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

Comparative Analysis of Person-Category and Household-Category Trip Generation Models

Author:

Aashish Pokhrel

Supervision:

Prof.Dr.-Ing Rolf Moeckel

Date of Submission: 21st June 2017

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

of Aashish Pokhrel

Date of Issue: 22nd December 2016

Date of Submission: 21st June 2017

Topic: **Comparative Analysis of Person-Category and Household-Category Trip Generation Models**

Trip generation which is the first step of the conventional four steps model estimates the number of trips to and from a traffic analysis zone. Conventional trip generation models developed in 1960's were aggregated to zonal level and are highly flawed as they are only valid for the homogenous zones. Household level models are considered most suitable as they overcome many limitations of zonal models. At the same time, it is also argued that individual should be the basis of analysis as it is individuals that make the trips, not households and zones. The present study will perform the comparative analysis of person-category and household category trip generation models using the data collected in the household survey. Further, relevant literature on the use of personal and household level trip generation model will be reviewed. The estimation of trips and result analysis will be conducted using two widely used approaches namely multiple regression and cross-classification.

The student will present intermediate results to the mentor Prof. Dr.-Ing. Rolf Moeckel in the fifth, tenth, 15th and 20th week.

The student must hold a 20-minute presentation with a subsequent discussion at the most two months after the submission of the thesis. The presentation will be considered in the final grade in cases where the thesis itself cannot be clearly evaluated.

Prof. Dr.-Ing. Rolf Moeckel

Abstract

The evolution of the travel demand model started with aggregate travel demand models in 1960's. However, studies revealed that aggregation to zonal level are flawed as it is only valid when the zones are perfectly homogenous. Later, disaggregate models, particularly household based models, emerged in 1970's as the best unit of analysis. Arguments were made that model should be analyzed at individual level since it is individuals who make trips not households and zones.

Debates concerning the superiority of person and household unit of analysis in trip generation modeling have been undergoing since last few decades. A literature review of the past studies showed researchers are divided on this issue. Moreover, for most of the studies in which personal level was considered, travel data were collected in the household-travel-behaviour surveys. Travel survey has undergone tremendous changes. Thus, it was expected that newer travel survey data would produce different results. In this context, this thesis compared the trip production models at household and personal level using new travel survey data. For this, MiD 2008 which is the most recent travel data was selected. MiD 2008 is a comprehensive account of mobility behavior in Germany. It gives insight into daily travel behaviour of individuals and households of Germany.

The data considered for the analysis excluded the weekend survey data. Trips were categorized as home-based work, home-based education, home-based shopping, home-based other (for remaining home-based trips) and non-home-based trips. Two conventional techniques namely, multiple regression and cross-classification were used to model each of the five trip categories. Later, the benefits of these techniques were also analyzed. Socio-economic and demographic attributes of the household were employed in multiple regression of household-category models while person-category models employed both household and personal attributes. However, final estimation results of person-category models did not have any household attributes. Likewise, two-way cross-classification were carried out for both cases. In the case of household-category models, household attributes such as household income, household size, the number of cars owned by the household, unemployed members were considered. For person-category models, age group was retained as one dimension of categorisation; the other dimensions were sex, employment status, student status and ownership of the license.

Later, household-category and person-category trip production models were compared based on the coefficient of determination (R^2), RMSE, and NRMSE values. In the case of multiple regression, except home-based shopping trips, all other trips yielded higher R^2 values at the household level. In the same way, all person-category trip production models yielded lower RMSE and only home-based shopping and home-based education yielded lower NRMSE at a personal level. In addition, this trend from the multiple regression was replicated by cross-classification. Thus, with these results, the thesis concluded that neither household nor person-category could be claimed to be superior in trip production modelling.

Acknowledgement

I would like to thank all the people who contributed in the completion of the thesis. Firstly, I would like to thank my thesis supervisor Prof. Rolf Moeckel whose guidance, assistance and encouragement throughout the execution of the thesis were immense in accomplishing the thesis.

I would like to extend my sincere thanks to Michael Okrah for accepting to be a second reader of the thesis. I would also like to acknowledge Hema Sheranya Rayaprolu who provided me the data and her translated description of the data.

Special thanks go to SOS Kinderdorf International for taking care of me since childhood. I am highly indebted to them for the continuous support and trust they have shown towards me.

Finally, I express my deep gratitude towards my family especially, my mom, for providing me with unfailing support and love throughout my study.

Table of Contents

1 Introduction

1.1 Background

Trip Generation model is the crucial portion of the travel demand model as the accuracy of subsequent stages of transport modelling are highly influenced by the accuracy of trip generation output (Badoe and Steuart 1997p,267-268). There has been a continuous study in the trip generation analysis since a long time. Conventional travel demand models developed in the 1960s are based on the aggregate characteristics of geographic areas called zones which involve the estimation of trips made directly from one zone to another (Ben-Akiva, Jon et al. 2007,p.118).

In the past, the general procedure involved collection of data from simple households and aggregation of the information to the zonal level in order to develop zonal level models, and the major reason for the aggregation was that conventional trip distribution functions at zonal level and the output of the trip generation is used as input to the trip distribution phase (Pas 1974,p.29). However, as McFadden and Reid (1975,p.24) argued, the aggregation to the zonal data is flawed as it is only valid when the zones are perfectly homogenous. In practice, it is very difficult to demarcate the zones homogenously as zones are a mixture of all kinds of people such as rich, poor, healthy, handicapped etc. Yet, aggregate models often yielded higher R^2 value which led to many researchers trusting in it (Pas 1974,p.29). Dissatisfaction developed with this method because of the inconsistency of equations and variations of explanatory variables, their signs, and magnitude of regression coefficients explaining same dependent variables in different areas (Douglas 1973,p.54).

During early 1970’s, the household was regarded as the most appropriate unit of analysis (Ortúzar and Willumsen 2011). Despite ubiquitous use of household level as the unit of analysis in trip generation models, it was argued that it is individuals that make trips rather than households or zones (McFadden 1976,p.2). Thus, this led to the push for the individual as the unit of analysis. Contrary to this, McCarthy (1969,p.35) reported a study by Bureau of Public Roads (1965) that argued although the individual is the basic trip making unit, the magnitude of the unexplained variation inherent in the individual’s travel behaviour cause trip generation analysis at individual level impractical.

Nevertheless, as reported by Ortúzar and Willumsen (2011,p.162) , Supernak (1979) proposed person-category trip generation model as an alternative to household-category trip generation model for the first time. Arguments for person-category as a better alternative to household-category are listed below (Supernak, Talvitie et al. 1983,p.74).

1. Person-level trip generation model is compatible with other components of the four-step travel demand model system that is based on trip makers rather than on households
2. Unlike, household-based cross-classification, person-based cross-classification allows use of all important variables with manageable number of classes
3. The person-category model can be developed with much smaller sample size (10 to 40 times) than for the household-category model

4. Demographic changes can be more easily accounted for in the person-category model like some demographic variables (such as age) which are virtually impossible to define at the household level.
5. Future forecasting is easier in the case of person categories as household categories require forecasts about the household formation and family size.

Person-category model is not aloof of limitations. It has the difficulty of introducing household interaction effects, household money costs and money budgets into a person-based model. Supernak, Talvitie et al. (1983,p.74) added that it is not clear how vital these considerations are and how they can effectively be introduced in household level. It is, however, essential to further study the trip generation model at personal level before dismissing its relevance.

1.2 Research Goals

The household level has dominated the research in disaggregate trip generation modelling. On top of that, as Badoe and Chen (2004,p.273) reported a study by Pas (1978) for most of the studies in which personal level was considered, travel data were collected in the household-travel-behaviour surveys. Badoe and Chen (2004,p.275) argued that in order to model at a personal unit of analysis, when the sampling unit is household, one has to assume independence in trips made by each household member which is misleading. Further, Kitamura (1996,p.122) pointed out that since 1950s and 60s significant changes have taken place in demographic and socio-economic characteristics of households. Some important changes are the involvement of women in workforce, preference of small household size and single parent households. Hence, it is expected that person-category trip generation model developed with new travel data would perform better than earlier studies. Moreover, travel survey methods have undergone tremendous changes as new travel surveys are sampled at both household and personal level.

Thus, the main aim of the research was to compare the household-category and person-category trip generation models using MiD 2008 (travel survey data from Germany). Further, only trip production model was analysed. There were several reasons for selecting this data. MiD 2008 is the latest travel data from Germany, replication of MiD 2002 which was the first comprehensive German travel survey. Further, it is sampled at both household and individual level and gives detailed account of individual and household travel behaviour.

Specific objectives of this thesis are:

1. To examine the predictive performance of trip production models with household and personal unit of analysis and determine which one best describes the travel behaviour

2. To determine the respective benefits of use of multiple regression and cross-classification approaches in trip production models

The research was conducted using two common modelling techniques namely multiple regression and cross-classification.

1.3 Structure of the Thesis

This thesis has seven chapters. Chapter one mainly deals with the introduction, research goals, and structure of the thesis. Chapter two gives an overview of the literature review related to person-category and household-category trip generation models. Chapter three describes the data. In this chapter, household and individual characteristics of the residents in Germany are analysed. Similarly, chapter four describes the modelling methodology. Model building of both multiple regression and cross-classification are discussed. Chapter five and six sheds light in the trip production analysis by multiple regression and cross-classification respectively. All five trip categories are analysed by both methods and results are discussed. Finally, chapter seven ends with the conclusions and recommendation.

2 Literature Review

2.1 Personal-Category Vs Household-Category Trip Generation Model

Downes, Johnsen et al. (1978) combined household survey data in the Reading area in 1962 (3,836 households) and 1971 (3,370 households) to create a single data bank of over 60,000 trips and compared trip generation model at household and personal level. Separate models were formed for home-based trips made for all purposes combined and for home-based work trips. The estimation results carried out by regression approach revealed similar levels of performance in both household and personal level of analysis. However, examination of the residual error confirmed that person trip rate model was better of the two. The authors at the end made the conclusion that on the basis of statistical validity and practical utility person trip rates are better.

Badoe and Chen (2004,p.274) reported a study carried out by Pas (1978) on an empirical comparison of trip generation models at personal, household and zonal level. In the paper, the personal and household level models were found to give very satisfactory estimates of zonal level trip generation for the study area in which they were derived. Further, it says that the personal model performed marginally better in that respect. The paper, however, suggested that the further research on the use of person unit in trip generation model is required.

Supernak, Talvitie et al. (1983) presented person-category trip generation model as an alternative to household-based trip generation models using cross-classification. The study employed data from Baltimore home interview conducted in 1977 by the FHWA and from Minneapolis-St. Paul home interview data collected in 1970. The data at first were superficially cleaned. Work day records were separated from weekend records, and outliers were removed. The study classified trips into home-based origin (if the origin of the trip), home-based destination (if the destination of the trip and non-home-based trips. Trips were further divided by trip purpose as work, education, shopping, personal business and social recreational. Age, employment status, and car availability were found to be the significant contributors of person's mobility. On the other hand, household-category cross-classification found household size as the only variable that gave consistent results, and thus appeared to overshadow other variables. The authors termed this as trivial as more people means more trips, and indicated it as the inefficiency of the household category models. Further, they noted that person-category model avoids such trivialities. Other arguments for preferring person-category models were that it classifies people in a logical manner and avoid the necessity of predicting household formation. Similarly, lesser data and use of fewer categories in person-category were also argued as advantage over household category. Moreover, person-category was reasoned as a better behavioural model since traveling unit is the analysis unit.

Badoe and Chen (2004) used the 1986 household-travel-survey data from the Greater Toronto Area of Canada to estimate the forecast performance of household and person trip generation models. Two-person-trip generation models were specified: first-person trip generation model included variables that describe only the individual and variables that describe individual's household while the second-person-trip generation model included variables that describe only the individual. The comparison of trips estimated for 1996 and observed 1986 trips showed that predictions given by the household models were found to be better than either of the person trip models. On top of that, the goodness-of-fit measures also indicated that household-based model is marginally better compared to the person-based trip generation model. The paper concluded that when the data for the modelling are collected in household-behaviour-survey, the appropriate analysis unit should be household.

So far, from the limited researches reviewed, it was found that common consensus on the better unit of analysis in trip generation modelling is still not developed. As cross-classification results revealed that trip generation models at the household level are affected by the trivial variable like household size by overshadowing other effects. Moreover, the need of more categories and data for the cross-classification and its inability to incorporate behavioural variables such as age, sex have been touted as important disadvantages of trip generation model at the household level. On the other hand, it is also true that person-category trip generation models failed to excel household-category models in statistical measures in some studies. Thus, amid this continuous debate and unsolved issue, it appears that it is unlikely that one or other unit dominate in all circumstances as Heggie and Jones (1978) suggested.

