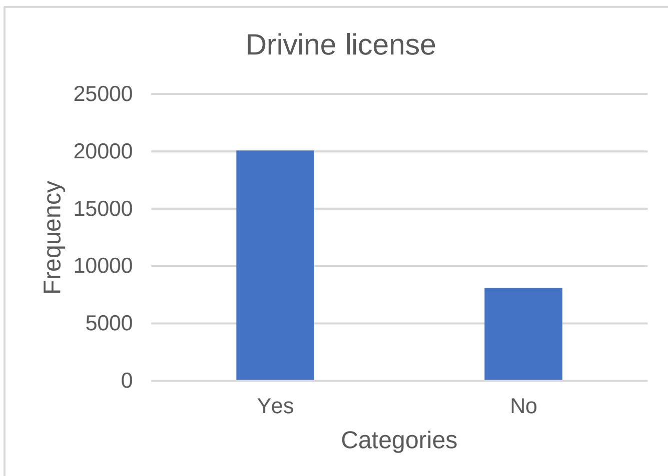


Figure 6.5 Driving license categories

Occupation and Age

MiD 2008 contains eight categories of occupation which were merged into three. The category worker contains both full time and part time professionals, category student contains school pupils and students and lastly, the category other is a combination of children, home maker, pensioner and others, as reported in MiD 2008. Figure 6.6 shows that most of the population in the data set consists of workers followed by other and students. It can be inferred that probably the MCA model will better represent workers as compared to other two categories. The categories of age were formed in such a way that it should correspond to occupation categories as well as each category should have similar number of records. Four categories of age were formed as shown in Figure 6.7. The first category represents students, as normally students complete their bachelors by the age of 22 years. The other two age groups from 23 years to 64 years represents mostly workers, while the last category was formed based on the fact that people mostly retire at the age of 65 in Germany. The category other can be in any of the age category.

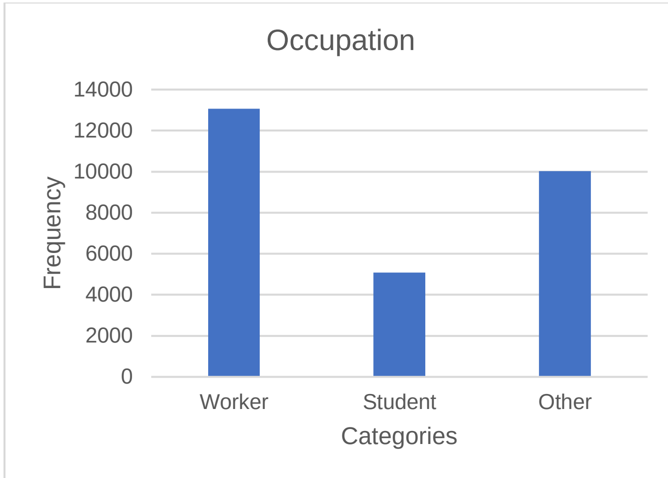


Figure 6.6 Occupation categories

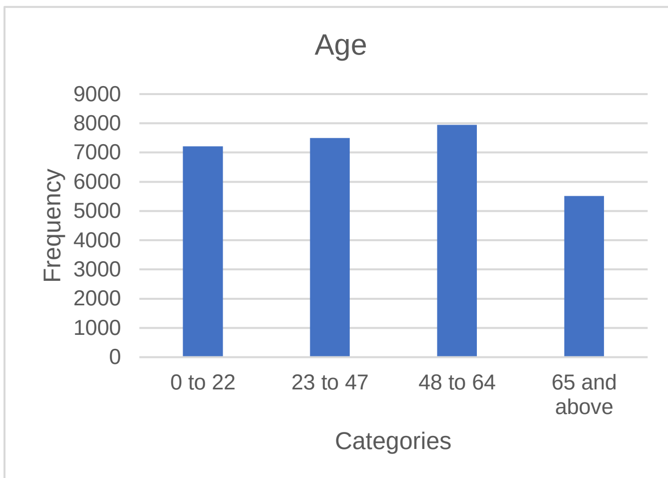


Figure 6.7 Age categories

It is important to note that workers and students can also be in any of the age category, this classification is just to have a simplified relation among variables. There were few cases observed which showed a student with a very high age and vice versa.

6.3 Model estimation and results

6.3.1 Estimation using *FactoMineR* R package

The *FactoMineR* package was developed for multivariate exploratory data analysis. It is developed by François Husson, Julie Josse, Sébastien Lê, d'Agrocampus Rennes, and J. Mazet. The detail description about the package is available at (Lê et al., 2008). The package contains a lot of functions, but the function used for multiple correspondence analysis is called 'MCA'. The dataset was prepared, and the function was called as described below.

Dataset preparation

The dataset was prepared in a way that each row represented an individual and each column represented socioeconomic variables and a trip chain, as reported in the MiD 2008. These variables naturally fall into two groups, one defining the person and the other defining travel behaviour. The current model uses socioeconomic variables as active variables and trip chains as supplementary variables. It is because an individual is defined by its socioeconomic variables and not by trip chain, as trip chain was recorded for one day, and an individual can have different trip chains for any other day. The dimensions were formed using active variables i.e. socioeconomic variables, which signifies that principal dimensions of variability were formed using these variables. Then supplementary variables i.e. trip chains were added which showed the relationship between the dimensions of variability and supplementary variables.

To use both types of variables (i.e. socioeconomic and travel behaviour) as active variables, some other methods such as multiple factor analysis are required to achieve a certain equilibrium between the two types of variables (Husson, 2016). Furthermore, addition of trip chains as active variable would result in very large number of dimensions. All variables used in the dataset are categorical variables.

Function for MCA

The function used from *FactoMineR* for estimation is described in Equation 6.1 below:

Equation 6.1

$$MCA(dataset, quali.sup, ncp)$$

Where:

$dataset$ = input file in the format as described above
$quali.sup$ = defining qualitative supplementary variables
ncp = number of dimensions

For current model, qualitative supplementary variables are defined as trip chains and number of dimensions were set to maximum possible.

6.3.2 Estimation Results

The model was estimated with different categories of variables, and finally after performing few iterations, the variable categories defined in section 6.2.2 were used for model estimation because the results were more plausible with these categories. The results of the model estimation are described below:

Eigenvalues

The model estimation resulted in the formation of 15 principal dimensions* as shown in Table 6.3 below.

Table 6.3 Variable decomposition per dimension

Dimension	Eigenvalue variance	Percentage of variance	Cumulative percentage
1	0.399	18.599	18.599
2	0.335	15.640	34.239
3	0.199	9.285	43.524
4	0.165	7.691	51.215
5	0.148	6.896	58.111
6	0.144	6.702	64.812
7	0.138	6.430	71.242
8	0.137	6.411	77.653
9	0.119	5.537	83.190
10	0.097	4.509	87.699
11	0.076	3.545	91.243
12	0.068	3.188	94.432
13	0.047	2.201	96.633
14	0.043	1.993	98.626
15	0.029	1.374	100.000

The number of dimensions to be formed is estimated by the MCA method itself. Each dimension explains certain amount of variance in the dataset. For e.g. dimension 1 explains 18.6% of the variance in the dataset, while dimension 2 explains 15.6% of the variance in dataset. Together both the dimensions (first and second) explain 34% of variance in dataset. By default, every consecutive dimension explains less variance then the previous one.

Variables

Each variable category can be represented in all the 15 dimensions. Seven variables with a total of 22 categories were modeled as described in section 6.2.2. Table 6.4 shows the estimation results for categories in first two dimensions. The column 'Dim.1' and 'Dim.2' shows the coordinate points of each category for the first and second dimension. The column 'ctr' represents contribution of each category in the formation of the respective dimension. For instance, AG1(age group 1) contributes the most in the definition of dimension 1 with the share of 14%, while G_M (gender male), contributes the least in the formation of dimension 1 with the share of only 0.065%. Similarly, for dimension 2 DL_N (no driving license) contributes the most (17%), while HHS2 (household size 2) contributes the least (almost 0% as rounded). The column v.test is a test statistic which follows a gaussian distribution, the value of below -1.96 or above 1.96 for a

*The concept of dimension is explained in section 5.3.4

category shows that the coordinate is significantly different from zero.

Table 6.4 Results of categories in first two dimensions

Categories	Dim.1	ctr	v.test	cos2	Dim.2	ctr	v.test	cos2
HHS1	1.345	4.753	63.490	0.143	0.695	1.508	32.791	0.038
HHS2	0.846	8.913	103.628	0.381	-0.013	0.003	-1.642	0.000
HHS3	-0.678	9.542	-133.474	0.632	-0.080	0.158	-15.727	0.009
AO_0	0.996	2.743	48.334	0.083	1.179	4.576	57.246	0.116
AO_1	0.384	2.408	59.008	0.124	0.224	0.980	34.514	0.042
AO_2	-0.500	3.286	-63.889	0.145	-0.311	1.516	-39.792	0.056
AO_3+	-0.691	1.697	-38.474	0.053	-0.799	2.698	-44.486	0.070
AG1	-1.220	13.664	-120.149	0.512	0.980	10.478	96.483	0.330
AG2	-0.112	0.119	-11.301	0.005	-0.848	8.158	-85.712	0.261
AG3	0.309	0.967	32.529	0.038	-0.616	4.566	-64.823	0.149
AG4	1.302	11.904	107.861	0.413	0.759	4.805	62.835	0.140
G_F	0.062	0.067	10.179	0.004	0.127	0.336	20.845	0.015
G_M	-0.059	0.065	-10.179	0.004	-0.122	0.322	-20.845	0.015
O1	-0.079	0.104	-12.319	0.005	-0.869	14.922	-135.583	0.653
O2	-1.410	12.857	-111.038	0.438	1.139	9.975	89.687	0.286
O3	0.817	8.527	102.014	0.369	0.554	4.662	69.171	0.170
IG_1	0.860	7.406	89.854	0.287	0.586	4.090	61.233	0.133
IG_2	-0.093	0.099	-10.740	0.004	-0.069	0.066	-8.024	0.002
IG_3	-0.495	2.717	-55.607	0.110	-0.312	1.286	-35.076	0.044
IG_4	-0.650	1.329	-33.833	0.041	-0.509	0.969	-26.495	0.025
DL_N	-0.688	4.871	-73.291	0.191	1.180	17.055	125.756	0.561
DL_Y	0.277	1.963	73.291	0.191	-0.476	6.874	-125.756	0.561

Table 6.4 shows that all variable categories can be considered significant for dimension 1, while for dimension 2, all variable categories except HHS2 (household size category2) can be considered significant. The column ‘cos2’ represents the quality of representation of a category with respect to dimension. It shows the degree of association between variable category and dimension as defined by squared cosine(cos2) or the squared correlations. The quality of representation is with respect to the factor map as shown in Figure 6.9 (which is described later). ‘cos2’ value can be between 0 and 1. The sum of all cos2 value for a category is equal to 1. If a category is well represented in the first two dimensions than the sum of cos2 for the two dimensions is close to 1. Not all categories are displayed well in any particular dimension. Figure 6.8 shows the squared correlation between categories for all 15 dimensions. It can be seen that HHS3 (household size category 3) is well represented in dimension 1 while IG_3 (income group 3) is best represented in dimension 6. Overall it can be seen in Figure 6.8, that most of the categories are well represented in the first two dimensions (summation of cos2 for the first two dimensions) as compared to other dimensions. Therefore, a factor map of first two dimension is

shown in Figure 6.9. The parameters described in Table 6.4 are provided in Appendix B.1 for all 15 dimensions.