Above studies used either cross-classification or multiple regression techniques to compare the performance of household-category and person-category trip generation models. Further, besides study by Supernak, Talvitie et al. (1983), no other studies evaluated the trip generation model for broadly classified trip categories. Thus, in this context, the thesis intends to use both modelling techniques for broadly classified trip purposes based on home-based and non-home-based at both household and personal level and investigate the performance of the models.

2.2 Review of Modelling Techniques

2.2.1 Multiple Regression

Regression analysis is simply termed as a method for investigating functional relationships among variables (Chatterjee and Hadi 2006). In other words, regression relates a variable y to the variables x_1, x_2, x_3,..,x_n . While the variable y is termed as the dependent variable, the variables x_1, x_2, x_3,..,x_n are known as independent variables or explanatory variables as they describe the variations in the value of y. Regression analysis helps us to generate the best surface to the observed data (Pas 1974,p.19). The best surface is described by the principle of least squares i.e. the squared sum of the residuals or deviations from the estimated line is minimised.

Usually, the relationship between y and x_1, x_2, x_3,.., x_n is assumed to be linear. Mostly, more than one independent variables are considered in the trip generation modelling. Hence, the regression is termed as multiple least square regression. For trip generation model based on linear regression, the number of trips generated by the individuals, household or zone is given by:

Equation 2.1

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n + \varepsilon \qquad \text{(Pas 1974,p.19)}$$

Where,

y = dependent variable (number of trips produced by the individual, household or zone)

x= independent variable (size of the household, number of cars, household income etc.)

a= model coefficient estimated by the regression

ε= error terms or disturbance terms of the equation. Since the model is not able to take account of all variables that influence y, ε incorporates the aggregate effect of excluded variables and random deviations

Using the principle of Least-squares, the model parameters may be estimated as:

Equation 2.2

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$ (Pas 1977,p.19)

Where, b_0, b_1,…b_n are the least square estimators of the unknown model parameters a_0, a_1,…a_n. These estimators are selected in such a way that the squared sum of deviations from the estimated regression surface is minimised.

Least-square regression techniques involve some important assumptions (Douglas and Lewis 1971). These are as follows:

- Mean and covariance of the error or the disturbance terms are zero
- The variance of the error terms is constant and their distribution is normal
- The independent variables are not correlated with one another

In late 1950's and 1960's, regression techniques were extensively used to generate equations relating urban trip volumes to land use and socioeconomic characteristics (Federal Highway Adminstration 1975,p.2). Regression analysis helped greatly in understanding travel and formed the basis for further development (Federal Highway Adminstration 1975,p.2).

Regression approach has some notable limitations. The assumption that error variance remains constant is frequently violated (Draper and Smith 1966). If the error variance is not constant, the data is said to be heteroscedastic which results in an overstatement of the accuracy of the regression (Douglas and Lewis 1971). Similarly, regression regards the number of trips as a continuous random variable although it is a discrete one (Badoe 2007,p.456). In addition, the dependent variables have the possibility to take the negative values due to the assumption of a normal distribution (Badoe 2007,p.456). This is not possible as for persons who make trips, trip rates can be measured but for those who do not make trips, the value is set zero. Furthermore, as the regression techniques simply establish a relationship between independent variable and a set of independent variables, the model does not give right insight into the trip generation (Wooton and Pick 1967,p.137).

2.2.2 Cross-classification or Category Analysis

In late 60's, cross-classification technique was forwarded as an alternative to regression analysis in the United Kingdom (Ortúzar and Willumsen 2011,p.157). The shift of emphasis from aggregated zonal analysis utilizing regression approach to a disaggregated household resulted in preferring cross-classification approach (Federal Highway Adminstration 1975,p.2). However, the preference of cross-classification over regression might have arisen due to the inability to use the regression technique in right way (Douglas 1973,p.56, Pas 1974).

Cross-classification method involves division of all households into defined categories and estimation of a trip production rate for each category which is assumed to remain constant over time (Golding 1972,p.309). The multidimensional matrix in which each dimension is represented by one of the independent variables is built to form different categories. After that, mean value of the dependent variable (i.e. trip rate) is calculated for each category. The basic assumption of the cross-classification is that trip production rates are assumed to remain unchanged over time (Ortúzar and Willumsen 2011,p.157).

Let us suppose a cross-classification based on household sizes and income groups. Then, the number of household categories will be m x n. As categories are defined, households are allocated to each cell. For each cell, the observed trips of the households that fall in that category are aggregated. Finally, the total number of trips in each cell is divided by the number of households in it to give the trip rate.

Mathematically, trips produced by the individuals of the household is given by:

Equation 2.3

$$t_p(h) = T_p(h)/H(h)$$ (Ortúzar and Willumsen 2011,p.157)

Where,

$t_p(h)$= the number of trips for purpose p made by the household members of type h

$T_p(h)$= total number of trips produced by the households of type h for purpose p

H(h)= total number of households in category h

The first application of cross-classification in trip generation analysis was in the Puget Sound Regional Transport Study by Wooton and Pick (1967). They studied three different variables: 1) household structure based on the number of employed and unemployed people in the household 2) income 3) car ownership.

The cross-classification method frees researchers of many assumptions present in regression method like no shape has to be defined between the trip rate and explanatory variable, no assumptions need to be made concerning the distribution of the error terms (Pas 1974,p.23). Further advantages of this method are: 1) raw data obtained from home interviews may be directly used 2) the effects of socioeconomic characteristics of trip makers are directly reflected and 3) human behaviour is simulated more realistically at the household level (Rengaraju and Satyakumar 1994,p.930).

Despite having numerous advantageous over regression method, cross-classification has some drawbacks as well. This method lacks statistical goodness-of-fit measures to test the significance of the various explanatory variables that influence the dependent variables (Chang, Jung et al. 2014,p.79). Likewise, with the increasing number of explanatory variables to be considered or increasing the number of classes where the variables are stratified, increases the demands made on data collected. Thus, the calculation requires large sample sizes which make it expensive and time consuming (Ortúzar and Willumsen 2011).

2.2.3 Some Previous Studies

Since the research focussed on disaggregate trip generation modelling by multiple regression and cross-classification, literature was reviewed for household and person-category trip generation models developed by multiple regression and cross-classification techniques.

Wooton and Pick (1967) discussed alternative approach called cross-classification to model household trip generation using data from London and West Midlands. The study used three variables namely income (stratified into six categories), household size (stratified into six categories) and cars owned by the household (stratified into three categories). Thus, 108 household classes were formed and trip rates were calculated for each class.

Goulias, Pendyala et al. (1990) developed trip generation and trip chaining model based on a dataset from Detroit Metropolitan Area. Six trip categories namely work, school, shopping, social, personal business and serving passengers were considered in the study. The study, however, did not estimate home-based and non-home-based trip generation by purpose. School trip generation model was the best performing model with R^2 value of 0.545 followed by work trips which yielded R^2 value of 0.313. Likewise, trip generation model for shopping trips yielded R^2 value of 0.079 and social trips yielded 0.087. The R^2 value of the remaining personal business and serve-passenger trip generation models were 0.114 and 0.05 respectively.

(Takyi 1990) used cross-classification technique to investigate household trip rate analysis in Kumasi, Ghana and presented it as a developing country context. Four independent variables were selected for the analysis namely, household income, household size, the number of cars owned per household, and the number of employed persons per household. Household income and size were each classified into six groups, and the number of employed persons and car ownership were classified into three and four groups respectively.

Prevedouros and Schofer (1991) analysed trip characteristics and travel patterns of suburban residents in Chicago based on mid-1989 mail-back survey in a quest to learn causes of and variations in traffic congestion. They employed linear regression to investigate factors affecting all trips, work trips and non-work trips per person. The estimate of the all trips generation model yielded R^2 value of 0.08 which is very low. Similarly, R^2 value yielded by non-work and work trip generation models were 0.28 and 0.33 respectively.

Badoe and Steuart (1997) discussed the urban and travel changes in the Greater Toronto area between 1964 and 1986 and temporal transferability of trip generation model. Multiple regression was used for the study with the household as the unit of analysis. Trips considered in the study were total home-based trips (THB), home-based work trips (HBW), home-based shopping trips (HBS), home-based personal business (HBPB) and home-based social and recreational (HBSR). The THB trip model yielded R^2 value of 0.266 for 1964 and 0.348 for 1986. Similarly, R^2 value yielded by HBW was 0.483 for 1964 and 0.584 for 1986. Further, HBS trip model yielded R^2 value of 0.045 for 1964 and 0.039 for 1986 and HBSR yielded 0.0345 and 0.0308 for 1964 and 1986 respectively. Finally, HBPB trip model yielded R^2 value of 0.022 and 0.025 for 1964 and 1986 respectively.

Hu (2010) calibrated trip generation model for work trips per household using National Travel Survey data, a household survey of travel covering residents of Great Britain. In addition, the author developed trip generation model for shopping trips per person using Edinburgh Household Survey (HS) data. In shopping trip generation model, parking costs reflecting the impact of transport policies was also included along with other attributes. Linear regression analysis and logistic analysis (binary, MNL, and NL models) were employed for both investigations. The linear regression trip generation model of work trips yielded R^2 value of 0.326 and shopping trips yielded 0.135, which is very low.

Chang, Jung et al. (2014) investigated home-based work trips per household in Seoul metropolitan area using six models regression, tobit, poisson, ordered logit, category, and multiple classifications and compared the results. They considered regional characteristics such as demographics, regional economy, transport system also in addition to household characteristics. The data for household characteristics and trip generation rates was taken from a household travel dairy survey of the Seoul metropolitan area and regional characteristics were based on the data from the Korea Regional Development Total Information System (REDIS). The Regression model gave R^2 value of 0.4. The comparison of six models based on RMSE and NRMSE, measures of difference between observed value and model value revealed category-type models to be superior of all. Further, regression model also produced acceptable performance.

Thus, it is apparent from the above literature review that disaggregate trip generation modelling is well researched. It appears that some studies did not segregate trip based on home-based and non-home-based. Comparing the performance of household-based regression models of some important trip categories, it is observed that R^2 value of home-based work trip model varied from 0.4 to 0.584. On the other hand, R^2 value of simply work trip model varied from 0.313 to 0.33. In the same way, home-based shopping trip model in the same place in two different years yielded low R^2 values of 0.039 and 0.045 which are lower than 0.079 yielded by simply shopping trip model. School trip model yielded R^2 value of 0.545. Similarly, home-based social trips yielded 0.0345. In the case of person based models, shopping model yielded R^2 value of 0.135 and work trip model yielded R^2 value of. 0.33. As can be observed, work trip models and school trip model performed far better than the rest. Cross-classification models at the household level of the two models used same three variables, household size, income, and cars. Cross-classification model at a personal level which has been already discussed in the prior section reported that age, employment, and car availability are important variables. For more information, see Appendix A.

2.3 Independent Variables in Trip Generation Modelling

Bruton (1985) opined that trip making is a function of trip maker's socio-economic characteristics, land use and the developments in the study area. Further, socio-economic characteristics of the trip makers which are readily observable are regarded as the significant determinants of travel behaviour (Pas 1986,p.73). Trip generation analysis takes into account the number of independent variables that have an impact on the trip making procedures. Type of model being used determines the level of detail and disaggregation required for these independent variables (Ortúzar and Willumsen 2011,p.490). For instance, variables used for the aggregate (zonal) and disaggregate travel demand models (household and person) could be different. Some commonly used independent variables have been reviewed below:

Family Income

The popular consensus is that the number of trips generated by the households tends to rise with an increase in household income. This is corroborated by the findings of Davies (1969) who studied the role of household income on the frequency of shopping movements. Hanson (1982) also reported that trip frequency is influenced by the household income.

On the other hand, income seems to influence car ownership, as higher income groups have high tendency to own more cars (Kockelman 1997). This is supported by the study of trip generation analysis in Kumasi, Ghana by Takyi (1990). The study revealed that when household income was included in the same model with car ownership, its influence in trip making was significantly reduced. Pas (1974,p.35) suggested that it is imperative to distinguish if the household income itself affects the trip generation rate or it is influencing the trip frequency because of its connection with the car ownership.