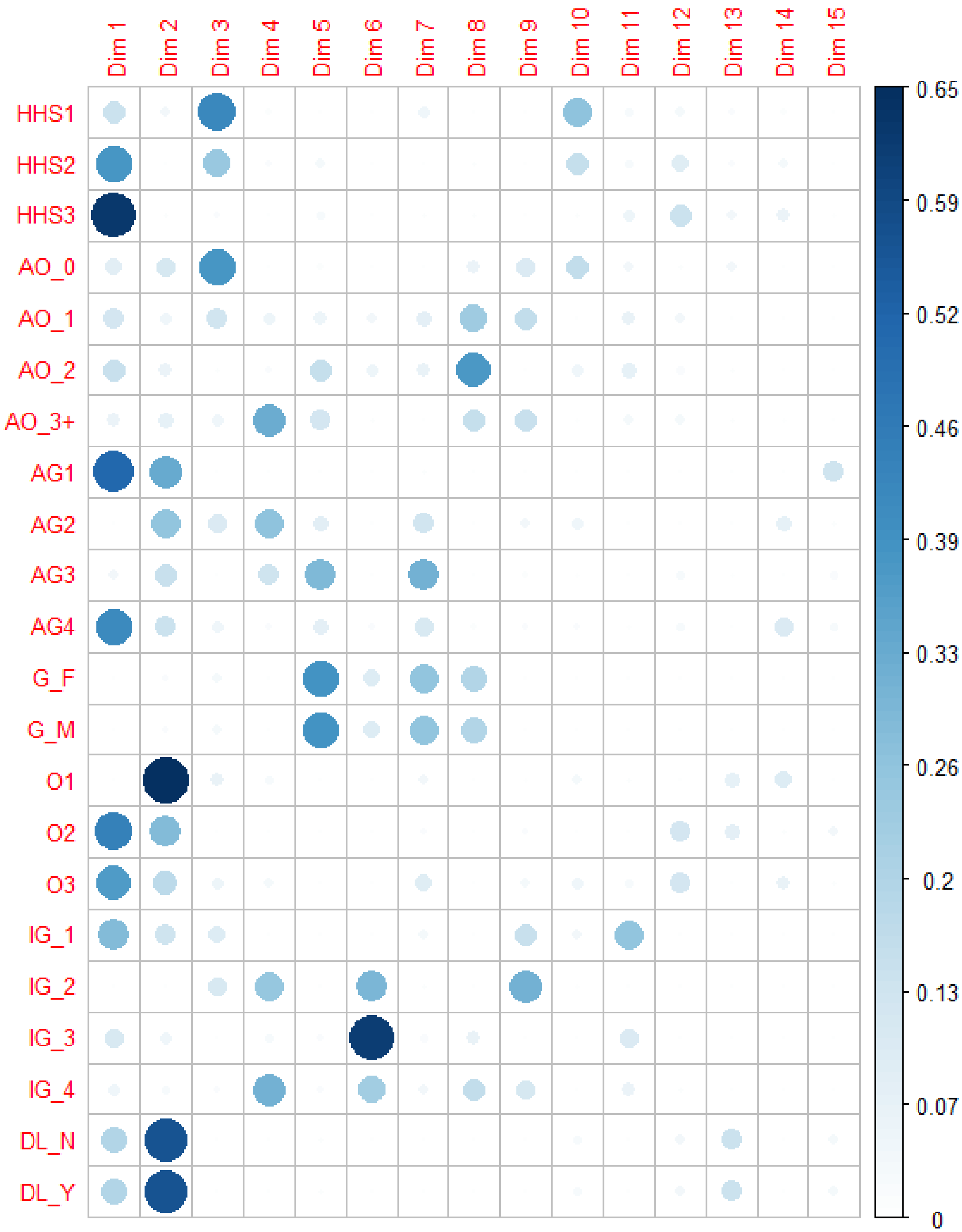


Figure 6.8 Squared correlation between category and dimension

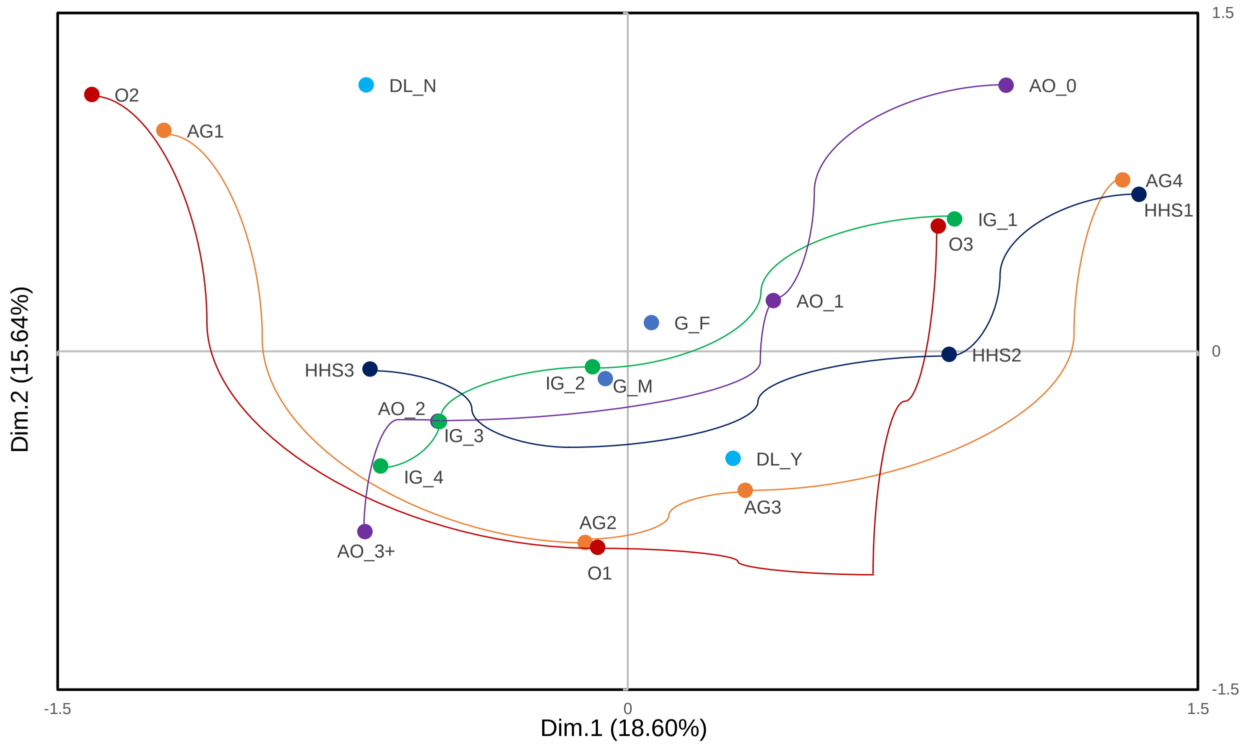


Figure 6.9 Factor map of variable categories represented in first two dimensions (cum.var 34%)

Figure 6.9 shows the relationship among variable categories in the first two dimensions. The closer the categories the more related they are with each other. The categories are coloured, and the lines show the distribution of each category along the map. The top left part of the map shows that students (O2) are more likely to fall between 0-22 years (AG1) and without having a driving license (DL_N). It is not very close as people with age 17 or older can have a driving license in Germany. Therefore, this part of the map mostly represents students. The middle portion of the map shows that workers (O1) are more likely to be between 23 to 64 years (AG2 and AG3), having high categories of incomes starting from 2000 euros and above (IG_2, IG_3 and IG_4), with more chances of owning a car and a driving license (DL_Y). The top right portion of the graph mainly show retired people (O3) with low income (probably due to pension) and high age (AG4) of 65 years and above. They could also be homemakers or currently jobless people with lower income and more likely to not own a car.

The relationship of household income and auto ownership is well depicted. It shows that people with low income (IG_1) either do not own a car (AO_O) or have chances to own only one car (AO_1). The people with medium income (IG_2) are more likely to own one or two cars, while people with high income (IG_3 and IG_4) have more chances of owing two and more cars (AO_2 and AO_3+).

The distribution of household size is also interesting. The students (O1) are more likely to stay with their parents (which represents household size of 3 or more people, HHS3) until they complete their basic education and leave for higher studies or start working. Workers (O2) are more likely to be couples (HHS2) or might have start their family having kids (HHS3). There is also a lower probability of a worker to live alone. Senior citizens (AG4) are more likely to live alone (HHS1) probably due to the fact that the children move out for study, work or starting a new family, or probably most of them become isolated when their spouses died.

The gender categories are close to each other and with the origin, which shows that gender of a person does not influence other characteristics. It can be inferred that male population of workers are slightly higher that female workers. The interpretation of the above map seems quite close to the reality, but it should be kept in mind that the Figure 6.9 is only showing the relationship among variable categories in two dimensions. The other dimensions can also influence the position of the categories as shown in Figure 6.8

Trip chains

The distribution of trip chains in first two dimensions are shown in Figure 6.10. The MCA parameters of the trip chains are attached with Appendix B.1.2. These parameters have similar interpretation as described in section of variables except that parameter ctr is not included for trip chains as they did not contribute in the formation of dimensions. The map below show that the trip chains are broadly grouped into three clusters. The one at the top left are the trip chains that mostly contain education activity, the cluster at the centre bottom mostly contain the work activity and the cluster in the middle contain every type of activity modeled.

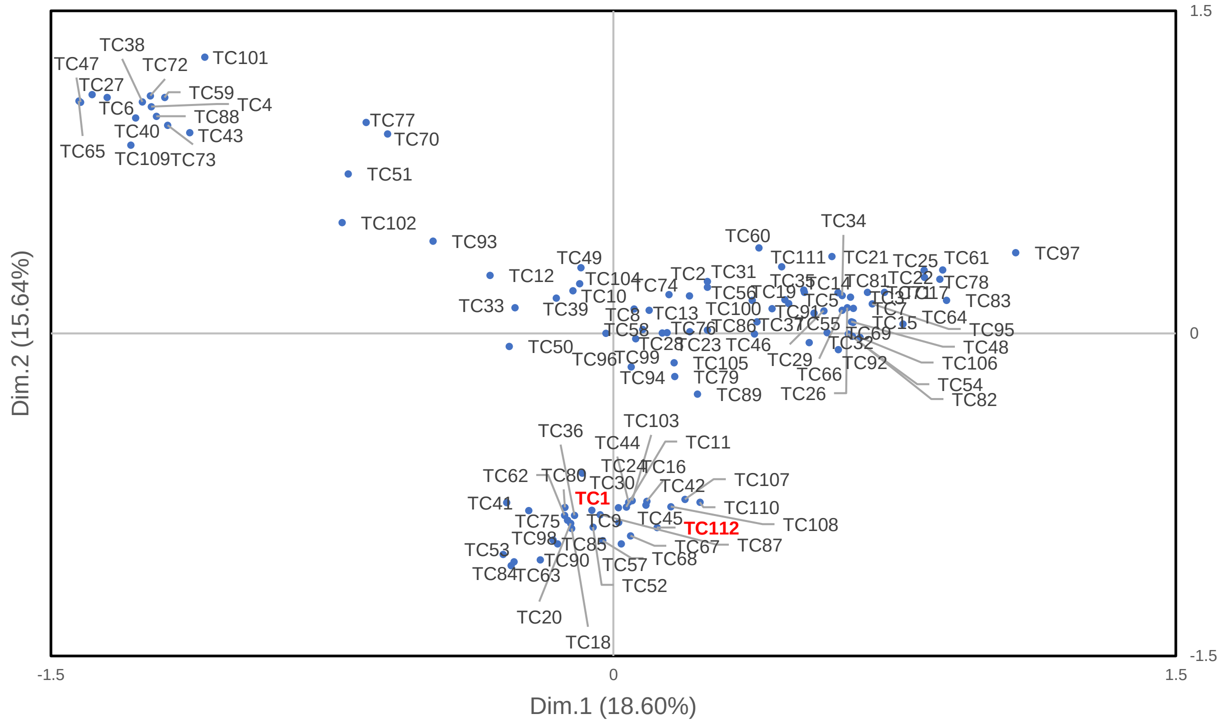


Figure 6.10 Factor map of trip chains

The relationship among trip chains cannot be interpreted based on the proximity, as in the case of relationship among variable categories described in the previous section. Trip chains should be interpreted in conjunction with variable categories as trip chains were used as supplementary variables and are place on the factor map based on individuals and categories. Therefore, a combined map of variable categories and trip chains is shown in Figure 6.11.

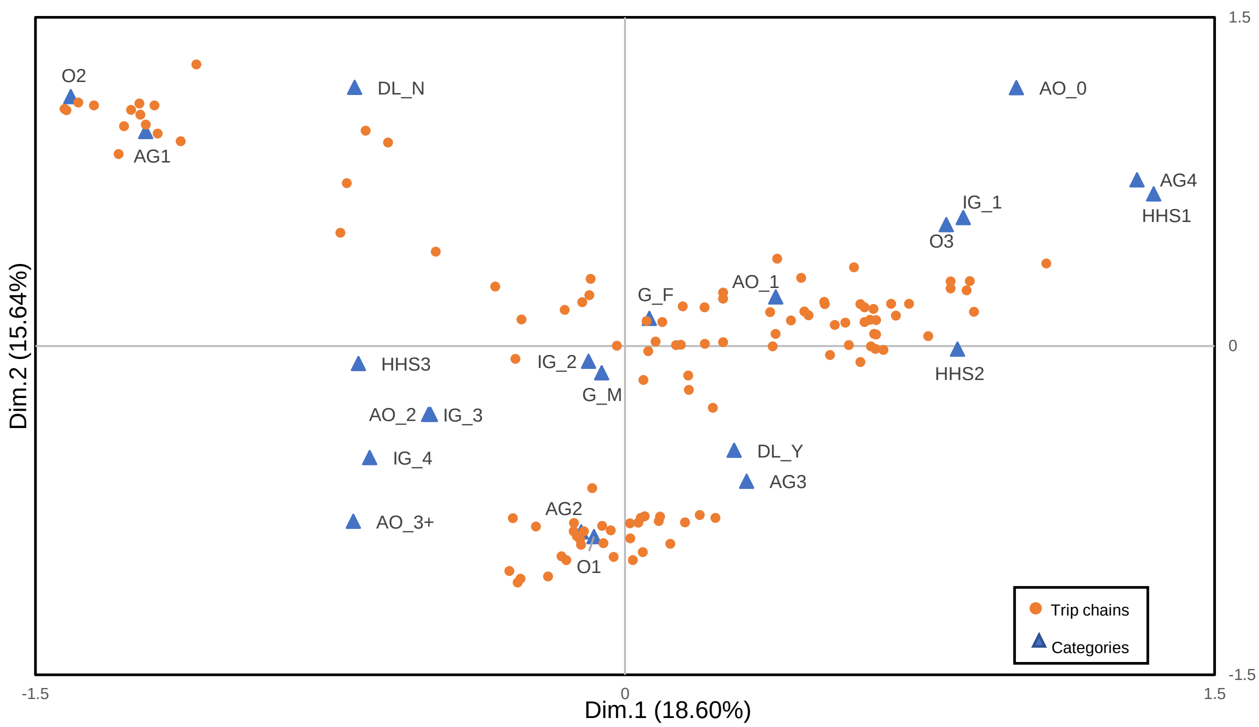


Figure 6.11 Factor map showing relationship between categories and trip chains

There are 15 trip chains that contain an education activity, and all of them are located near O2 which is an occupation category of students. Similarly, there are 31 trip chains with work activity involved and all of them are mapped near occupation category of workers (O1) as shown in centre bottom of Figure 6.11 above. Rest of the trip chains are mostly in the centre which shows that these trips chains can be performed by any occupation category. It is also interesting to note that trip chain TC1, which is the most occurring trip chain in the dataset is very close to trip chain TC112, which is the least occurring trip chain in the dataset as shown bold in Figure 6.10. It is probably due to the fact that both contains work activity. From Figure 6.11, it can be inferred that people with the occupation category of other (O3), are more likely to perform trip chains that does not contain either work or education category.

Individuals

The relationship among individuals can also be studied with the help of multiple correspondence analysis. The parameters for individuals, as described for variable categories are attached in Appendix B.1.3. Their interpretation can also be made in a similar way as of categories. However, those parameters are not of interest here. Figure 6.12 shows that individuals are well distributed over the whole map (in the first two dimensions), which shows variability in trip chain choice. The green ellipse shows people who mostly selected education related trip chains while red ellipse shows people who chose mostly work-related trip chains. If it is analysed together with Figure 6.9,

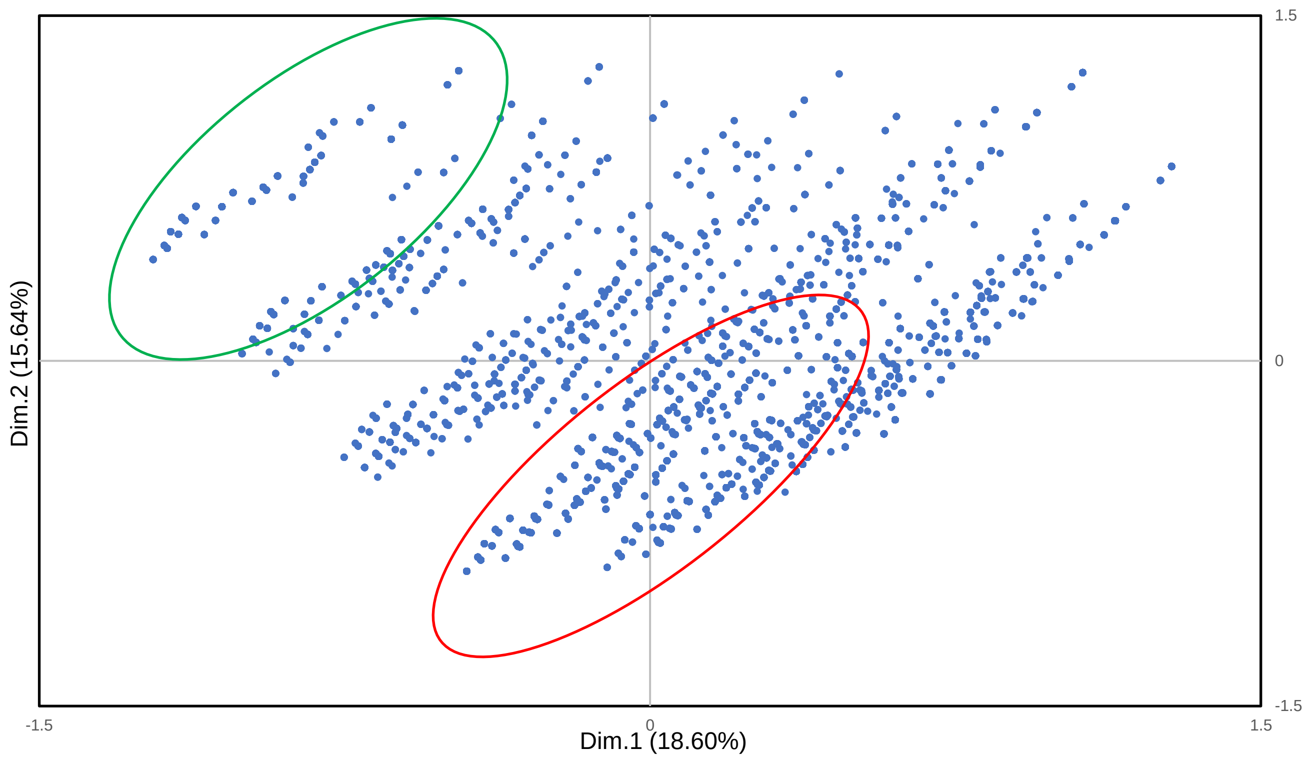


Figure 6.12 Factor map of individuals

it shows that green ellipse shows students, while red ellipse shows workers. However, there were some people found to be students and chose work trips. It is possible as a student can also work part time, so he/she made a work trip. Similarly, some workers were found making education trips. It is also realistic as someone can be doing part time education program (such as PhD).

By analysing Figure 6.9, Figure 6.10 and Figure 6.12 together, it shows that model has performed well in explaining the travel behaviour of people based on their socioeconomic variables.

6.4 Predictions from MCA

Multiple correspondence analysis is a descriptive technique used to analyse relationships in the data. To the extent of the knowledge of author, there is no definite procedure to perform predictions from MCA results. *FactoMineR* package has a function "*predict.mca"* which can be used to predict the coordinates of new individuals, but not the trip chains. This study intends to develop a model that can estimate the trip chains for population. Therefore, a procedure was developed to estimate the trip chain of individuals given their socioeconomic characteristics, based on MCA results.

6.4.1 Conceptual framework

The prediction model is based on the similar concept of the utility maximization theories. An individual will select a trip chain that would maximize his utility. The utility for an individual to select a trip chain depends upon the impedance or distance between individual and trip chain because the points that are close together in MCA dimensions are more related to each other.

The distance calculation is based on the coordinates in MCA dimensions. The coordinates of trips chains are estimated by MCA, while the coordinates of new individuals can be predicted. For an individual with a set of variable categories, the coordinates in all dimensions can be predicted based on developed MCA model, by using the following function from *FactoMineR* package.

Equation 6.2

$$predict.mca(object, newdata)$$

where;

$object$ = MCA model
$newdata$ = characteristics of persons to be predicted

Once the coordinates of individuals are predicted, the distance between individual and all trip chains can be calculated using Equation 6.3.