Car ownership/License ownership

Car ownership in the household is one of the important factors that influence the trip generation. Since early days, most of the travel demand models have been employing car ownership as an independent variable. However, Kitamura and Kostyniuk (1986) reported that the impact of car ownership on trip frequency is declining. They studied the impact of motorization on the travel behaviour of nuclear households in New York with data from 1963 and 1974. They argued that the decline in the effect of car ownership in New York could be largely due to the homogeneity in car availability across households caused by the affordability of the cars. Moreover, multi-car ownership does not essentially mean the household tend to use it more.

Berechman and Paaswell (1977,p.122) stressed that car ownership alone does not influence the trip frequency. Rather, availability of the car is crucial. For instance, a household may own a car but it is mostly used by the head of the house to commute. Once the car is parked at the work place, other members of the household are deprived of the opportunity to use it. In that case, another member of the household might use a car from car sharing or by other means to carry out their daily activity.

Prevedouros and Schofer (1991) maintained that high ownership of cars does not necessarily explain the variation in frequency of the trips. They conducted research on differences in travel pattern between outer ring, low density and growing suburbs and inner ring, high density and stable suburbs in Chicago. Despite higher car ownership among the growing suburbs, the commute to the central business district was comparatively lesser than the stable suburbs. This was largely due to the substantial portion of growing suburbs people being employed in the suburb they live in.

In the same way, Paez, Scott et al. (2007) studied the elderly trip generation study in Hamilton CMA, Canada. It revealed that license ownership is also positively correlated with the trip frequency.

Household size

The frequency of trips is the result of both individuals as well as household decisions. Hence, as some trips are the result of the collective decision of the household, it is very likely that the trip rates do not increase linearly with increase in the size of the household (Pas 1978). A study in Modesto, California by Shuldiner and Institute (1962) reported that average trip frequency increases with increasing persons per household, at approximately 0.8 trips per day for each additional person. This rate holds true mainly for the non-work trips.

A study in Kumasi, Ghana Takyi (1990), representing a developing country context revealed that together with car ownership and the number of employed persons in the household, household size significantly influenced the trip making. He argued the prevalence of extended family system in Ghana behind this finding.

Age

Prevedouros and Schofer (1991) asserted that the trip frequency and the age of the trip maker are non-linear. The research further explains that the age of a person until 80, is positively correlated with the number of trips with the highest number of trips made by the individual at the age of 40. Evidently, the effect is more for non-work trips. This is further supported by Paez, Scott et al. (2007) in their study in Canada. In addition, they revealed that although elderly people are less mobile, they tend to produce more trips if they have access to vehicles.

In a different context, Pettersson and Schmöcker (2010) studied the travel patterns by those aged 60 or over in Metro Manila, The Philippines a developing country, and compared the results from developed cities, particularly, London. While it was observed that the total trips decreased with increasing age, for specific trip purposes, trends were found similar to developed countries.

Employment status and sex

Boarnet and Sarmiento (1998) acknowledged that sex is a strong predictor of non-work trip frequency as they report that females are likely to make more non-work trips. In an attempt to examine whether the gender gap is narrowing with increasing participation of women on workforce, Hanson and Hanson (1980) weighed the gender impact on trip frequency. The study assessed whether this trend has altered the traditional role of the female in the household amid similar constraints for male and female. However, the result depicted that men still significantly undertake work and recreation trips while female dominated in making shopping trips. Likewise, the presence of children in the family has no impact on unemployed men or women; however, for employed women, the children in the family denies them the time in pleasure travel as they dedicate more time in shopping (Allaman, Tardiff et al. 1982). These instances support the notion that despite increment of shares of women in workforce, the disparities of travel behaviour based on the sex have not altered.

Further, home-based work trips increase with the increase in the number of employed members of the household while home-based other trips are negatively correlated with the higher number of employed members (Paez, Scott et al. 2007).

3 Data Description

The data used for this thesis is from Mobilität in Deutschland 2008 or simply MiD 2008. MiD 2008 a comprehensive account of mobility behaviour of people living in Germany, was conducted jointly by Institute of Social Sciences in Bonn and Institute for Transport Research at German Space center (DLR) in Berlin (infas and DLR 2010). KONTIV which stands for a continuous survey on traffic behaviour started as early as mid-seventies in former West Germany and had survey in 1976, 1982 and 1989. The unified Germany had its first travel survey entitled "Mobilität in Deutschland" in 2002 (MiD 2002), and second one MiD 2008, a replication of 2002 in 2008 (DLR 2012). However, the samples of survey are incomparable as in 1976,1982, 1989 and 2008 surveys included German-speaking residents as basic population, while in 2002, whole residential population including foreigners were included, and the lower limit of age was set at six years in 1989, at ten in 1976 and 1982, but at zero in 2008 (Scheiner, Sicks et al. 2011,p.374).

MiD 2008 gives insight into social demography of individuals and households and daily travel behaviour of residents of Germany (DLR 2012). Randomly selected, about 25,000 households with around 100,000 household members of all age groups were interviewed about their everyday travel behaviour. The sample data was then weighted and expanded to represent the overall travel behaviour of the entire population in Germany (DLR 2012). In addition to travel data, the information on socio-demographic characteristics of the household and individuals were also collected.

The survey was conducted by CATI (Computer Aided Telephone Interview), PAPI (Paper and Pencil Interview) and CAWI (Computer Assisted Web Interview) among which telephone interview was the predominant approach in the survey (DLR 2012). It was held in two consecutive stages (infas and DLR 2010):

- In the first phase, the household survey was carried out that involved collection of information on household size, the profile of individual members, existing transport mode in the family and other features. This information was regarded as the representative of all the household members.

- In the second phase, all household members were individually quizzed about the personal characteristics

3.1 Model Dataset

MiD 2008 provides ample information about the travel behaviour in Germany. Before commencing the analysis task, it was crucial to screen unnecessary data from the original dataset. The thesis intended to investigate trip production modelling from the weekday's data only, therefore, households whose survey was conducted on weekends were omitted at first.

Next step was to segregate all the trips on the basis of trip purpose. To perform this, trips which were recorded as trip chains, needed to be separated initially. "Trip chaining" refers to going from destination to destination without returning home from each trip and, thereby executing series of short trips (Ross, Meyer et al. 1995,p.340-341). Individuals whose trips were recorded without origin of the first trip could not be separated from the trip chains. Likewise, individuals whose trip purpose and trip destination were missing were also considered surplus to requirements as such trips could not be put into any trip categories. Since trip production models were going to be analysed at both household and personal level, exclusion of individual only could result in biases of the results. Therefore, all the households of the individuals which had incomplete information were discarded. Out of 25,922 households with 60,713 members and 193,290 trips, only 14,367 households with 31,479 members and 117,829 trips turned out to be final sample for the study.

Primarily, on the basis of the purpose, MiD 2008 has divided the trips into work, business, training, shopping, errands, leisure and accompanying or escorting someone. However, at analysis stage, some of these categories were aggregated. Moreover, the thesis dealt separately with home-based and non-home-based trips. Home-based trips are coined for the trips where the home of the trip maker is either the origin or the destination of the journey and, non-home-based trips are coined for the trips where neither end of the journey are the home of the trip maker (Ortúzar and Willumsen 2011,p.140). While return home trips have been ignored, remaining trips have been broken down into following categories:

i. Home-based work (HBW) trips: Trips from home to workplace as well as personal business

ii. Home-based education (HBE) trips: Trips from home to all kind of learning places.

iii. Home-based shopping (HBS)trips: It includes trips from home to shopping place

iv. Home-based other (HBO) trips: It includes all trips from home to any destination except work and business, education and shopping e.g. recreation, social

v. Non-home-based trips (NHB): Trips that have neither ends at home, regardless of purpose.

Home-based work and business trips were aggregated to form home-based work trips. Home-based education and shopping trips were taken as it is. Remaining home-based errands, leisure and accompanying trips were combined to form home-based other trips. Finally, all the trips with neither end of the journey at home were grouped under non-home-based trips.

3.2 Trip Distribution by Purpose

Tab. 3.1 presents the distribution of weighted trips by different trip production categories. The table shows that home-based trips with more than 70% comprise the maximum amount of trips. Overall, females were observed to produce more trips compared to males as they made 52.35% and males made 47.65% of the total trips. While it appears males dominated HBW trips and HBE trips, females were predominant producers of HBS, HBO and NHB trips. Further, it can be observed that HBE trips comprising 6.71% of total trip production were least produced trips followed by HBS trips with 13.35% and HBW with 15.6%.

Tab. 3.1 Trip production distribution

Trip Production Type	All (%)	Trips by male (%)	Trips by female (%)
Home-based work trips	15.6	55.40	44.60
Home-based education trips	6.71	52.40	47.60
Home-based shopping trips	13.35	44.90	55.10
Home-based other trips	35.91	45.70	54.30
Non-home-based trips	28.4	46.30	53.70
Total trips	**100**	**47.65**	**52.35**

3.3 Household characteristics

Household characteristics mainly include household size, household income, the number of cars and license owned by the household, the number of children, the number of student etc.

Household Size

It is apparent that trip frequency of the household increase with the increase in household size. Tab. 3.2 gives the overview of household structure for the selected dataset (14,367). In the dataset, only one household did not disclose the household size. The table shows that the proportion of single person households with 39% constitute the largest share of households. Similarly, two-person household comprises 34% of total households and less than 30 % of the households have three or more members of the household.

Tab. 3.2 Household structure

Household Size	Percentage
1	38.5
2	33.9
3	14.1
4	10.1
5+	3.4

Household Income

On the basis of monthly income of the household, MiD 2008 has categorised household income into 15 groups ranging from less than € 500 per month income to greater than € 7000 per month. Fig. 3.1 below presents the distribution of household income per month for the selected dataset (14,367). The household income data was not available for 2,060 households in the data set. About 30 % of the households have income below € 1,500 euros per month. In the same way, more than 40% of the household's income is from € 1,500-3,000 per month and upper income ranging from € 3,000-€ 5,000 per month accounts to 17% of the households. Remaining 5% households earn more than € 5,000 euros per month. Household income influence positively in the frequency of trip making. In the thesis, the household income below € 2600 has been termed low income and above €2,600 has been termed as medium/high income.

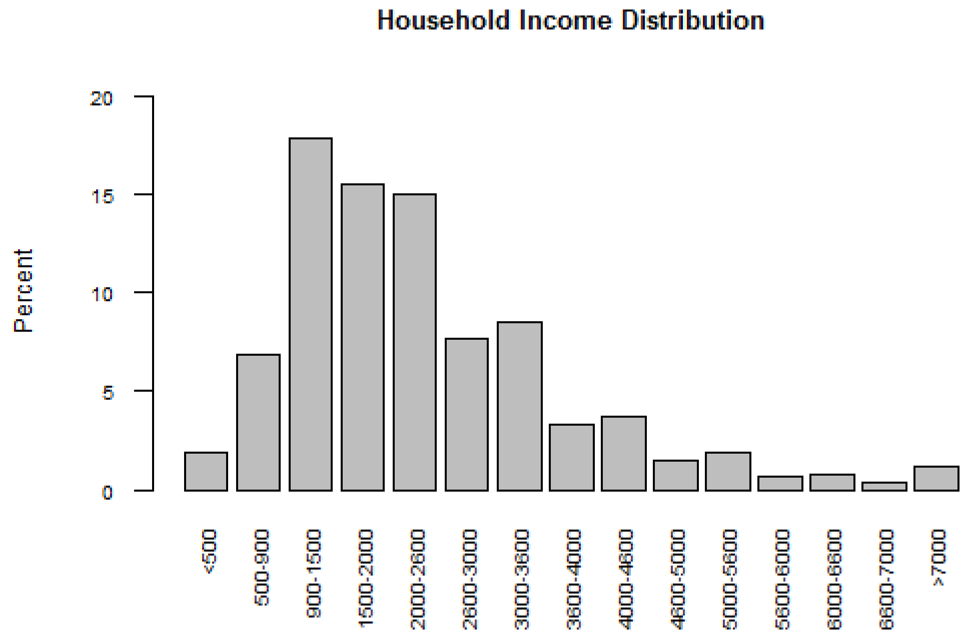


Fig. 3.1 Household income distribution

Car and License Ownership

Tab. 3.3 presents the number of cars and driving licenses owned by the households in the selected dataset (14,367). It depicts that 17.5% of the households do not possess cars, 53.4% households possess at least one car and about 30% of the households own multiple cars. On the other hand, only 8.5% of the households were reported to have zero licenses, 40.5% of households own at least one license, while households owning multiple driving licenses exceeds 50%. The chances of households making trips increase significantly with increase in the number of cars and license owned by the household members. In the research, both attributes have been included separately.