Equation 6.3

$$d_{ij} = \sqrt{\sum_{k=1}^{n} \left[\left(I_i - TC_j \right)_k^{\,2} \times w_k \right]}$$

Where;

d_{ij} = distance between individual i and trip chain j

I_i = coordinate value of individual i in k^{th} dimension
TC_j = coordinate value of trip chain j in k^{th} dimension
w_k = weight of k^{th} dimension (equal to percentage of variance explained by dimension k)
k = dimension number
n = total number of dimensions

Not all the dimensions explain same amount of variability, which implies that some dimensions are more important than the other ones. Therefore, a weighted factor (w_k) is included in Equation 6.3 to have a weighted calculation of distance. The distance should be computed considering all the dimensions in order to capture the whole variability of data. By using Equation 6.3, a skim matrix can be formed between individuals and trip chains.

Calculation of Utility

The utility of a trip chain depends upon the impedance or distance between individual and trip chain. The utility for a trip chain closer to the individual is more as compared to the trip chain that is far from the individual (in terms of distance). For this purpose, exponential function is used. Given an individual i, the utility of each trip chain j can be calculated by Equation 6.4 below:

Equation 6.4

$$U_{ij} = e^{-\beta \times d_{ij}}$$

Where;

U_{ij} = utility of individual i for trip chain j
d_{ij} = distance between individual i to trip chain j
β = estimated parameter

By using Equation 6.4, a utility matrix can be formed between individuals and trip chains.

Calculation of Probabilities

Based on the calculated utilities, the probability of individual i choosing trip chain j can be calculated using Equation 6.5 below:

Equation 6.5

$$P_{ij} = \frac{e^{u_{ij}} \times f_j}{\sum_{j=1}^{n} \left(e^{u_{ij}} \times f_j\right)}$$

Where;

P_{ij} = probability of individual i choosing trip chain j
u_{ij} = utility of individual i for trip chain j
f_j = relative frequency of trip chain j as observed in the survey data
n = number of trip chains

In Equation 6.5, the factor f_j, which is the relative frequency of trip chain j as observed in the survey data is multiplied to compute the probability. It is because two trip chains can be at the same distance from the individual based on the characteristics, but one trip chain could be more frequent then the other one. It indicates that the frequent trip chain should have more chances to be selected as compared to the other one. For example, trip chain TC1(H-W-H) is the most occurring trip chain, while TC112 (H-W-H-S-H-PE-H) is the least occurring trip chain in the dataset. From Figure 6.10, it can be seen that both are placed closer to each other. They are placed correctly as both have 'work' activity involved. From a worker's perspective, it does not matter if the worker selects TC1 or TC112, but the fact that from choice set, there are more TC1 available as compared to TC112, the worker should select more often TC1 rather than TC112. For the trip chains that are not modeled in MCA, they can be grouped together. The group can be assigned a probability value and the selection of trip chain from that group can be made randomly without assigning specific probabilities. However, this part is not completed in the current thesis, but can be implemented later.

6.5 Model Calibration

The model was calibrated based on top 10 most occurring group of individuals. For that purpose, the coordinates of the individuals that were used to build the MCA model were predicted using Equation 6.2. Given the coordinates of trip chains from MCA model, the distance between individuals and trip chains was calculated using Equation 6.3. The β parameter for utility function in Equation 6.4 was adjusted to minimize the error. Coefficient of determination (R^2) and root-mean-square-error (RMSE) were used for calibrating the parameter β. The value of β was used such that it maximizes R^2 and minimizes RMSE.

Table 6.5 Combination of individuals used for model calibration

Combination	Description	Frequency
Comb1	HHS2, AO_1, AG4, G_M, O3, IG_1, DL_Y	881
Comb2	HHS3, AO_2, AG2, G_M, O1, IG_3, DL_Y	544
Comb3	HHS2, AO_1, AG4, G_M, O3, IG_2, DL_Y	543
Comb4	HHS3, AO_2, AG2, G_M, O1, IG_2, DL_Y	464
Comb5	HHS3, AO_2, AG1, G_M, O2, IG_3, DL_N	450
Comb6	HHS2, AO_1, AG4, G_F, O3, IG_1, DL_Y	441
Comb7	HHS3, AO_2, AG2, G_F, O1, IG_3, DL_Y	421
Comb8	HHS3, AO_2, AG1, G_F, O2, IG_3, DL_N	419
Comb9	HHS3, AO_1, AG2, G_M, O1, IG_2, DL_Y	401
Comb10	HHS3, AO_1, AG1, G_M, O2, IG_2, DL_N	387

A total of 899 combinations of individuals were found from the data. Combination represents the set of variables that represents a person. The model was calibrated based on top 10 combinations. These 10 combinations along with their description and frequency of occurrence and are shown in Table 6.5 above. For instance, in Table 6.5, 'Comb1' is the combination 1 which represents individuals with household size category 2 (2 persons in household), auto ownership category 1 (owns one car), age group category 4 (age 65 and above), male person with occupation category of 3 (others), income group 1 (0 < 2000 euros) and owning a driving license. There are 881 individuals in the dataset with these characteristics.

Using these combinations, the probabilities of all trip chains were calculated using Equation 6.5 by varying β parameter for utility function in Equation 6.4. These probabilities were compared with the actual probabilities from the dataset. Table 6.6 shows results of some of the iterations of β parameter.

Table 6.6 Adjustment of β parameter

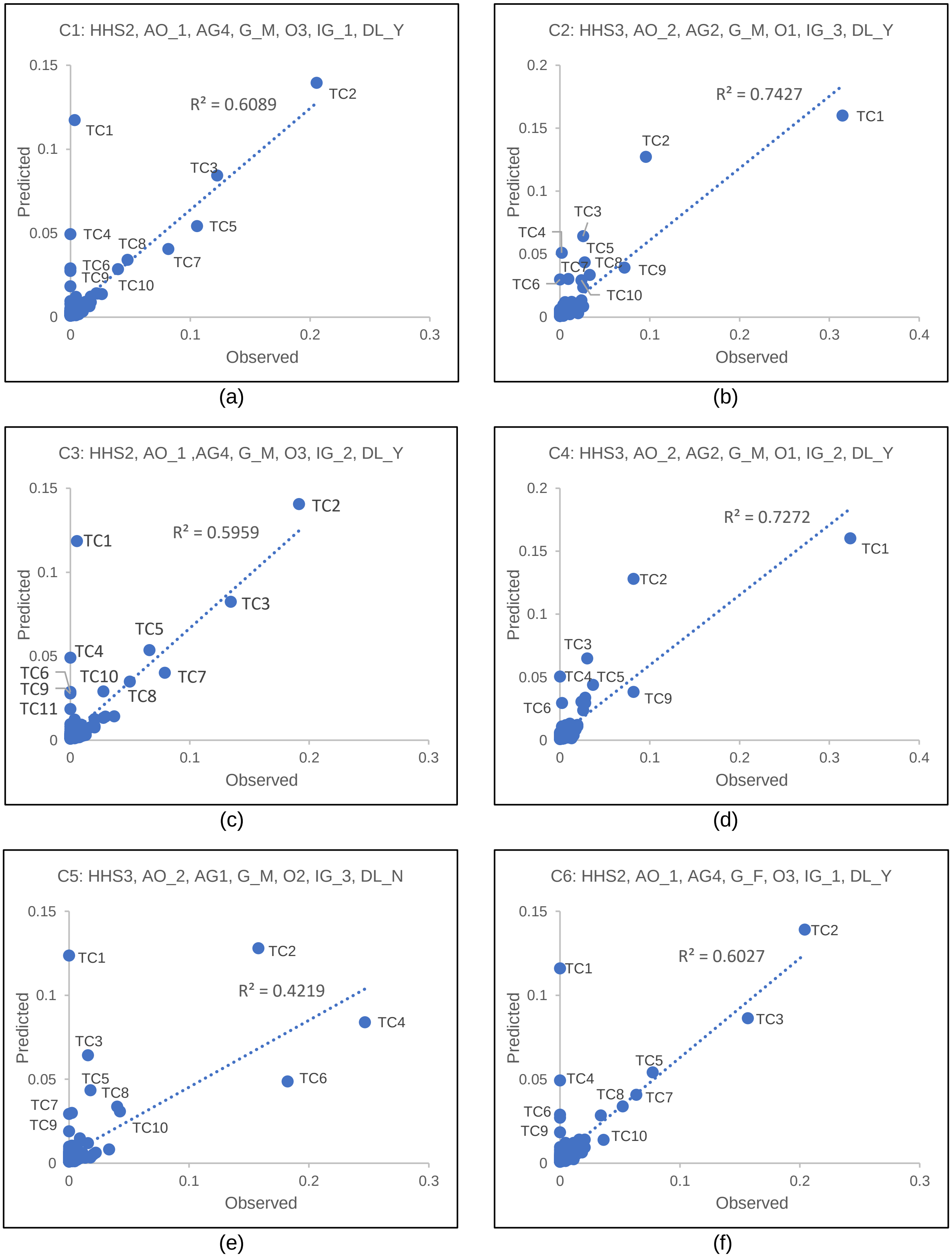
$\boldsymbol{\beta}$	$\boldsymbol{R^2}$	**RMSE**
-0.001	0.461990	10.39882
-0.01	0.477323	10.24387
-0.05	0.527718	9.717681
-0.1	0.562647	9.336483
-0.15	0.578694	9.156441
-0.2	0.583401	9.103018
-0.25	0.581268	9.127252
-0.3	0.575123	9.196751

The R^2 and RMSE is the average of the 10 combinations. The optimum value of β was estimated to be -0.2. The result with the optimum value of β for the combinations are shown in Table 6.7

Table 6.7 Estimated results with β = -0.2

Combinations	$\boldsymbol{R^2}$	**RMSE**
comb1	0.61	14.23
comb2	0.74	9.39
comb3	0.60	8.44
comb4	0.73	8.39
comb5	0.42	11.11
comb6	0.60	7.28
comb7	0.83	3.65
comb8	0.34	10.84
comb9	0.71	6.96
comb10	0.37	10.73

Figure 6.13 below depicts graphically, the observed vs predicted trip chains for the combinations.



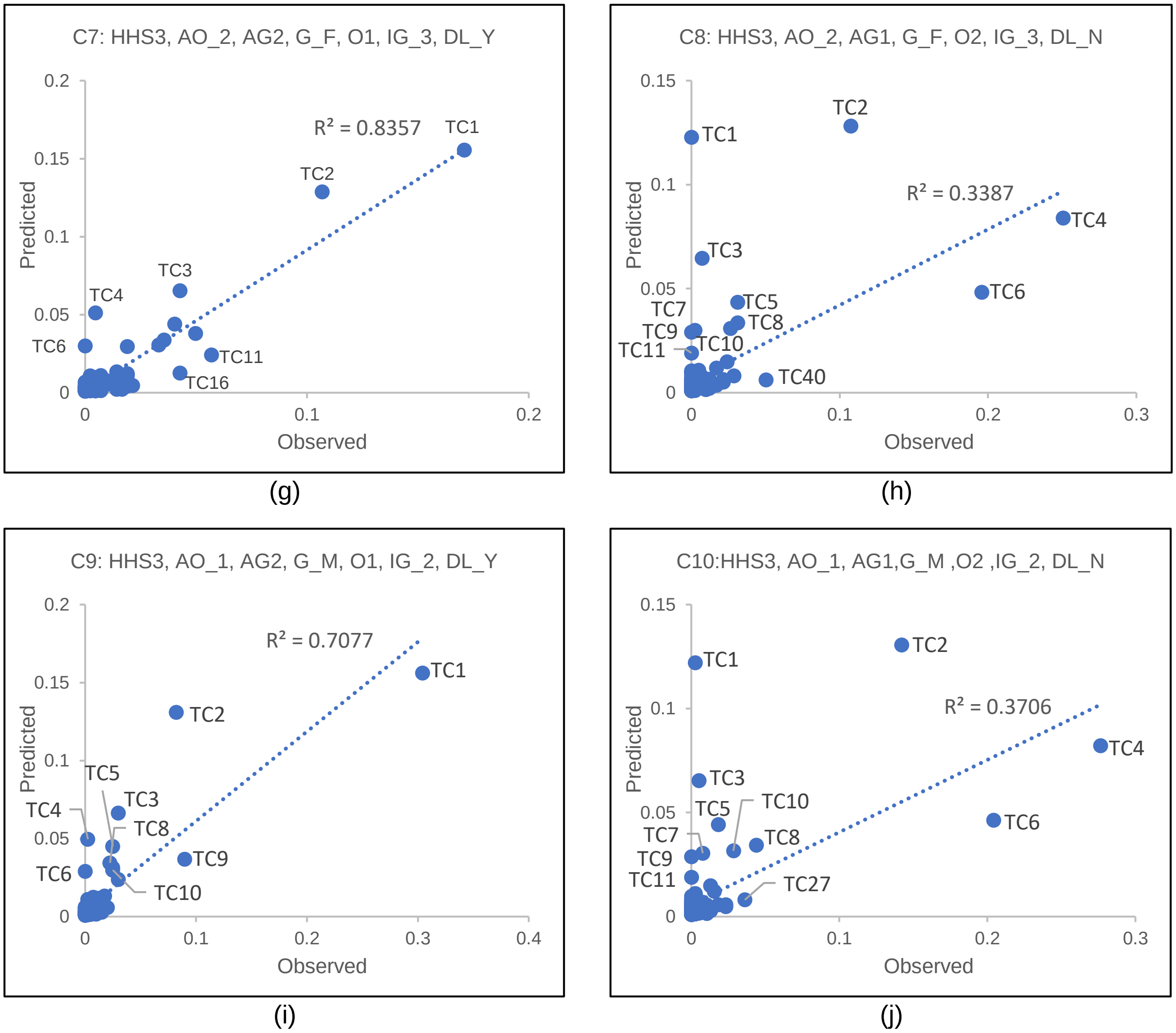


Figure 6.13 (a-j) Model calibration results for top 10 combinations

Figure 6.13 also points the outliers. Both Table 6.7 and Figure 6.13 show that the model performs well for workers (comb 2,4,7,9). For workers, the outliers mainly involve the trip chains that are related to education. For example, combination 2, 4, 7 and 9 has a common outlier of TC4 (H-E-H) and TC6 (H-E-H-L-H). Similarly, for students, the trip chain which is not predicted well at all is TC1 (H-W-H). For occupation category of others, both work and education trips are not predicted well. It is important to note here that the pointed trip chains (outliers) are only a few among the 31 work trip chains and 15 education trip chains. But their frequency is more, as TC1 is the most occurred while TC112 is the least. It shows that most work and education trips are well predicted, besides the outlier. The fact that this data was captured for only one day, so a person who chose a particular trip chain on that day could also choose any other trip chain the other day. Another important point is that all the trip chains, that are modeled, are not observed for every combination. For example, the dataset shows 62 trip chains observed for individuals belonging to combination 1, while the model predicts from 112 trip chains. Similarly, 47 trip chains are observed for combination 10 while 112 are considered while predicting by the model. This could also have an impact on the graphs shown in Figure 6.13. The overall model performance is shown in Figure 6.14. It only represents the overall number of trip chains observed versus predicted by the population.

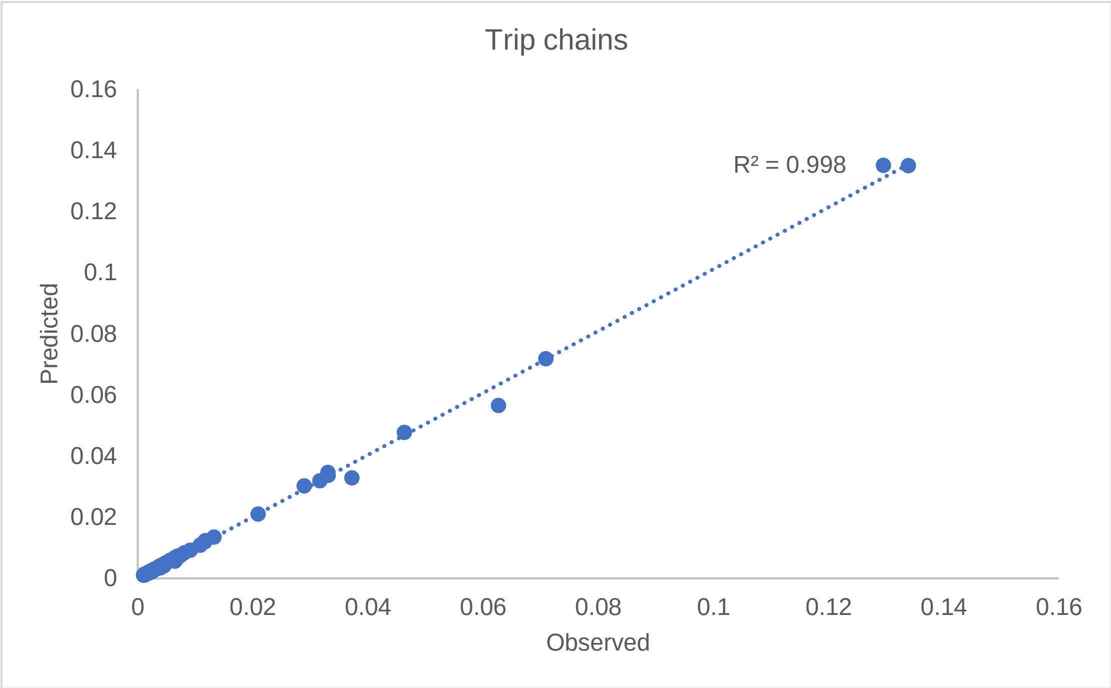


Figure 6.14 Observed versus estimated trip chains

It shows that the model has performed well overall for whole of population. This graph is influenced by the factor f_j (relative frequency of trip chain j as observed in the survey data) in Equation 6.5. However, it is important to apply this factor, as to have an effect of the number of trip chain choices an individual must select from. MCA can give valuable results describing the relationship between individuals and trip chains but not the frequency of trip chain.

6.6 Model application example

The previous sections described modeling structure intended to model trip chains. In this section, the application of the model is depicted with the help of an example. A sample simulation is provided which shows how the model can be used to predict trip chains for individuals. For this purpose, a hypothetical individual is considered. The process steps are shown with the help of flow chart in Figure 6.15 below:

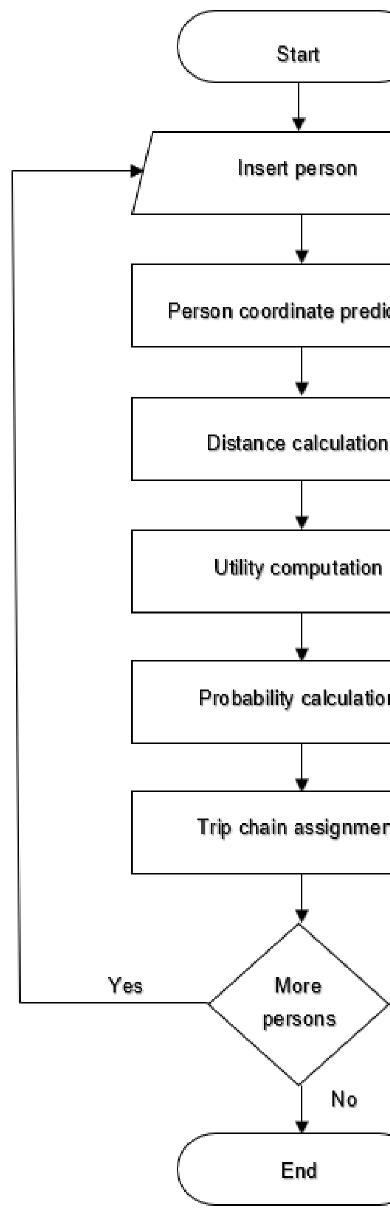


Figure 6.15 Flow diagram of trip chain assignment

The characteristics of the considered individual is as *HHS2, AO_1, AG2, G_M, O1, IG_2 AND DL_Y.* The steps shown in Figure 6.15 were followed to calculate probabilities. The model was run 100 times. The predicted trip chains along with the frequency are shown in Table 6.8 below:

Table 6.8 Example result

Sr no.	Trip chains	Description	Predicted no.
1	TC1	H - W - H	16
2	TC2	H - L - H	12
3	TC9	H - W - H - L - H	7
4	TC11	H - W - H - S - H	5
5	TC5	H - PE - H	4
6	TC8	H - L - H - L - H	4
7	TC15	H - S - H - S - H	4
8	TC7	H - S - H - L - H	3
9	TC10	H - L	3
10	TC18	H - W - H - PE - H	3
11	TC25	H - PE - S - H	3
12	TC4	H - E - H	2
13	TC6	H - E - H - L - H	2
14	TC14	H - PE - H - L - H	2
15	TC16	H - W - S - H	2
16	TC19	H - S - S - H	2
17	TC33	H - L - H - L	2
18	TC42	H - W - PE - H	2
19	TC70	H - AC - AC - H	2
20	TC90	H - W - H - PE - H - L - H	2
21	TC3	H - S - H	1
22	TC13	H - L - L - H	1
23	TC20	H - W - H - W - H	1
24	TC27	H - E - H - PE - H	1
25	TC29	H - PE - H - PE - H	1
26	TC30	H - W - L - H	1
27	TC32	H - S - H - L - H - L - H	1
28	TC36	H - W	1
29	TC48	H - S - H - S - H - L - H	1
30	TC55	H - PE - PE - H	1
31	TC75	H - W - H - W - H - L - H	1
32	TC80	H - L - H - W - H	1
33	TC83	H - PE - H - S - H - PE - H	1
34	TC86	H - PE - H - L	1
35	TC95	H - S - H - S - H - S - H	1
36	TC96	H - L - H - L - L - H	1
37	TC98	H - AC - H - W - H	1
38	TC102	H - E - S - H	1

Trip chains are selected through Monte Carlo random sampling based on calculated probabilities. The result showed that mostly trip chain TC1 and TC2 is predicted, because both have high probabilities. Others have low numbers as the probability is low. As the person is the worker, therefore, first 11 predicted trip chains don't have education trip and the most predicted trip chains have work activity involved. Similarly, the model can be used for any type of person combination to obtain trip chains.