Tab. 3.3 Number of cars and license per household

Number of cars/License	Cars (%)	License (%)
0	17.5	8.5
1	53.4	40.5
2	24.1	39.7
3+	5.0	11.3

Further, the superposition of household income and ownership of cars is demonstrated in Fig. 3.2 below. It can be observed that propensity of owning one or more vehicles increases with the increase of household income. As presented, more than 60% of the households below € 500- income have no cars and more than 65% of the households with income more than € 4000 per month possess multiple cars.

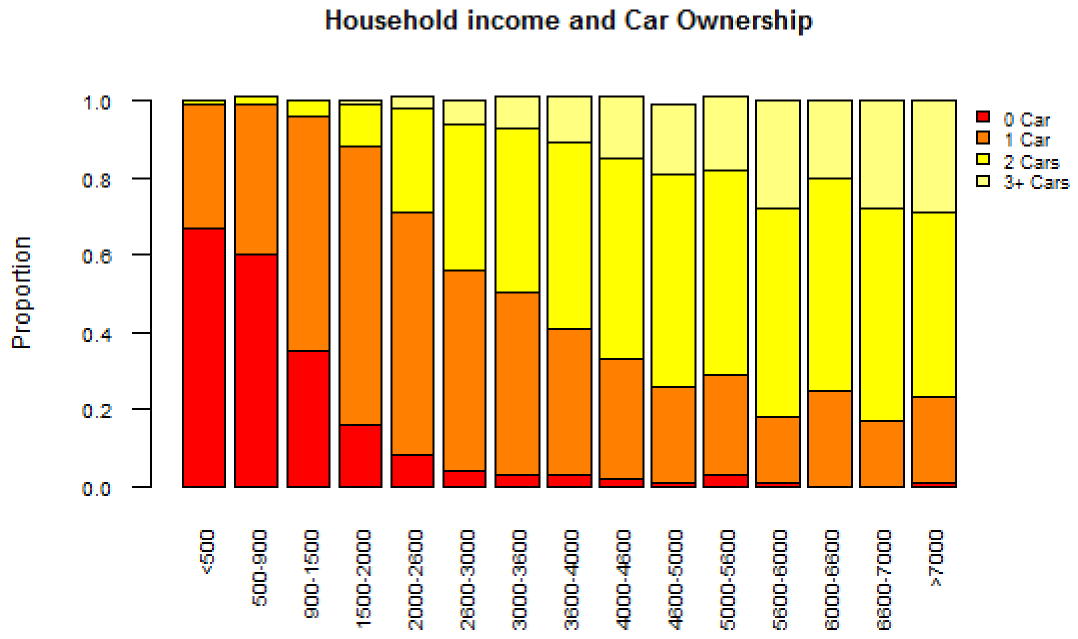


Fig. 3.2 Household income and car ownership

Number of Workers in the Household

Tab. 3.4 demonstrates the number of employed people in the household for the selected dataset (14,367). 40.2% of the household have no employed people, 32.4% of the households have one worker, either full-time or part-time or trainee and 27.4% households have multiple workers. Similarly, 48.4% of the households have no full-time workers and more than 50% of the households have at least one full-time worker. On the other hand, only about 22 % of the households have at least one part-time worker. In addition, only about 5% of the households have at least one trainee and the vast majority of the households do not have a single trainee. To find the impact of full-time workers, part-time workers and trainees in the work trip, all three have been included separately.

Tab. 3.4 Number of workers per household

Number of Workers	Full-time Workers (%)	Part-time Workers (%)	Trainees (%)	All (%)
0	48.4	77.0	95.4	40.2
1	39.8	22.1	4.2	32.4
2+	11.9	0.1	0.4	27.4

Number of Children in the Household

Tab. 3.5 below presents the distribution of children aged below six and children from 6 to 17 in the selected dataset. It depicts that about 85% of the households do not have children age 6 to 17 while about 95% of the household do not have children age below six. Likewise, households with at least one children age below six and age 6 to 17 are 3.6% and 8.8% respectively. Further, households having multiple children below six is about 1% while children 6 to 17 is about 7%. Children below six who are mostly expected to be preschool and children above 6 to 17 are expected to be students. Since travel behaviour among these two groups is not similar, they have been used separately in the modelling process.

Tab. 3.5 Number of children in the household

Number of Children	Children Age <6	Children Age 6-17
0	95.4	84.7
1	3.6	8.8
2	0.9	5.3
3+	0.1	1.3

Number of Students in the Household

Tab. 3.6 presents the distribution of the number of school pupil and university student in the selected household dataset (14,367). It demonstrates that more than 80% of the households do not have any student. 9.1% of the households have one school pupil and 7.1% of the households have multiple students. Further, less than 5% of the households have university student. Both attributes have been used in the analysis of education trips in the later chapter.

Tab. 3.6 Number of students in the household

Number of Students	School Pupil (%)	University Student (%)
0	83.8	95.2
1	9.1	4.3
2+	7.1	0.5

3.4 Individual Characteristics

Individual characteristics or attributes involves sex, age, license ownership status, employment status, student status etc.

Sex and Age

Fig. 3.3 below presents the frequency distribution of males and females in the dataset by different age groups. In the selected dataset (31,479), the age of three individuals and the sex of 57 individuals are missing. The percentage of male and female trip maker stand at 49.21 and 50.79 respectively which is almost similar. Further, as expected individuals above age group, 25 are a significant contributor of trips for both male and female. Both sexes, as well as six age groups, have been included in the models.

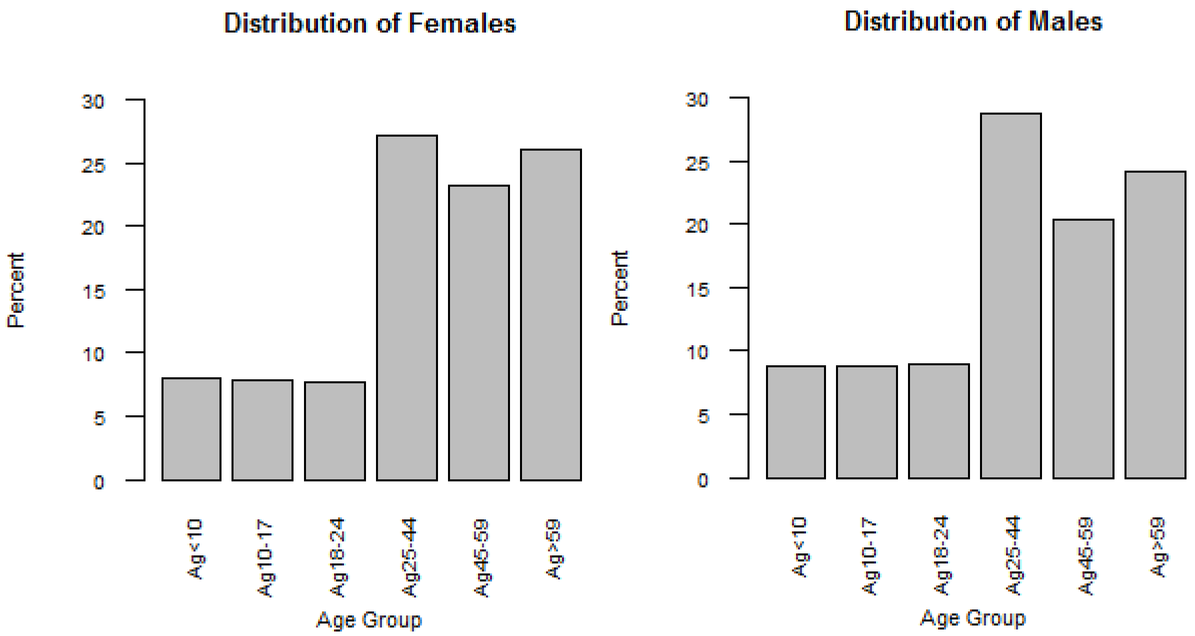


Fig. 3.3 Distribution of male and female by age

License Ownership

The acquisition of license plays a crucial role in mobility behaviour. In the selected dataset, among 31,479, individuals owning and not owning license constitute 23,433 and 8,037 respectively. Remaining 174 individuals did not mention whether they own license or not.

Employment Status

Employment status of the individuals is an important predictor of trip production analysis at a personal level as well. In the selected dataset (31,479), employment status of 174 individuals was missing. 14,229 individuals were reported employed and 17,076 individuals were reported unemployed. In addition, among employed individuals, full-time workers, part-time workers, and trainee comprise 9,561, 3,945 and 845 respectively. Remaining individuals though employed did not reveal the scope of employment. Similar to the household-category trip production analysis, full-time workers, part-time workers, and trainees have been included in the person-category trip production modelling.

Student Status

MiD 2008 has grouped the student into school pupil and university student. The dataset reveals that 4,697 individuals are school pupil and 751 individuals are a university student. As the thesis deals with education trips as well, and both school pupil and university student could have a different impact on education trips, both attributes have been considered separately.

4 Modelling Methodology

4.1 Trip Calculation

The thesis compared the trip production model results at household and personal level developed through multiple linear regression and cross-classification approach. Before applying the modelling techniques, trips from the model dataset were assigned to each trip categories. Trip data consisting of origin and destination of the trip and trip purposes were used from the model dataset.

Trip production for each trip categories was calculated based on the known activities and the sequence of the activity participation. For example, within a day, a traveller is involved in three different activities. At first, he leaves home for work which is his first trip and a home-based work trip (HBW). In the afternoon, he produces his second trip, a non-home-based trip (NHB), when he goes to a restaurant for lunch and again come back to the workplace. Later in the evening before leaving for home, he goes to the grocery store for shopping which is again another non-home-based trip and his 3rd trip of the day.

During trip production calculation, if a person leaves for an activity and returns back to the origin of the tip after completing the activity then the trip is counted as a single trip. For instance, in the Fig. 4.1 below, when the person goes for lunch from work and returns back again to his workplace after having lunch, the trip is counted as a single non-home-based trip. Additionally, the trip back home in the evening is not counted. The figure below gives the pictorial overview of the calculation of trip production.

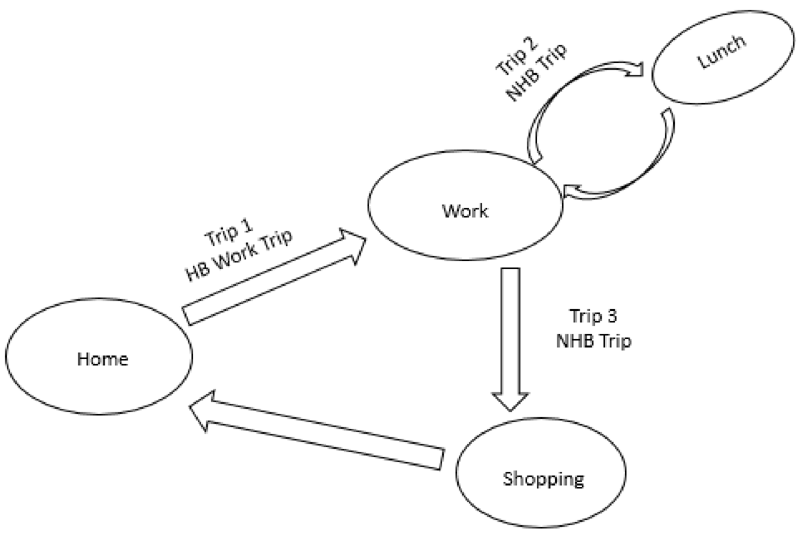


Fig. 4.1 Calculation of trip production from daily activity pattern

To clarify the idea, few examples of the trip production calculation are illustrated below:

1. Home → Work + Work → Home = 1 HBW Trip
2. Work → Shop + Shop → Home = 1 NHB Trip
3. Home → Work + Work → Shop + Shop → Home = 1 HBW +1 NHB Trip

4. Home → Education + Education → Other + Other → Shop + Shop → Home = 1 HBE Trip + 2 NHB Trip

4.2 Multiple Regression

In multiple regression of the trip generation model, the dependent variable is the number of trips produced by the household or person, while independent variables are predominantly household and personal characteristics. The independent variables differ for personal and household level analysis. The household level trip generation analysis included household characteristics such as, income, household size, the number of children, the number of licenses owned by the household members and the number of employed members.