7 Conclusion

This thesis attempted to model trip chains from household travel survey data (MiD 2008). The goal was to build a model that can input a person with socioeconomic characteristics and return a trip chain. Trip chains were generated from MiD 2008. 5136 unique trips chains were generated form 39080 persons. 112 unique trip chains from 28172 persons were selected for modeling in MCA because the number of records for these trips chains were 30 or more and 72% of the dataset was represented. Household and person variables and categories were defined based on the sate-of-the-art practice in transportation modeling and data analysis.

MCA model resulted in the formation of 15 principal dimensions. The factor map with dimension 1 and 2 clearly showed patterns and explained choice of trip chain with respect to socioeconomic characteristics. No definite method was found to use MCA results for prediction. Therefore, a prediction framework was proposed based on the distance between trip chains and persons in all 15 dimensions. The utility function was described according to which the trip chain that is closer to the person in 15 principal dimensions has more utility. The probability of every trip chain was calculated for the person. Therefore, different persons with same definition can have different trip chains. The prediction methodology for the rest of the observed trip chains that were not modeled in MCA was also discussed. However, it was not applied in current work but can be applied later.

Model calibration has been performed based on top ten most occurring definition of individuals. The results showed that overall performance of the model has been good with R^2 value of 99.8%. The model performed best for workers, then for others category and finally for students. The demonstration shows that multiple correspondence analysis is applicable for modeling of trip chains. However, the methodology also has some limitations which are discussed in the following section. The recommendations for future work are also listed in section 7.2

7.1 Limitations

- The model was built at individual level. Therefore, intrahousehold interactions are lost or ignored. The *accompanying* trips are difficult to explain. For e.g., if one member of a household has *accompanying* activity, it is difficult to say which household member the person is traveling with.
- *Business/service* activities were removed from trip chain. Hence, the model is not able to represent business trips.
- The data for trip chains were observed for a day. So, it was assumed that the trip chain represents average week day. In reality, trips chains are dependent on days as well. For example, if a person does grocery shopping today, it is more like that he will not go for grocery shopping the next day.

- Multiple correspondence analysis (MCA) gives an idea about the relationship of trip chain with the socioeconomic characteristics by the calculation of distance. It does not explicitly provide any coefficients to show which variable has more weight or is positively or negatively effecting as in case of regression analysis.
- Predictions cannot be performed from multiple correspondence analysis, therefore an additional framework had to be designed for that purpose.
- The validation of the model is difficult to perform because the new dataset can have entirely different set of trip chains.
- 112 trip chains were used for modeling through MCA, rest of the trip chains were grouped in a single cluster. For those trip chains the relationship with socioeconomic characters might not have been valid.

7.2 Future work

- Inclusion of more trip chains in MCA model can be made possible with the availability of more sufficient data for trip chains. However, the methodology proposed for the trip chains that were not modeled can also be implemented.
- MCA results can further be tested for significance by applying different measure of fit methods as suggested by (Greenacre, 2006), (Greenacre and Blasius, 2006) and (Asan and Greenacre, 2008).
- The time of activities in trip chains can be modeled through some other technique to better represent the travel behaviour.
- Integration of the model to SILO-MATSim is the next step to implement the model practically.

8 References

Adler, T., Ben-Akiva, M., 1979. A theoretical and empirical model of trip chaining behavior. Transp. Res. Part B Methodol. 13, 243–257.

Asan, Z., Greenacre, M.J., 2008. Measures of fit in multiple correspondence analysis of crisp and fuzzy coded data. Econ. Bus. Work. Pap. Ser. 13.

Badoe, D.A., Chen, C.-C., 2004. Unit of analysis in conventional trip generation modelling: an investigation. Can. J. Civ. Eng. 31, 272–280. https://doi.org/10.1139/l03-098

Bäumer, M., 2017. German national travel survey “mid 2016 – mobility in germany”: new challenges – new approaches 1–12.

Bhat, C., 2001. Modeling the commute activity-travel pattern of workers: formulation and empirical analysis. Transp. Sci. 35, 61–79.

Bhat, C.R., Koppelman, F.S., 1999. Activity-based modeling of travel demand. Int. Ser. Oper. Res. Manag. Sci. 35–62.

Bhat, C.R., Koppelman, F.S., 1993. A conceptual framework of individual activity program generation. Transp. Res. Part A Policy Pract. 27, 433–446.

Bowman, J.L., 1998. The day activity schedule approach to travel demand analysis. Massachusetts Institute of Technology.

Bowman, J.L., Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with activity schedules. Transp. Res. Part A Policy Pract. 35, 1–28.

Castiglione, J., Bradley, M., Gliebe, J., 2015. Activity-Based Travel Demand Models: A Primer. https://doi.org/10.17226/22357

Cluster Analysis, 2007. , in: Applied Multivariate Statistical Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 271–288. https://doi.org/10.1007/978-3-540-72244-1_11

Cyganski, R., Justen, A., Schulz, A., Köhler, K., 2013. Generation of coherent trip chains for travel behaviour analysis and modelling, in: European Transport Conference.

Downes, J.D., Johnsen, S., Morrell, D., 1978. Household and Person Trip Generation Models.

Gangrade, S., Pendyala, R.M., McCullough, R.G., 2002. A nested logit model of commuters’ activity schedules. J. Transp. Stat. 5, 19–36.

Golob, T.F., 2000. A simultaneous model of household activity participation and trip chain generation. Transp. Res. Part B Methodol. 34, 355–376.

Golob, T.F., 1986. A nonlinear canonical correlation analysis of weekly trip chaining behaviour. Transp. Res. Part A Gen. 20, 385–399.

Greenacre, M., 2006. Tying up the loose ends in simple, multiple, joint correspondence analysis. Compstat 2006-Proceedings Comput. Stat. 163–185.

Greenacre, M., Blasius, J., 2006. Multiple correspondence analysis and related methods. CRC press.

Hägerstraand, T., 1970. What about people in regional science? Pap. Reg. Sci. 24, 7–24.

Härdle, W.K., Simar, L., 2015. Applied Multivariate Statistical Analysis, Fourth. ed. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45171-7

Heggie, I.G., Jones, P.M., 1978. Defining domains for models of travel demand. Transportation (Amst). 7, 119–125. https://doi.org/10.1007/BF00184635

Huang, Z., 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Min. Knowl. Discov. 2, 283–304. https://doi.org/10.1023/A:1009769707641

Husson, F., 2016. Mutltiple Correspondence Analysis (Part 1/4: Data - issues) [WWW Document]. URL https://www.youtube.com/watch?v=gZ_7WWEVlTg (accessed 11.30.17).

Husson, F., Lê, S., Pagès, J., 2017. Exploratory multivariate analysis by example using R. Chapman and Hall/CRC.

Jiang, S., Ferreira, J., González, M.C., 2012a. Clustering daily patterns of human activities in the

city. Data Min. Knowl. Discov. 25, 478–510.
Jiang, S., Ferreira, J., González, M.C., 2012b. Clustering daily patterns of human activities in the city. Data Min. Knowl. Discov. 1–33.
Jones, P.M., 1977. New Approaches to Understanding Travel Behavior: The Human Activity Approach. Oxford University.
Kitamura, R., 1988. An evaluation of activity-based travel analysis. Transportation (Amst). 15, 9–34.
Kitamura, R., Chen, C., Pendyala, R.M., Narayanan, R., 2000. Micro-simulation of daily activity-travel patterns for travel demand forecasting. Transportation (Amst). 27, 25–51.
Lê, S., Josse, J., Mazet, F., 2008. Package ‘ FactoMineR .’ J. Stat. Softw. 25, 1–18. https://doi.org/10.1007/978-3-540-74686-7
Mcnally, M.G., 2007. The Four Step Model. Handb. Transp. Model. 35–52. https://doi.org/https://escholarship.org/uc/item/0r75311t
Mladenovic, M., Trifunovic, A., 2014. The Shortcomings of the Conventional Four Step Travel Demand Forecasting Process. J. Road Traffic Eng.
Moeckel, R., Huntsinger, L.F., Donnelly, R., 2015. Microscopic trip generation: Adding fidelity to trip-based travel demand models, in: International Conference on Computers in Urban Planning and Urban Management.
Ordóñez Medina, S.A., 2015. Recognizing personalized flexible activity patterns, in: 14th International Conference on Travel Behavior Research (IATBR 2015).
Ort, D., Willumsen, L.G., 2011. MODELLING TRANSPORT, 4th ed. WILEY.
Pas, E.I., 1984. The effect of selected sociodemographic characteristics on daily travel-activity behavior. Environ. Plan. A 16, 571–581.
Pas, E.I., 1984. The effect of selected sociodemographic characteristics on daily travel-activity behavior. Environ. Plan. A 16, 571–581. https://doi.org/10.1068/a160571
Pas, E.I., 1983. A flexible and integrated methodology for analytical classification of daily travel-activity behavior. Transp. Sci. 17, 405–429.
Pas, E.I., 1983. A flexible and integrated methodology for analytical classification of daily travel-activity behavior. Transp. Sci. 17, 405–429. https://doi.org/10.1287/trsc.17.4.405
Pinjari, A.R., Bhat, C.R., others, 2011. Activity-based travel demand analysis. A Handb. Transp. Econ. 10, 213–248.
Pokhrel, A., 2017. Comparative Analysis of Person-Category and Household-Category Trip Generation Models. Technical University of Munich.
Ralambondrainy, H., 1995. A conceptual version of the K-means algorithm. Pattern Recognit. Lett. 16, 1147–1157. https://doi.org/https://doi.org/10.1016/0167-8655(95)00075-R
Srinivasan, N., Santos, A., 2011. Uses of National Household Travel Survey Data in Transportation.
Strathman, J.G., Dueker, K.J., 1995. UNDERSTANDING TRIP CHAINING. CHAPTER 1, VOLUME III SPECIAL REPORTS ON TRIP AND VEHICLE ATTRIBUTES-IN: 1990 NPTS SPECIAL REPORTS.
Strathman, J.G., Dueker, K.J., Davis, J.S., 1994a. Effects of household structure and selected travel characteristics on trip chaining. Transportation (Amst). 21, 23–45.
Strathman, J.G., Dueker, K.J., Davis, J.S., 1994b. Effects of household structure and selected travel characteristics on trip chaining. Transportation (Amst). 21, 23–45. https://doi.org/10.1007/BF01119633
Supernak, J., Talvitie, A., Dejohn, A., 1982. Person-Category Trip-Generation Model 74–83.
Trivedi, S., Pardos, Z.A., Heffernan, N.T., 2015. The Utility of Clustering in Prediction Tasks. arXiv 1–11.
Wen, C.-H., Koppelman, F.S., 2000. A conceptual and methdological framework for the generation of activity-travel patterns. Transportation (Amst). 27, 5–23.

Appendix A.1 - List of 112 trip chains modeled

Trip Chains	TC code	Freq.
H - W - H	TC1	3772
H - L - H	TC2	3649
H - S - H	TC3	1997
H - E - H	TC4	1765
H - PE - H	TC5	1305
H - E - H - L - H	TC6	1048
H - S - H - L - H	TC7	933
H - L - H - L - H	TC8	931
H - W - H - L - H	TC9	892
H - L	TC10	816
H - W - H - S - H	TC11	590
H - AC - H	TC12	374
H - L - L - H	TC13	332
H - PE - H - L - H	TC14	329
H - S - H - S - H	TC15	328
H - W - S - H	TC16	307
H - S - H - PE - H	TC17	306
H - W - H - PE - H	TC18	258
H - S - S - H	TC19	228
H - W - H - W - H	TC20	210
H - S	TC21	208
H - PE - H - S - H	TC22	200
H - L - H - L - H - L - H	TC23	196
H	TC24	193

Trip Chains	**TC code**	**Freq.**
H - PE - S - H	TC25	187
H - L - H - S - H	TC26	185
H - E - H - PE - H	TC27	182
H - L - L - L - H	TC28	182
H - PE - H - PE - H	TC29	182
H - W - L - H	TC30	166
H - PE	TC31	162
H - S - H - L - H - L - H	TC32	159
H - L - H - L	TC33	153
H - S - L - H	TC34	140
H - L - S - H	TC35	138
H - W	TC36	137
H - L - H - PE - H	TC37	134
H - E - L - H	TC38	130
H - AC - H - AC - H	TC39	127
H - E - H - L	TC40	126
H - W - H - L	TC41	123
H - W - PE - H	TC42	123
H - E - H - S - H	TC43	119
H - S - H - W - H	TC44	119
H - W - L - W - H	TC45	114
H - S - H - L	TC46	113
H - E - H - L - H - L - H	TC47	111
H - S - H - S - H - L - H	TC48	111
H - L - L	TC49	110

Trip Chains	TC code	Freq.
H - AC - H - L - H	TC50	105
H - AC	TC51	99
H - W - H - S - H - L - H	TC52	97
H - W - H - AC - H	TC53	94
H - L - H - S - H - L - H	TC54	84
H - PE - PE - H	TC55	83
H - PE - L - H	TC56	81
H - W - S - W - H	TC57	80
H - L - L - L	TC58	78
H - E - H - AC - H	TC59	75
H - S - L	TC60	75
H - S - PE - H	TC61	75
H - W - H - L - H - L - H	TC62	74
H - AC - W - H	TC63	73
H - S - H - PE - H - L - H	TC64	72
H - E - H - L - L - H	TC65	70
H - S - S - S - H	TC66	69
H - W - S - H - L - H	TC67	69
H - W - PE - W - H	TC68	68
H - S - S - H - L - H	TC69	67
H - AC - AC - H	TC70	64
H - PE - H - S - H - L - H	TC71	64
H - E	TC72	62
H - E - H - E - H	TC73	62
H - L - L - L - L - H	TC74	60

Trip Chains	TC code	Freq.
H - W - H - W - H - L - H	TC75	60
H - L - H - AC - H	TC76	56
H - AC - AC	TC77	55
H - PE - S - H - L - H	TC78	55
H - S - H - AC - H	TC79	54
H - L - H - W - H	TC80	51
H - PE - H - L - H - L - H	TC81	51
H - PE - H - PE - H - L - H	TC82	51
H - PE - H - S - H - PE - H	TC83	50
H - W - AC - H	TC84	49
H - AC - W - AC - H	TC85	48
H - PE - H - L	TC86	47
H - PE - H - W - H	TC87	47
H - E - PE - H	TC88	45
H - AC - H - S - H	TC89	43
H - W - H - PE - H - L - H	TC90	43
H - S - H - L - L - H	TC91	42
H - PE - H - PE - H - PE - H	TC92	41
H - AC - L - H	TC93	40
H - L - L - H - L - H	TC94	40
H - S - H - S - H - S - H	TC95	39
H - L - H - L - L - H	TC96	38
H - PE - S - S - H	TC97	37
H - AC - H - W - H	TC98	36
H - AC - S - H	TC99	36

Trip Chains	TC code	Freq.
H - L - PE - H	TC100	36
H - E - H - AC	TC101	34
H - E - S - H	TC102	33
H - W - H - L - L - H	TC103	33
H - L - AC - H	TC104	32
H - L - H - L - H - L - H - L - H	TC105	32
H - L - S - H - L - H	TC106	32
H - PE - W - H	TC107	32
H - S - W - H	TC108	32
H - E - H - S - H - L - H	TC109	31
W - H	TC110	31
H - PE - L	TC111	30
H - W - H - S - H - PE - H	TC112	30

Appendix B.1 - MCA results in all dimensions

ctr.	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15
HHS1	4.75292	1.50776	28.0328	1.01621	0.16447	0.00212	3.93759	0.0087	0.0083	36.4204	3.81214	5.66608	3.78871	3.43851	0.11337
HHS2	8.91346	0.00266	11.372	0.97052	2.04738	0.01258	0.00102	0.49263	0.30989	14.6679	2.6077	12.0768	3.06506	7.03974	1.65565
HHS3	9.54183	0.15754	0.53171	0.16376	0.93008	0.00497	0.46412	0.33296	0.15914	0.67303	3.78658	12.5288	4.19869	7.37447	1.24724
AO_0	2.74293	4.57593	25.2901	0.01069	1.05971	0.24765	0.06243	5.30141	11.5972	21.6221	6.40961	1.72526	9.63094	0.77719	1.2264
AO_1	2.40823	0.97981	4.91175	2.15342	2.71504	1.85822	4.31663	13.2498	10.6367	0.31465	6.32558	3.75565	0.39163	0.32911	0.01278
AO_2	3.28571	1.51573	0.44078	0.86883	9.3156	2.94333	4.24792	24.6582	0.0001	3.8374	7.8434	2.44129	1.56867	0.11051	0.18748
AO_3+	1.69691	2.69799	2.54792	25.3692	10.562	0.66779	0.50599	14.3066	15.8032	0.75937	5.02096	5.33985	2.79334	1.87605	0.14944
AG1	13.6635	10.4784	0.21064	0.08984	0.03101	0.00485	0.66893	0.03659	0.09149	0.0693	0.16497	0.79188	0.31167	0.38764	47.3887
AG2	0.11925	8.15773	5.47943	16.6948	5.75803	0.04369	9.79543	0.19446	3.20832	4.61847	0.00277	0.30322	0.03333	16.4967	2.47932
AG3	0.96675	4.56554	0.24677	8.21333	20.1071	0.38752	22.9855	0.06506	1.07087	0.63131	0.21972	3.4559	1.24086	0.933	6.71617
AG4	11.9044	4.80451	2.58616	0.96415	5.67306	1.14567	9.23901	1.07952	1.42201	1.56515	0.93238	3.44128	0.23559	26.8578	8.56963
G_F	0.0673	0.33562	1.09644	0.00103	19.0814	4.81085	13.8123	10.3544	0.19205	0.17679	0.03157	0.40438	0.08854	0.5414	0.05304
G_M	0.06454	0.32185	1.05146	0.00098	18.2986	4.61348	13.2456	9.92963	0.18417	0.16954	0.03027	0.38779	0.08491	0.51919	0.05087
O1	0.10359	14.9218	2.30019	1.09104	0.26288	0.13334	2.11584	0.03644	0.39893	2.27189	0.98794	0.43262	11.1967	16.4177	0.97451
O2	12.8565	9.97485	0.09534	0.00773	0.01675	0.05856	1.21215	0.05077	1.40602	0.09149	0.45227	21.096	19.0119	2.07554	13.5477
O3	8.52749	4.66248	2.28252	1.57384	0.24297	0.34689	5.97181	0.14308	2.44917	3.74548	2.60185	16.1665	0.50958	12.9447	2.23254
IG_1	7.40552	4.08997	5.03962	0.04663	0.00342	0.20827	2.05444	0.2656	12.8743	3.49373	35.3641	0.46372	0.03134	0.72698	0.02176
IG_2	0.09926	0.06589	5.38419	14.9097	0.07865	19.8376	0.56443	0.25305	25.3651	0.56666	0.13959	0.34176	0.00282	0.02916	0.00015
IG_3	2.71666	1.28552	0.06739	1.21009	1.02367	42.4145	1.37294	4.3316	0.02391	0.50801	13.4764	0.36931	0.01653	0.22717	0.0071
IG_4	1.32861	0.96894	0.88314	24.6351	1.58013	20.1653	3.23351	14.9073	12.6915	0.30169	8.96085	1.42956	0.03104	0.08851	0.01655
DL_N	4.87119	17.0552	0.10668	0.00648	0.74701	0.06613	0.13719	0.00157	0.07672	2.49145	0.59113	5.26159	29.7693	0.57661	9.5146
DL_Y	1.96337	6.87425	0.043	0.00261	0.30109	0.02665	0.05529	0.00063	0.03092	1.0042	0.23826	2.12073	11.9988	0.23241	3.83494

Appendix B.1

Coord.	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15
HHS1	1.345	0.69466	2.30787	-0.39992	0.15235	0.01706	-0.7198	0.03378	0.03066	-1.83326	0.52585	0.608	0.41314	0.37449	-0.05645
HHS2	0.84576	-0.0134	-0.67496	0.17946	-0.24682	-0.01907	0.00532	0.11673	-0.08604	0.53422	0.1997	0.40759	0.17063	0.24605	-0.09907
HHS3	-0.67803	-0.07989	0.11309	-0.05712	0.1289	0.00929	0.08792	-0.07436	0.04778	-0.08867	-0.18646	-0.32167	-0.15474	-0.19513	0.06662
AO_0	0.99559	1.17918	2.13592	-0.03997	-0.37681	-0.17957	-0.08831	0.81259	1.11693	1.37636	0.66439	-0.3269	-0.64182	-0.17348	-0.18093
AO_1	0.38368	0.22442	-0.38714	-0.23331	0.24806	0.20231	-0.30202	-0.52835	-0.43994	-0.06829	0.27146	-0.19837	-0.05323	-0.04643	-0.0076
AO_2	-0.49954	-0.31112	-0.12927	-0.16518	-0.51217	-0.2838	0.33396	0.80341	0.00151	-0.26582	-0.33693	0.17827	0.11875	0.02999	0.03243
AO_3+	-0.6914	-0.79944	0.59859	1.71909	1.05033	0.26036	0.22199	-1.17861	1.1512	0.22774	-0.51919	0.50779	0.30519	0.23798	0.05576
AG1	-1.22002	0.97971	-0.10703	0.06362	0.03539	0.0138	-0.15872	-0.03707	-0.05447	-0.04278	-0.05852	0.1216	0.06339	-0.06727	-0.6175
AG2	-0.1118	-0.84797	0.53547	-0.85068	0.47306	0.04062	0.59579	-0.08382	-0.31641	0.3426	0.00745	-0.07381	0.02034	0.43047	0.13855
AG3	0.30929	-0.61634	-0.1104	0.57971	-0.85889	0.11754	-0.88673	-0.04711	0.1776	-0.12307	-0.06437	-0.24211	-0.12055	0.09946	0.22156
AG4	1.3024	0.75872	-0.4289	0.23835	0.54746	-0.24253	0.67461	0.23025	0.24559	-0.23253	0.15912	0.28991	0.06303	-0.64038	0.30032
G_F	0.06193	0.12682	0.17662	-0.00492	-0.63498	0.31431	0.52166	-0.45099	-0.05708	-0.04942	0.01852	0.06285	-0.02444	-0.0575	-0.01494
G_M	-0.05939	-0.12162	-0.16937	0.00472	0.60893	-0.30142	-0.50026	0.43249	0.05474	0.0474	-0.01776	-0.06027	0.02344	0.05514	0.01433
O1	-0.07896	-0.86901	0.26288	-0.16478	-0.07659	0.05377	-0.20982	-0.02749	-0.08454	0.18208	0.10645	0.06681	0.28242	-0.3254	-0.06582
O2	-1.40981	1.13873	-0.08578	-0.02223	0.03099	0.05711	-0.25452	-0.05201	-0.25437	0.05856	0.11543	0.74769	-0.58982	0.18543	0.39332
O3	0.81749	0.5543	-0.29882	0.22584	0.08402	-0.09897	0.40223	0.06217	0.23903	-0.26677	-0.19713	-0.46602	-0.06875	0.32971	-0.11368
IG_1	0.86037	0.58632	0.50147	0.0439	0.01126	-0.08661	-0.26644	-0.09566	-0.61894	0.29098	-0.82077	0.08914	0.01926	-0.08824	0.01268
IG_2	-0.09251	-0.06911	-0.48136	-0.72905	0.05014	0.78499	-0.1297	0.08671	0.80681	-0.10883	0.04789	0.07106	0.00536	0.01641	0.00099
IG_3	-0.49487	-0.31216	-0.05507	0.21238	-0.18497	-1.17374	0.20685	-0.36686	0.02533	-0.10537	0.48116	-0.07554	-0.01328	0.04684	-0.00688
IG_4	-0.64981	-0.50887	0.37432	1.79933	0.43151	1.51963	0.59605	1.27789	-1.09578	-0.15247	0.73671	-0.27907	-0.03417	0.0549	-0.01971
DL_N	-0.6878	1.18017	0.07192	-0.01613	-0.16401	0.0481	-0.06787	0.00726	0.0471	0.24221	0.1046	-0.29596	0.58498	0.07747	0.26125
DL_Y	0.27722	-0.47568	-0.02899	0.0065	0.0661	-0.01939	0.02735	-0.00293	-0.01898	-0.09762	-0.04216	0.11929	-0.23578	-0.03122	-0.1053