On the other hand, for person-category trip production analysis, the explanatory variables have been dominated by the personal characteristics such as age, sex, ownership of the license, student and employment status etc. In addition to this, relevancy of household characteristics such as the size of the household, income, and number of cars owned by the household members have also been examined.

Models were developed for each trip classification using the multiple regression approaches. In model building, identifying the best subset of many variables is arguably the most difficult task. Rawlings, Pantula et al. (2001) suggested a number of statistical procedures to help in the selection of the best regression equation which is as follows:

i. All possible regressions
ii. Stepwise regression
iii. Forward selection
iv. Backward elimination

In real life, the independent variables are not orthogonal (statistically independent), which causes the least square of each independent variable to be dependent on other variables in the model. Hence, in that case, all possible subset regression ensures the best model for each subset size.

In the same way, forward selection method involves selection of best possible models by adding one variable at a time to the previously chosen subset. Contrary to forward selection, backward elimination begins with the full model and step by step eliminates one variable whose omission will result in the least increment of the residual sum of squares. Finally, stepwise regression generates the subset models by adding or deleting the variable based on the greatest impact on residual sum of squares. Compared to forward selection and backward elimination, stepwise regression needs more computation but offers an advantage in checking the number of possible subsets before deciding the model. For detail information see (Rawlings, Pantula et al. 2001).

In this research, stepwise regression is adopted as it incorporates both forward selection and backward elimination approaches. The author is well aware that the stepwise regression doesn't guarantee the best subset size and it is not the goal of the research to dig deep into the limitations of it. Hence, researcher judgment would be also key to the selection of variables. As Rawlings, Pantula et al. (2001) rightly state "No variable selection procedure can substitute for the insight of the researcher".

Multicollinearity Problem in Regression

Multicollinearity refers to the presence of linear or near linear relationships between the independent variables (Silvey 1969,p.539). In other words, multicollinearity means two or more independent variables explaining the same effect in regression. Multicollinearity is a threat, as it results in the estimation of model parameter that is sensitive to changes in model specification and to sample coverage (Farrar and Glauber 1967,p.94). If all the multicollinear variables are employed in regression analysis at the same time, it is very likely that the individual effect of the independent variables will be lost.

Despite its severity, “no proper treatment” is available to tackle the problem of multicollinearity (Farrar and Glauber 1967,p.92). Hence, personal judgement could be crucial in this situation. The model estimates were closely examined to avoid problems caused by multicollinearity of variables.

Thus, model development is summarised in following steps:

1. Variables expected to be important in explaining the trip frequency were selected. For this, previous studies, as well as different attributes in the dataset were tested.
2. For continuous variables, the correlation among the variables was tested to determine the expected effect of the independent variables on dependent variables as well as the correlation between the independent variables. Multicollinear variables were noted.
3. All variables were entered in stepwise regression. Variables failing to comply with 5% significance level were discarded. In addition, regression coefficients were examined and discarded one at a time if wrong signs were spotted.
4. Finally, the variables included in the model were examined whether they made sense.

4.3 Cross-classification

To make the cross-classification simpler, two-way cross-classification was performed. Cross-classification tables of trip rates were developed for each trip purposes. Such tables demonstrate how dependent are mean household or person trip rates on independent variables used in the cross-classification. As independent variables are causal variables that influence trip rates, the independent variables used in the model for different purposes were not same. Since the aim of the thesis is to compare the performance of the trip production models at household and personal level, the emphasis was given to categorise households that ideally distinguish trip making behaviour rather than predefined household types. Further, importance was given on selecting the similar variables which were found significant in multiple regression.

The ideal case in cross-classification model in trip generation analysis would be when there is enough data for each cell of the matrix (Badoe and Mwakalonge 2011,p.264). As state of practice stress that every household type should have at least 30 household records in the survey, in order to ensure sufficient household records in each cell, household types need to be aggregated (Moeckel, Huntsinger et al. 2017,p.34). For instance, a trip production model is cross-classified by four categories of cars and five household sizes. With these combinations, 20 household types would be formed. However, it is very likely that the households with just one member possess more than three cars are very few. In that case, aggregation would compensate the paucity of the household records. Fig. 4.2, elaborate the aggregation phenomenon. See (Moeckel, Huntsinger et al. 2017) for a detailed description. The thesis has adopted this approach of aggregation in both person-category as well as household-category cross-classification.

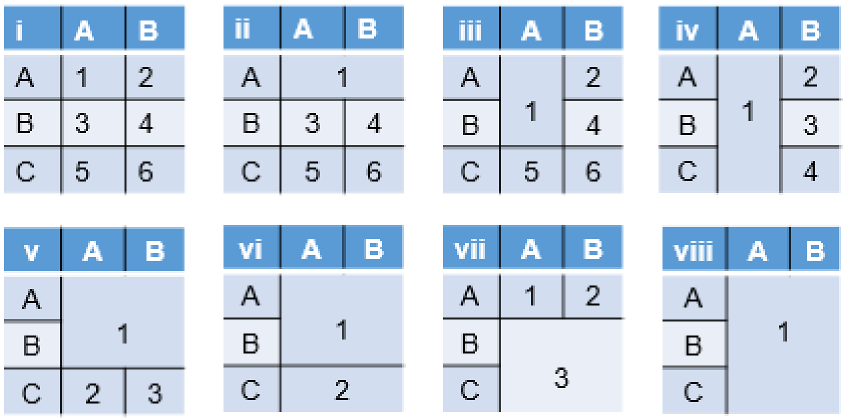


Fig. 4.2 Selected aggregations of six household types with two attributes (Source: (Moeckel, Huntsinger et al. 2017,p.35))

The model development is summarised briefly in following steps:

1. All attributes expected to be significant contributors of trip frequency were determined. For household-category cross-classification, in addition to commonly used variables such as cars, household size, and household income, variables such as, unemployed and children were also tested. The variable combination that performed well in terms of R^2, RMSE, and NRMSE was chosen for the model.
2. During cross-classification, as described above, if any household type or person type were short of 30 records, they were aggregated with another nearest household type or person type.
3. The results were presented graphically to observe the heterogeneity and impact of variables used in the cross-classification.

4.4 Survey Weights

Survey weights are required for survey sample to use it in model development and statistical analysis so that the weighted sample represent the whole population. Suitable survey weights are available in the model dataset for a person, household, and trips. Thus, during analysis, personal survey weights were used for the person-category trip production models and household survey weights were used for the household-category trip production models.

4.5 Model Validation and Comparison

In this research, the coefficient of determination R^2, root mean square error (RMSE) and normalised root mean square error (NRMSE) have been employed for the comparison of person-category and household-category trip production models. RMSE and NRMSE are two measures of difference between observed and predicted trip rates. NRMSE is an alternative measure of RMSE normalised by the standard deviation. Mathematically, they are given by:

Equation 4.1

$$RMSE: \sqrt{\sum_{i}^{N} \frac{(Observed_i - Predicted_i)^2}{N}}$$

Equation 4.2

$$NRMSE: \sum_{i}^{N} \frac{RMSE}{sd(Observed_i)}$$

Where,
$Observed_i$ = observed trip rate,
$Predicted_i$ = predicted trip rate,
$sd(Observed_i)$ = standard deviation of the observed trip rate

For the better performance of the model, higher R^2 value and lower RMSE and NRMSE are desired. All three measures were adopted for both multiple regression and cross-classification cases.

4.6 Data Analysis Software

Both multiple regression and cross-classification were performed using R, a free software environment for statistical computing and graphics. R offers extensive statistical (like linear and non-linear modelling, classical statistical tests, time-series analysis, classification etc.) and graphical techniques (The R Foundation 2017).

Besides standard sets of packages in R, packages can be downloaded and installed in R depending on the types of analysis. Regression was carried out using R function *lm*. In the thesis, some additional packages were installed to aid the modelling task, for instances, "*corrplot*" for visualising correlation among variables, "*hydrGOF*" for estimating RMSE and NRMSE, "*weights*" for incorporating survey weight.

5 Trip Production Modelling by Multiple Regression

This section delves deeper into the trip production modelling of all trip categories both at household and personal level by regression approach. During regression, dependent variables were related to the number of independent variables and relationship between them were produced. After completing the analysis, estimation results were compared with previous disaggregate trip generation studies.

5.1 Household-Category

Before commencing the analysis, dependent and independent variables were defined. Dependent variables were the trips produced by households for each of the five trip categories. They are listed below with their names used in the modelling:

i. hwtrip: Home-based work trips per household
ii. hetrip: Home-based education trips per household
iii. hstrip: Home-based shopping trips per household
iv. hotrip: Home-based other trips per household
v. nhbtrip: Non-home-based trips per household

On the other hand, independent variables used in this study were socio-economic and demographic variables at the level of the households which were obtained from the model dataset. The selection was based on previous studies as well as testing of attributes available in the model dataset. These independent variables with their range, mean and standard deviation are presented in Tab. 5.1.

Tab. 5.1 Independent variables used in household-category trip production models

Independent Variables	Description	Range	Mean	SD
HSIZE	A continuous variable that describes the number of household members in the household	1-11	2.500	1.400
Log(HSIZE)	A continuous variable that describes the logarithmic of number of household members in the household	-	-	-
INKhigh	A dummy variable that takes the value of 1 if the income of the household is above €2000, 0 otherwise	-	-	-
CARS	A continuous variable that describes the number of cars owned by the household members	0-8	1.356	0.817

LICN	A continuous variable that describes the number of license owned by the household members	0-8	1.844	0.872
CHILD5	A continuous variable that describes the number of household members aged below 6	0-3	0.080	0.331
CHILD6_17	A continuous variable that describes the number of household members aged from 6 to 17	0-6	0.343	0.725
SCHstud	A continuous variable that describes the number of school pupil in the household	0-8	0.365	0.747
UNIstud	A continuous variable that describes the number of university student in the household	0-8	0.067	0.28
FTIME	A continuous variable that describes the number of full-time workers in the household	0-8	0.750	0.76
PTIME	A continuous variable that describes the number of part-time workers in the household	0-8	0.300	0.487
TRAINEE	A continuous variable that describes the number of trainee in the household	0-8	0.069	0.279
UNEMP	A continuous variable that describes the number of unemployed in the household	0-8	1.372	1.010

Household size was tested as continuous variables in the study. Further, the logarithm of household size was also used as another continuous variable. Stepwise regression was performed in two steps. In first step, household size was entered with other independent variables, and in second step, logarithm of household size was entered with other independent variables. Later, model that performed better was chosen. In all trip categories, household size is expected to influence positively in trip frequency. Since income was recorded in the dataset as a range rather than a definite value, it was decided to include income variable as a dummy variable. Income was categorised into low-income household (below € 2,000 per month) and medium to higher income (above 2,000 per month). Low-income group has been considered as the reference group. Income is also expected to have a positive impact on all trip categories.

It is very likely that the household may not have cars but the members in the household could have a license. Hence, both variables were considered in the analysis and were tested as continuous variables. Like, previous two variables, it is predicted that the higher number of cars and license in the households increase the chances of making all kind of trips.

The number of children is an important factor in determining the number of trips made by the household. It is obvious that the younger children and elder children do not show similar travel behaviour. Therefore, children below six and children above five were considered separately, both as continuous variables. The number of children particularly, children below six was expected to hinder work trips. However, their presence is likely to impact positively on education trips.

Students in the household increase the number of education trips. School going student and university student may not exhibit the same kind of travel behaviour. Thus, both were entered separately into the regression as continuous variables.

Similarly, the number of employed members in the household promotes higher work trips. On the other hand, being employed should restrict the participation in non-work trips. In the dataset, the employment status was branched into full-time workers, part-time workers, trainees, employed but the scope of work unspecified and unemployed. Full-time workers, part-time workers, trainees and unemployed members were considered separately while the number of employed people whose scope were not mentioned was considered as the reference group. Since trainees are those who are undergoing training for a particular job or profession, it is expected to impact on education trips as well in addition to work trips.

All above mentioned independent variables may not be truly independent to each other. Thus, it is imperative to learn their interdependencies, so that highly correlated independent variables could be screened during analysis. The correlation among these variables is presented in Fig. 5.1. Through the close examination of the figure, it can be observed that the household size is highly correlated with the number of licenses owned by the household, the number of school going members, the number of employed members and the number children 6 through 18.

Likewise, the number of licenses owned by the households have a higher correlation with the number of cars and full-time workers in the household. Unsurprisingly, the number of school going members and the number of children 6 to 18 are highly correlated. Finally, the number of unemployed members in the household appear to have a high correlation with school going members and the number of children of age 6 to 18. These highly correlated independent variables were closely examined to avoid error in the models.