Appendix B.1

v.test	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15
HHS1	63.49	32.7912	108.942	-18.8782	7.19147	0.80524	-33.9776	1.59459	1.44739	-86.538	24.8223	28.7003	19.502	17.6777	-2.66492
HHS2	103.628	-1.64154	-82.7005	21.9888	-30.2417	-2.33713	0.65159	14.3027	-10.5423	65.4559	24.469	49.9402	20.9065	30.1473	-12.1382
HHS3	-133.474	-15.7269	22.2614	-11.2442	25.3742	1.82927	17.3081	-14.6379	9.40481	-17.4544	-36.7059	-63.3221	-30.461	-38.4114	13.1151
AO_0	48.3336	57.2463	103.694	-1.9404	-18.2932	-8.71772	-4.28726	39.4493	54.2243	66.8191	32.2545	-15.8704	-31.159	-8.42209	-8.78367
AO_1	59.0077	34.5141	-59.5406	-35.8813	38.1505	31.1138	-46.4498	-81.2577	-67.6609	-10.5023	41.7486	-30.5086	-8.18665	-7.14075	-1.16832
AO_2	-63.8892	-39.7915	-16.5333	-21.1264	-65.5046	-36.2976	42.7124	102.753	0.19292	-33.9971	-43.0921	22.8004	15.1875	3.83554	4.14769
AO_3+	-38.4743	-44.4864	33.3096	95.6617	58.4475	14.488	12.3528	-65.5858	64.0604	12.6729	-28.8913	28.2569	16.9828	13.2427	3.10302
AG1	-120.149	96.4832	-10.5401	6.26505	3.48535	1.35879	-15.6308	-3.65025	-5.3643	-4.21314	-5.76335	11.9754	6.24305	-6.62476	-60.8126
AG2	-11.3009	-85.7121	54.1246	-85.9857	47.8169	4.1062	60.222	-8.47249	-31.982	34.6298	0.75255	-7.46088	2.0555	43.5116	14.0047
AG3	32.5292	-64.8229	-11.6117	60.9705	-90.3325	12.3625	-93.26	-4.95437	18.6793	-12.9434	-6.76986	-25.4635	-12.679	10.461	23.3021
AG4	107.861	62.8351	-35.5202	19.7392	45.3392	-20.0857	55.8698	19.069	20.3395	-19.2575	13.1777	24.01	5.22035	-53.035	24.872
G_F	10.1795	20.845	29.0294	-0.8082	-104.368	51.6613	85.742	-74.1265	-9.38182	-8.12358	3.04339	10.3305	-4.01694	-9.45113	-2.4561
G_M	-10.1795	-20.845	-29.0294	0.8082	104.368	-51.6613	-85.742	74.1265	9.38182	8.12358	-3.04339	-10.3305	4.01694	9.45113	2.4561
O1	-12.3194	-135.583	41.0153	-25.7094	-11.9498	8.38978	-32.7357	-4.28976	-13.1902	28.4074	16.6083	10.4232	44.0638	-50.7691	-10.2692
O2	-111.038	89.6871	-6.756	-1.75121	2.44064	4.49825	-20.0466	-4.09645	-20.0347	4.61216	9.09163	58.8886	-46.4549	14.6047	30.9785
O3	102.014	69.1709	-37.2899	28.1821	10.4851	-12.3506	50.1943	7.75783	29.8287	-33.2899	-24.5993	-58.1536	-8.57954	41.1444	-14.1862
IG_1	89.8539	61.233	52.3714	4.58484	1.1758	-9.04516	-27.8265	-9.99023	-64.6392	30.3888	-85.7181	9.30914	2.01102	-9.21586	1.32378
IG_2	-10.7397	-8.02382	-55.885	-84.6405	5.82094	91.1359	-15.0577	10.0672	93.6684	-12.6348	5.5598	8.25044	0.62237	1.90558	0.11515
IG_3	-55.6068	-35.0764	-6.18798	23.8651	-20.7847	-131.89	23.2428	-41.2226	2.84646	-11.8401	54.0667	-8.4884	-1.49209	5.26385	-0.77277
IG_4	-33.8333	-26.4948	19.4893	93.6842	22.467	79.1214	31.0338	66.5345	-57.0529	-7.93844	38.3577	-14.53	-1.77921	2.85864	-1.02634
DL_N	-73.2908	125.756	7.66328	-1.71846	-17.4763	5.12595	-7.2317	0.77363	5.01842	25.8088	11.1457	-31.5363	62.3339	8.2545	27.8384
DL_Y	73.2908	-125.756	-7.66328	1.71846	17.4763	-5.12595	7.2317	-0.77363	-5.01842	-25.8088	-11.1457	31.5363	-62.3339	-8.2545	-27.8384

Appendix B.1

cos2	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15
HHS1	0.14309	0.038169	0.42129	0.012651	0.001836	2.30E-05	0.040981	9.03E-05	7.44E-05	0.26583	0.021872	0.02924	0.013501	0.01109	0.000252
HHS2	0.3812	9.57E-05	0.24278	0.017163	0.032465	0.000194	1.51E-05	0.007262	0.003945	0.15209	0.021253	0.08853	0.015515	0.03226	0.00523
HHS3	0.63239	0.00878	0.01759	0.004488	0.022855	0.000119	0.010634	0.007606	0.00314	0.01081	0.047826	0.14233	0.032937	0.05237	0.006106
AO_0	0.08293	0.11633	0.38168	0.000134	0.011879	0.002698	0.000652	0.055243	0.104372	0.15849	0.03693	0.00894	0.034464	0.00252	0.002739
AO_1	0.1236	0.042285	0.12584	0.045702	0.051665	0.034364	0.076589	0.234383	0.162508	0.00392	0.06187	0.03304	0.002379	0.00181	4.85E-05
AO_2	0.14489	0.056206	0.0097	0.015843	0.152314	0.046769	0.06476	0.37479	1.32E-06	0.04103	0.065916	0.01845	0.008188	0.00052	0.000611
AO_3+	0.05255	0.070251	0.03939	0.324843	0.121264	0.007451	0.005417	0.152692	0.145672	0.0057	0.02963	0.02834	0.010238	0.00623	0.000342
AG1	0.51244	0.330447	0.00394	0.001393	0.000431	6.55E-05	0.008673	0.000473	0.001021	0.00063	0.001179	0.00509	0.001384	0.00156	0.131276
AG2	0.00453	0.260785	0.10399	0.262452	0.081163	0.000599	0.128738	0.002548	0.036309	0.04257	2.01E-05	0.00198	0.00015	0.06721	0.006962
AG3	0.03756	0.149161	0.00479	0.131959	0.289658	0.005425	0.308737	0.000871	0.012386	0.00595	0.001627	0.02302	0.005706	0.00388	0.019275
AG4	0.41298	0.140153	0.04479	0.013831	0.07297	0.014321	0.110803	0.012908	0.014685	0.01316	0.006164	0.02046	0.000967	0.09984	0.021959
G_F	0.00368	0.015424	0.02991	2.32E-05	0.386663	0.094739	0.260967	0.19505	0.003124	0.00234	0.000329	0.00379	0.000573	0.00317	0.000214
G_M	0.00368	0.015424	0.02991	2.32E-05	0.386663	0.094739	0.260967	0.19505	0.003124	0.00234	0.000329	0.00379	0.000573	0.00317	0.000214
O1	0.00539	0.652542	0.05972	0.023463	0.005069	0.002499	0.03804	0.000653	0.006176	0.02865	0.009791	0.00386	0.068923	0.09149	0.003743
O2	0.43767	0.285534	0.00162	0.000109	0.000211	0.000718	0.014265	0.000596	0.014248	0.00076	0.002934	0.1231	0.076605	0.00757	0.034066
O3	0.36942	0.169842	0.04936	0.028193	0.003903	0.005415	0.089435	0.002136	0.031584	0.03934	0.02148	0.12005	0.002613	0.06009	0.007144
IG_1	0.2866	0.133097	0.09736	0.000746	4.91E-05	0.002904	0.027486	0.003543	0.148317	0.03278	0.260821	0.00308	0.000144	0.00301	6.22E-05
IG_2	0.00409	0.002285	0.11086	0.254304	0.001203	0.294833	0.008048	0.003598	0.311447	0.00567	0.001097	0.00242	1.37E-05	0.00013	4.71E-07
IG_3	0.10976	0.043675	0.00136	0.020217	0.015335	0.617481	0.019177	0.060321	0.000288	0.00498	0.103767	0.00256	7.90E-05	0.00098	2.12E-05
IG_4	0.04063	0.024918	0.01348	0.311552	0.017918	0.222221	0.034188	0.157142	0.115545	0.00224	0.052228	0.00749	0.000112	0.00029	3.74E-05
DL_N	0.19068	0.561375	0.00208	0.000105	0.010842	0.000933	0.001856	2.12E-05	0.000894	0.02364	0.00441	0.0353	0.137926	0.00242	0.02751
DL_Y	0.19068	0.561375	0.00208	0.000105	0.010842	0.000933	0.001856	2.12E-05	0.000894	0.02364	0.00441	0.0353	0.137926	0.00242	0.02751