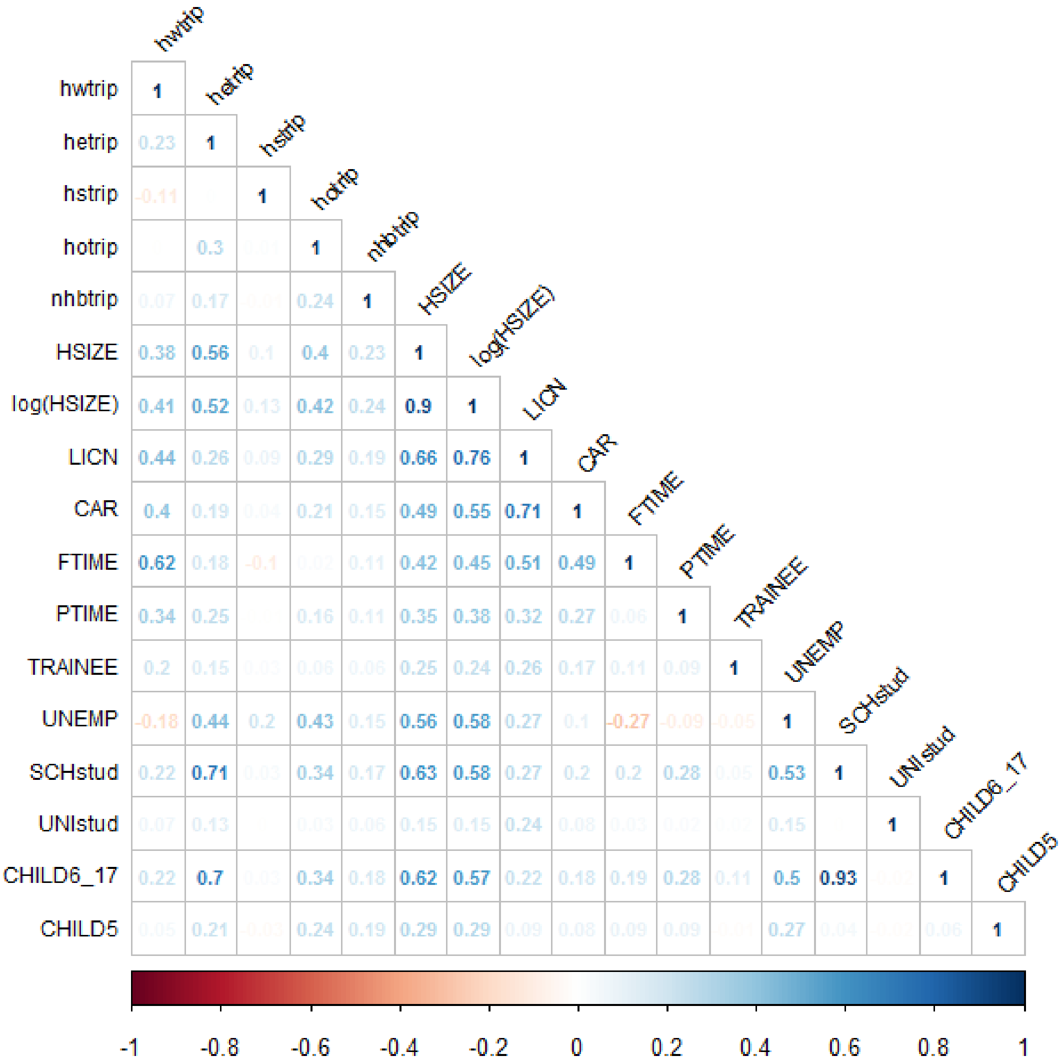


Fig. 5.1 Correlation among variables

5.1.1 Home-Based Work Trips

Initially, all variables were entered in regression equation in a stepwise procedure. It was conducted in two steps (once with household size and once with logarithm of household size). The variables given by the stepwise regression were the number of full-time workers (FTIME), the number of part-time workers (PTIME), the number of trainees (Trainee), the number of cars (Cars), the number of school pupil (SCHpupil) and the number of a university student (UNIstud). The variables were significant at 1% to 0.1% level. The model yielded R^2 value of 0.484. However, the latter two variables (SCHpupil and UNIstud) though significant, their effect could not be described. On omitting these two variables, the R^2 value was 0.483 which is almost equal to the previous value. Hence, the final model comprised FTIME, PTIME, TRAINEE, and CARS which is presented in Tab. 5.2. Neither of the two household size variables was found significant in this case.

The R^2 value 0.483 yielded by the model denotes 48.3% of the trip produced is influenced by the variation in the variables included in the model while remaining 51.7% is explained by other factors. RMSE and NRMSE of the final model were found to be 0.663 and 74.1 respectively.

It is no surprise that all employment variables were found to be significant in HBW trip modelling. Coefficients of the variables suggest that full-time workers are the most significant contributor of HBW trips followed by part-time workers and trainees. Finally, the number of cars though with positive coefficients has a least positive impact on HBW trips. What looks startling is that income variable failed to make any impact.

Tab. 5.2 Estimation results of household-category HBW trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.034	3.822	***
Number of full-time workers (FTIME)	0.655	83.229	***
Number of part-time workers (PTIME)	0.547	47.649	***
Number of trainee (TRAINEE)	0.371	17.586	***
Number of cars (CARS)	0.023	3.132	**

*** 0.1% significance level, **1% significance level, *5% significance level

Number of observations: 14,633, Adjusted R-Squared: 0.483, RMSE:0.663, NRMSE:74.1

5.1.2 Home-Based Education Trips

On entering all variables in stepwise regression, subset of variables obtained were full-time workers (FTIME), part-time workers (PTIME), trainee (TRAINEE), unemployed (UNEMP), school pupil (SCHpupil), university student (UNIstud), number of children 6 to 17 (CHILD6_17) and children below 6 (CHILD5). The result, however, was not statistically significant with these variables. Since school pupil and children age 6 to 17 is highly correlated, children age 6 to 17 was omitted at first. Then, full-time workers and part-time workers were omitted one after another. Finally, as shown in Tab. 5.3, SCHstud, UNIstud, TRAINEE and CHILD5 were found to be significant. The model yielded R^2 value of 0.579 which indicates that about 58% of the trip produced is influenced by the variation of variables in the model. All the variables have correct sign. The RMSE and NRMSE yielded by the model are 0.51 and 87.8 respectively.

Expectedly, school students appear to contribute more than university students in HBE trip making. Surprisingly, trainees which were found to influence HBW trips also appears to have a significant impact on HBE trips as well. It could be because learning is also involved in a traineeship. Despite only 10 children below 6 claiming to be a student in the dataset out of their 1004 in the dataset, it was observed that the number of children below 6 contributed significantly in HBE trips. The reason for this could be they are expected to be admitted in nursery and kindergarten which is also a learning place.

Tab. 5.3 Estimation results of household-category HBE trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.010	2.683	**
Number of school pupil (SCHstud)	0.711	131.151	***
Number of university student (UNIstud)	0.324	24.118	***
Number of trainee (TRAINEE)	0.321	22.642	***
Number of children aged below 6 (CHILD5)	0.408	33.119	***

*** 0.1% significance level, **1% significance level, *5% significance level
Observations: 14,637, Adjusted R-Squared: 0.5785, RMSE:0.51, NRMSE:87.8

5.1.3 Home-Based Shopping Trips

The subset of variables given by stepwise regression of HBO trips were income (INKhigh), full-time workers (FTIME), part-time workers (PTIME), trainees (TRAINEE), unemployed (UNEMP), school pupil (SCHstud), university student (UNIstud), children 6 to 17 (CHILD6_17) and children below 6 (CHILD5). However, the model run with these variables did not deliver the statistically significant result. CHILD6_17, SCHstud and UNIstud delivered opposite sign than expected. Step by step, insignificant variables were omitted from the model. Thus, final model comprised INKhigh, FTIME, UNEMP and CHILD5 which is demonstrated in Tab. 5.4**.**

The R^2 value of the final model was 0.0524 which shows that the variables in the model is able to describe merely 5.24% of the total variation in the model. Further, the model yielded RMSE value of 0.854 and NRMSE value of 98. Although the overall fit of the model was poor, all the variables involved, exhibited correct signs. The model indicates that the HBS trips increase with the increase in household income and the number of unemployed members in the household. On the other hand, the negative coefficient of full-time workers implies that full-time workers are less likely to make shopping trips which are obvious as their involvement in non-work trips is justified. Another important indication is that children below six have negative impact on HBS trips which looks logical as younger children always require somebody to look after them.

Tab. 5.4 Estimation results of household-category HBS trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.409	28.36	***
Medium/high household income (INKhigh)	0.111	6.754	***
Number of full-time workers (FTIME)	-0.076	-6.340	***
Number of unemployed members (UNEMP)	0.172	20.222	***
Number of children aged below 6 (CHILD5)	-0.246	-9.131	***

*** 0.1% significance level, **1% significance level, *5% significance level
Observations: 12,577, Adjusted R-Squared: 0.0524, RMSE:0.854, NRMSE:98

5.1.4 Home-Based Other Trips

Subset of variables given by the stepwise regression were income (INKhigh), full-time workers (FTIME), part-time workers (PTIME), trainees (TRAINEE), unemployed (UNEMP), cars (CARS), driving license (LICN), school pupil (SCHstud), university student (UNIstud), children

6 to 17 (CHILD6_17) and children below six (CHILD5). The model run with these variables did not give the desired result as FTIME, UNIstud and SCHstud produced opposite sign than expected. Due to the high correlation of SCHstud with CHILD6_18, SCHstud was omitted at first. Similarly, UNIstud, FTIME were removed in succession. Finally, the statistically significant model comprised PTIME, TRAINEE, UNEMP, LICN, CARS, CHILD6_17 and CHILD5. The model result is presented in Tab. 5.5. Although, LICN and CARS are highly correlated, they both were statistically significant in the model and thus both were retained. The R^2 value of the final model was 0.2586 which means the variables in the model describes 26% of the total variation in the model. Likewise, the model yielded RMSE value of 1.544 and NRMSE value of 87.8.

HBO trips include errands, leisure and escorting someone. The model coefficient shows that the household with a higher number of children makes more HBO trips. Part-time workers and unemployed could be parents and they are likely to get involved in escorting children to leisure and other activities. In the same way, the unemployed member could be elderly people as well, who have a higher tendency of making leisure trips. Further, the households with higher income and a higher number of licenses and cars are also likely to make more HBO trips. Trainees also contribute positively in making HBO trips.

Tab. 5.5 Estimation results of household-category HBO trip production model

Variables	**Coefficient**	**t-statistic**	**Significance**
Intercept	0.407	12.203	***
Medium/high household income (INKhigh)	0.168	4.224	***
Number of part-time workers (PTIME)	0.328	10.183	***
Number of trainees (TRAINEE)	0.181	3.276	**
Number of unemployed members (UNEMP)	0.536	29.640	***
Number of license owned (LICN)	0.140	5.948	***
Number of cars (CARS)	0.094	4.192	***
Number of children aged 6 to 17 (CHILD6_17)	0.297	11.177	***
Number of children aged below 6 (CHILD5)	0.730	15.329	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 12,577, Adjusted R-Squared: 0.2586, RMSE:1.544, NRMSE:87.8

5.1.5 Non-Home-Based Trips

On entering all the variables in stepwise regression, subsets of variables were logarithm of household size (log(HSIZE)), income (INKhigh), full-time workers (FTIME), part-time workers (PTIME), unemployed (UNEMP), driving license (LICN), school pupil (SCHstud), university student (UNIstud), children 6 through 17 (CHILD6_17) and children below six (CHILD5). However, on running a regression with these sets of variables, FTIME gave negative coefficient. Thus, it was removed and regression was again run. One by one, insignificant variables were removed. Finally, INKhigh, log(HSIZE), PTIME, UNEMP, LICN, and CHILD5 were found to be significant with the R^2 value of 0.0834. Although correlation of log(HSIZE) with LICN is high, both were statistically significant. Thus they both were included. The final results of the model are illustrated in Tab. 5.6. The logarithm of household size worked here unlike other four models.

R^2 value of 0.0834 shows that the model is poorly fit. The variables in the model describe only 8.2% of the total variation in the model. However, all the variables were significant and with correct coefficient signs. The households with children below six have a higher propensity of making NHB trips. As can also be seen, part-time workers and unemployed members are also positive contributors of NHB trips. It can be speculated that at least one parent in the households with children below six is either unemployed or part-time worker. Thus, it is likely that children below six get picked from the preschool or kindergarten and are escorted by their parents to shopping, recreation, and other activity places before returning to home. Likewise, positive coefficient of the logarithm of household size indicates NHB trips increase as household size increases. Other important variables with positive coefficients are household income and license owned by the household. The model yielded RMSE and NRMSE values 1.917 and 95.9 respectively.

Tab. 5.6 Estimation result of household-category NHB trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.487	12.470	***
Medium/high household income (INKhigh)	0.368	9.522	***
Logarithm of household size (log(HSIZE))	0.218	3.265	**
Number of part-time workers (PTIME)	0.178	4.303	***
Number of unemployed (UNEMP)	0.106	4.383	***
Number of license (LICN)	0.097	3.299	***
Number of children aged below 6 (CHILD5)	0.929	15.163	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 12,577, Adjusted R-Squared: 0.0834, RMSE:1.917, NRMSE:95.9

5.1.6 Summary

A total of five models were developed using multiple regression for all five purposes considered. Besides, HBW and HBE no trip production model could explain the variation of more than 45%. HBO trip production model accounted for about 26% of the total variability in HBO trips while remaining two HBS and NHB trip production model accounted less than 10% of the variation in corresponding trips.

Except, HBW trips and HBE trips, unemployed members appear to influence all other trips. The non-work trips by unemployed are justified. Likewise, children below six were found to occur in all models except HBW trip production model. Other variables that were repeated most were the number of licenses and household income.

5.2 Person-Category

For person-category trip production model, dependent variables were personal trip rates for each of the trip categories and independent variables were personal attributes as well as few household attributes of the individuals. The dependent variables with their names in the modelling are listed below:

i. pwtrip: Home-based work trips per person

ii. petrip: Home-based education trips per person

iii. pstrip: Home-based shopping trips per person

iv. potrip: Home-based other trips per person

v. pnhtrip: Non-home-based trips per person

Independent variables are depicted in Tab. 5.7. Besides personal attributes, household attributes such as household size, income, and cars were also considered in the model.

Tab. 5.7 Independent variables used in person-category trip production models

Independent Variables	Description
LIC	A dummy variable that takes the value of 1, if individual possess driving license, 0 otherwise
SEX_FEMALE	A dummy variable that takes the value of 1 if the individual is female, 0 otherwise
FTIME	A dummy variable that takes the value of 1 if the individual is full-time worker, 0 otherwise
PTIME	A dummy variable that takes the value of 1 if the individual is part-time worker, 0 otherwise
TRAINEE	A dummy variable that takes the value of 1 if the individual is trainee, 0 otherwise
UNEMP	A dummy variable that takes the value of 1 if the individual is unemployed, 0 otherwise
SCHstud	A dummy variable that takes the value of 1 if the individual is school pupil, 0 otherwise
UNIstud	A dummy variable that takes the value of 1 if the individual is university student, 0 otherwise
Ag_6	A dummy variable that takes the value of 1 if the age of the individual is less than 6, 0 otherwise
Ag6_9	A dummy variable that takes the value of 1 if the age of the individual is between 6 to 9, 0 otherwise
Ag10_13	A dummy variable that takes the value of 1 if the age of the individual is between 10 to 13, 0 otherwise
Ag14_17	A dummy variable that takes the value of 1 if the age of the individual is between 14 to 17, 0 otherwise
Ag18_24	A dummy variable that takes the value of 1 if the age of the individual is between 18 to 24, 0 otherwise

Ag25_44	A dummy variable that takes the value of 1 if the age of the individual is between 25 to 44, 0 otherwise
Ag45_59	A dummy variable that takes the value of 1 if the age of the individual is between 45 to 59, 0 otherwise
Aggr59	A dummy variable that takes the value of 1 if the age of the individual is greater than 59, 0 otherwise
HSIZE	A continuous variable that describes the number of household members in the household
INKhigh	A dummy variable that takes the value of 1 if the income of the household is above €2000, 0 otherwise
CARS	A continuous variable that describes the number of cars owned by the household

Talking about the expected effect of independent variables on personal trip rates, sex of the individual has a greater impact on the frequency of trip making which is supported by the trip distribution in Tab. 3.1, which reveals that although males produced more work trips, overall, females produced the most trips. In the models, the male has been taken as the reference group. Likewise, individuals possessing driving license has a higher propensity of making trips.

It is apparent that all age group individuals do not show similar travel behaviour. Children are likely to make education trips and accompanying trips with their parents, while adults have a higher tendency of making work trips, shopping trips etc. Hence, separate variables were assigned for different age groups to examine the impact of each of the age groups on trip making. The reference group for the age variables was chosen based on the model. Any age group whose effect in the model is insignificant was considered reference group.

Similar to the household-category analysis, employment status of the individuals was also segregated into a full-time worker, part-time worker, trainee and unemployed. Moreover, employed individuals whose scope was not defined was considered as the reference group. It is expected that full-time workers contribute most in work trips followed by part-time workers and trainees. Conversely, unemployed people are expected to perform non-work trips. Further, remaining three household characteristics are expected to have a positive impact on the frequency of trips.

5.2.1 Home-Based Work Trips

As discussed earlier, age group expected to be the least contributor in the trip making would be taken as the reference group. Thus, individuals below 6 have been considered as the reference group in the home-based work trip production model. The result of the regression is presented in Tab. 5.8 below. The result was produced by entering all the variables in stepwise regression and examining the signs of variable coefficients and significance of the variables. No household attributes were found to be significant. It can be observed from the result that full-time workers (FTIME), part-time workers (PTIME) trainee (TRAINEE), individuals of age 18 to 24 (Ag 18_24), age 25 to 44 (Ag25_44), age 45 to 59 (Ag45_59) and female (SEX_FEMALE) influenced the HBW trip production per person. The R^2 value of the model was found to be 0.450 which says that the variables in the model describe 45% of the total variation in the model. Further, RMSE and NRMSE yielded by the model are 0.388 and 74.6 respectively.

Unsurprisingly, full-time workers with higher coefficient show that they are the significant positive contributors of work trips followed by part-time workers and trainees. It is obvious that adults have a higher tendency of making work trips which are supported by the positive coefficients of individuals above 18 till 59. However, the coefficients seem to have lower coefficients than expected. Since adults have a higher propensity of getting employed, it is very likely that age variables and employment variables share the same effect. Likewise, the negative coefficient of female suggests that females are less involved in producing HBW trips.

Tab. 5.8 Estimation results of person-category HBW trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.022	5.210	***
Full-time worker (FTIME)	0.700	101.178	***
Part-time worker (PTIME)	0.535	62.594	***
Trainee (TRAINEE)	0.334	22.592	***
Age group 18-24 (Ag18_24)	0.084	9.296	***
Age group 25-44 (Ag25_44)	0.042	5.749	***
Age group 45-59 (Ag45_59)	0.088	11.642	***
Sex (SEX_FEMALE)	-0.016	-3.418	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 39,404, Adjusted R-Squared: 0.4501, RMSE:0.388, NRMSE:74.6

5.2.2 Home-Based Education Trips

Since elderly people are not expected to contribute in education trips, age above 59 was considered reference group for the home-based education trip production model. On entering all the variables in stepwise regression and examining the signs and significance of the variables, school pupil (SCHstud), university student (UNIstud), trainee (TRAINEE), age below 6 (Ag_6), age 6 to 9 (Ag6_9), age 10 to 13 (Ag10_13), age 14 to 17 (Ag14_17), age 18 to 24 (Ag18_24) and license ownership (LIC) were found to be significant. However, although significant, license ownership which is prevalent among individuals above 18 only was discarded. Even on discarding this variable there was not much change in the R^2 value. Like HBW trips, no household attributes were found to influence HBE trips. The results of the analysis are presented in Tab. 5.9**.** The R^2 value of the final model was 0.531 which shows that the variables in the model explain 53.1% of the total variation in the model. RMSE and NRMSE of the model are 0.254 and 65.8 respectively.

As can be observed from the Tab. 5.9**,** the coefficient of school pupil is the highest which indicates that individuals who are school going student are more likely to make more education trips. Likewise, university students, trainees are also a significant contributor of education trips. Besides age below six, all age variables have lower coefficient than expected though the sign is correct. This is because, individuals of age 6 to 24 are expected to be mostly student, and there are already two variables to represent the student. Hence, it is very likely that the effects are shared between these variables. On the other hand, individuals below six seem to have a higher coefficient than the individuals above six. As mentioned in household-category HBE trips, the reason could be individuals below six are expected to be enrolled in nursery or kindergarten.

Tab. 5.9 Estimation results of person-category HBE trip production model

Variables	**Coefficient**	**t-statistic**	**Significance**
Intercept	0.005	2.986	**
School pupil (SCHstud)	0.547	49.453	***
University student (UNIstud)	0.362	41.498	***
Trainee (TRAINEE)	0.293	27.782	***
Age group <6 (Ag_6)	0.403	57.448	***
Age group 6-9 (Ag6_9)	0.273	21.703	***
Age group 10-13 (Ag10_13)	0.231	17.455	***
Age group 14-17 (Ag14_17)	0.262	21.098	***
Age group 18-24 (Ag18_24)	0.092	13.552	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations 31,407, Adjusted R-Squared: 0.5315, RMSE:0.254, NRMSE: 65.8

5.2.3 Home-Based Shopping Trips

Since younger, as well as elderly individuals, were observed to have a contribution in home-based shopping trips, age 14 to 17 was considered as the reference group. The person-category HBS trip production model is depicted in Tab. 5.10. The variables impacting the model were part-time workers (PTIME), unemployed (UNEMP), age 45 to 59 (Ag45_59), age greater than 59 (Aggr59) and females (SEX_FEMALE). The model yielded R^2 value of 0.0781. The model with variables, part-time workers (PTIME), full-time workers (FTIME), trainees (TRAINEE), age 45 to 59 (Ag45_59), age greater than 59 (Aggr59) and females (SEX_FEMALE) was also significant and yielded R^2 value of 0.0782 which is similar to first one. Since, unemployed and part-time workers are better suited for shopping trips, first one was selected. R^2 value of 0.0781 means the model is able to describe about 8 % of the total variation. Further, RMSE and NRMSE yielded by the model are 0.388 and 74.6 respectively.

Although, poorly fit, all the coefficients have correct signs in the model. The positive signs of PTIME and UNEMP indicate that part-time workers and unemployed individuals have higher tendency to make HBS trips which is not surprising. High magnitude of Aggr59 suggests that elderly people tend to make more HBS trips. In addition, people of age group 45 to 59, also follow the same trend. Expectedly, positive coefficient of SEX_FEMALE hints that females have higher propensity of making HBS trips than males.

Tab. 5.10 Estimation results of person-category HBS trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.137	32.104	***
Part-time worker (FTIME)	0.106	10.319	***
Unemployed (UNEMP)	0.062	8.592	***
Age group 45-59 (Ag45_59)	0.128	16.863	***
Age group >59 (Aggr59)	0.328	43.309	***
Sex (SEX_FEMALE)	0.021	3.385	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 31,404, Adjusted R-Squared: 0.0781, RMSE:0.512, NRMSE:95.5

5.2.4 Home-Based Other Trips

For home-based other trip production model, again age 14 to 17 was considered as the reference group. Tab. 5.11 below illustrates the person-category HBO trip production model. The variables involved in the model are female (SEX_FEMALE), unemployed (UNEMP), age less than six (Ag_6), age greater than 59 (Aggr59) and driving license (LIC). While taking the subsets of a variable from the pool of independent variables, stepwise regression did not help much as it gave almost all variables. Hence, heat and trial in parallel with inspection of signs of the variables' coefficient and their significance led to the final sets of the variable. R^2 value yielded by the model was found to be 0.061, that means, the variables in the model explains about 6.0% of the variations in the model. The model yielded RMSE and NRMSE values of 0.92 and 102.4 respectively.

Although the overall fit of the model was not satisfactory, the variables in the model were significant and had correct sign. In the model, the high magnitude of the coefficient of unemployed variable suggests that unemployed individuals have a higher tendency of making HBO trips. The positive coefficient of individuals of age less than six and age greater than 59 shows that elderly and very small children are likely to make more home-based other trips. The higher rate of HBO trips produced by children below six could be owed to their tendency to accompany their parent. In addition, females and individuals having license have higher chances of making such trips.

Tab. 5.11 Estimation results of person-category HBO trip production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.309	20.076	***
Unemployed (UNEMP)	0.396	30.647	***
Age group <6 (Ag_6)	0.214	8.407	***
Age group >59 (Aggr59)	0.106	7.840	***
License ownership (LIC)	0.283	20.859	***
Sex (SEX_FEMALE)	0.105	10.862	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 31,397, Adjusted R-Squared: 0.0614, RMSE:0.92, NRMSE:102.4

5.2.5 Non-Home-Based Trips

For non-home based trip production model, again age 14 to 17 was considered as the reference group. The person-category NHB trip production model is presented in Tab. 5.12. The variables that affect the NHB trips was found to be ownership of driving license (LIC), age below six (Ag_6), age 25 to 44 (Ag25_44) and female (SEX_FEMALE). The R^2 value of the model is 0.0157 which is very low. The variables in the model are able to describe the variation of about 1.6%. The low R^2 value hints that it is not easy to capture the factors influencing non-home-based trips. Further, RMSE and NRMSE of the model are 1.044 and 99.4 respectively.

In the model, it appears younger individual aged below six contribute most in NHB trips. It can be that children are picked from the kindergarten or nursery and escorted to other destinations. Females and individuals of age 25 to 44 also have positive impact on NHB trips. Likewise, individuals owning license also are likely to make more NHB trips.

Tab. 5.12 Estimation results of person-category NHB production model

Variables	Coefficient	t-statistic	Significance
Intercept	0.373	24.857	***
Age group _6 (Aggr_6)	0.335	10.709	***
Age group 25 to 44 (Ag25_44)	0.162	11.479	***
License ownership (LIC)	0.220	14.205	***
Sex (SEX_FEMALE)	0.095	7.779	***

*** 0.1% significance level, **1% significance level, *5% significance level

Observations: 31,412, Adjusted R-Squared: 0.0157, RMSE:1.044, NRMSE:99.4

5.2.6 Summary

Similar to household-category, person-category HBW and HBE trip production models produced variability above 40% in their corresponding trips. However, remaining three trip production models; HBS, HBO, and NHB yielded very low R^2 value. It is apparent from the models that the trips produced by the individuals of different age groups, sex significantly varied with the trip purposes. It appears that younger individuals contribute significantly in education trips while other trips are dominated by adults. Further, repetition of age below six suggests that younger individuals are likely to follow their parents. Females are a dominant contributor of HBS, HBO and NHB trips.

5.2.7 Discussion

In this section, trip production models developed in this thesis have been related to previous studies. Although most of the previous studies being compared were not distinguished by home-based and non-home-based categories as in the thesis, and different variables have been included, some trends still can be compared.

The findings that the number of employed (full-time, part-time and trainees) positively affect the work trip frequency confirms to other studies. As Hu (2010) found full-time and part-time workers have a significant positive contribution in producing work trips. Similarly, Chang, Jung et al. (2014) also reported that as the number of employed members increases, households tend to make more home-based work trips. Further, person-category work trip model by

Prevedouros and Schofer (1991) also revealed that individuals have a higher propensity of making work trips when they are full-time workers.

The positive impact of household income and number of cars in work trips have been observed in (Goulias, Pendyala et al. 1990, Hu 2010, Chang, Jung et al. 2014). However, although the results from the thesis confirm the influence of cars, it failed to recognise the effect of household income on work trips. Surprisingly, the household size which was found significant in shopping trips and social trips in Detroit Metropolitan area (Goulias, Pendyala et al. 1990) was relevant in only one case i.e. NHB trips. That too was the logarithm of the household size. The logarithm of household size was found significant in Toronto (Badoe and Steuart 1997).

Goulias, Pendyala et al. (1990) reported that the school trips increase with an increase in the number of children aged 6 through 18 and decrease if there is presence of children below six in the household. Since the model dealt with school trips only, it is likely that children below six had a negative impact in the model. On the other hand, thesis included all education trips, therefore, school student, university student as well as children below six were found significant.

Person-category trip production revealed that sex is an important variable in explaining travel frequency. The sign of the sex variable disclosed that male is still dominant contributor of home-based work trips while females were found to have a positive impact on home-based shopping and home-based other trips. This confirms the study by (Hanson and Hanson 1980) that males lead in work trips while females have more penchant for shopping trips. Further, as in this study, male contributed positively in work trips and negatively in non-work trips in Chicago (Prevedouros and Schofer 1991). Moreover, in the same study, full-time workers were found to have a positive effect on work trips and negative effect in non-work trips which support the findings of the thesis.

Comparing the model performance with the previous studies, it appears satisfactory for household-category models. HBW trip model in the thesis with R^2 value of 0.483 is lower than 0.584 yielded by HBW trip model in Toronto (Badoe and Steuart 1997) and marginally higher than 0.40, yielded by HBW trip model in Seoul (Chang, Jung et al. 2014). Likewise, R^2 value of 0.531 yielded by HBE trip model in the thesis is slightly lower than that of 0.545 yielded by school trip model in Denver (Prevedouros and Schofer 1991). Further, HBS in the thesis yielded R^2 value of 0.0524 which is similar to 0.045 obtained by HBS in Toronto (Badoe and Steuart 1997). In the same way, concerning person-category model, HBW trip model in the thesis yielded R^2 value 0.4501 which is higher than 0.33 yielded by work trip model in Chicago (Prevedouros and Schofer 1991). Further, R^2 value yielded by HBS trip model in the thesis was 0.0782 against 0.135 obtained by shopping trip model in Edinburgh (Hu 2010).

On comparing the model results with the previous studies conducted in different places and different time period, it is apparent that the trip production models in the thesis are satisfactory. Further, as in this study, it can be observed in previous studies as well that work and school trip models were the best performing models while other models were poorly fit.

6 Trip Production Modelling by Cross-classification

In this section household-category and person-category trip production models were analysed by cross-classification. Two-way cross-classification were employed for both levels. Unlike, regression, findings were not compared with previous studies as models in the thesis and studies were defined differently.

6.1 Household-Category

Trip production rates per household for a given purpose by cross-classification are a function of household attributes. For each trip categories, four cross-classifications were tested. At the end, the RMSE and NRMSE were evaluated to choose the best cross-classification. Although more variables were included in the regression, it was not practical to incorporate all of them as it results in a very high number of categories.

Car ownership, household size, and household income are common variables that were used in every cross-classification models besides other significant variables taken from regression. Four categories of car ownership have been used; zero cars, one car, two cars, and three or more cars. Likewise, five categories of household size have been used; one member, two members, three members, four members and five or more members. Regarding income, it has been squeezed from 15 categories in the original dataset to five categories. Since it is difficult to incorporate all 15 income groups into cross-classification, it was felt essential to reclassify the income, and it was done based on the steepness of the income versus household trip rate which is illustrated in Fig. 6.1.

Ink1, first income category includes income less than € 1,500 as the increment of household trips with an increase in income appears to be similar in these segment. This group comprises about 26% of the households. Similarly, fourth and fifth income groups (€ 1,500 - € 2,600) with 32% households and sixth and seventh income groups (€ 2,600 - € 3,600) with 16% households seem to exhibit similar progression of trips with respect to increase in income were grouped into Ink2 and Ink3 respectively. The households having income beyond € 3,600 comprise below 15%. Thus, remaining six groups were compressed into further two income groups. Income € 3600 to € 5,000 with 9% households where trip rate appears to increase at first and decrease and again increase for the last time were grouped into income group Ink4 despite irregular progression of trips in these segment. Finally, income above € 5,000 comprising 5% households where household trips start decreasing with increasing income were grouped into final income group Ink5.

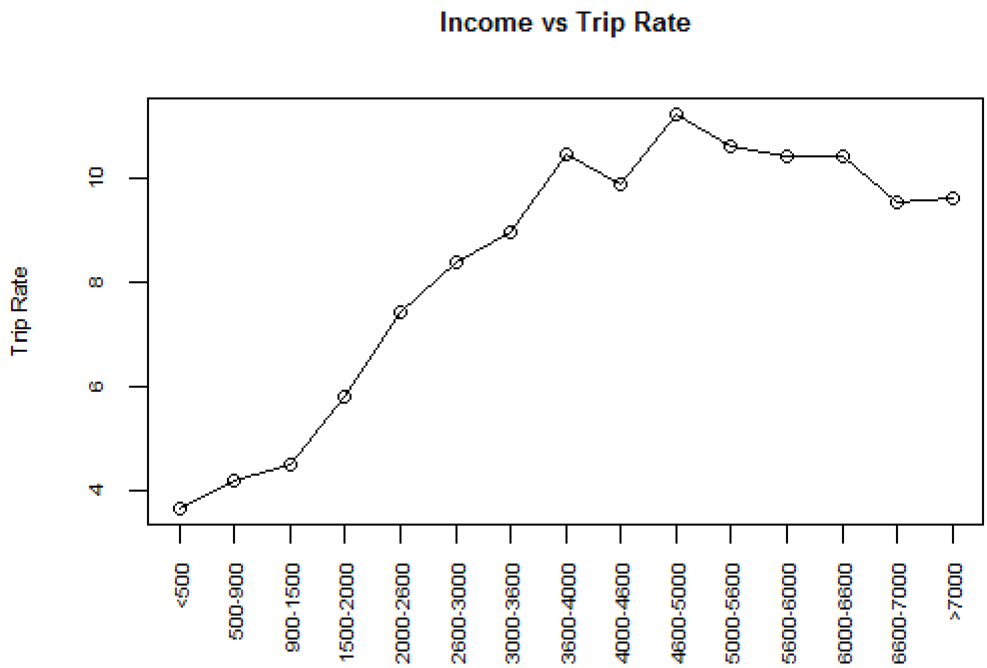


Fig. 6.1 Income vs household trip rate

6.1.1 Home-based Work Trips

From the regression estimate of HBW trip production regression, the number of workers (full-time, part-time and trainee all combined) had a higher effect. Hence, the number of workers categorised as zero workers, one worker, two workers, three or more workers was cross-classified once with income and once with car ownership in addition to two other cross-classification of household size with income and car ownership. Four cross-classifications with their respective RMSE and NRMSE values are presented in Tab. 6.1 below.

Tab. 6.1 HBW cross-classification stratifications

Cross-classification Stratifications	**RMSE**	% **RMSE**	R^2
Workers (0,1,2 and 3+) and income (5 groups)	0.659	74.1	0.45
Workers (0,1,2 and 3+) and cars (0,1,2 and 3+)	0.668	74.7	0.44
Household size (1,2,3,4 and 5+) and income (5 groups)	0.786	88.5	0.217
Household size (1,2,3,4 and 5+) and cars (0,1,2 and 3+)	0.790	88.3	0.22

Cross-classification of workers and income which appears to have the lowest RMSE and NRMSE was chosen as a final model. Tab. 6.2 below presents the HBW trip production rates for the chosen combination. The number of households with more than three workers and income below € 1,500 was found 13 which is very low. Hence, this cell was aggregated with the household type which has more than three workers and household income between € 1,500 and € 2,600 to give aggregated HBW trip production rates. HBW trip production varied from 0.032 trips per household to 1.973 trips per household. The results appear to be logical as it is apparent that in most cases, HBW increase as household and number of employed members in the household increase. Exceptions were observed in four cells which are marked with *. However, the increment is marginal.

Tab. 6.2 HBW trips per household

Monthly Income	Employed Members			
	Zero	One	Two	Three+
Ink1 (<€ 1,500)	0.032 (1609)	0.689 (754)	0.902 (110)	- (13)
Ink2 (€ 1,500 - € 2,600)	0.032 (1957)	0.734 (1532)	1.272 (1058)	1.831 (146+13)
Ink3 (€ 2,600 - € 3,600)	0.031* (748)	0.751 (777)	1.295 (1291)	1.970 (286)
Ink4 (€ 3,600 - € 5,000)	0.035 (186)	0.726* (366)	1.340 (855)	1.973 (244)
Ink5 (>€ 5,000)	0.025* (63)	0.749 (154)	1.391 (465)	1.916* (244)
Number of observations=12,577, R^2=0.45, RMSE=0.659, NRMSE=74.1				

Note: The figures in the parentheses indicate the number of households in each category

The results are presented graphically in Fig. 6.2. Graph of other three cross-classification can be viewed from Appendix D. It shows that due to marginal increment of the trips with an increase of cars, the line graphs appear to be almost flat.

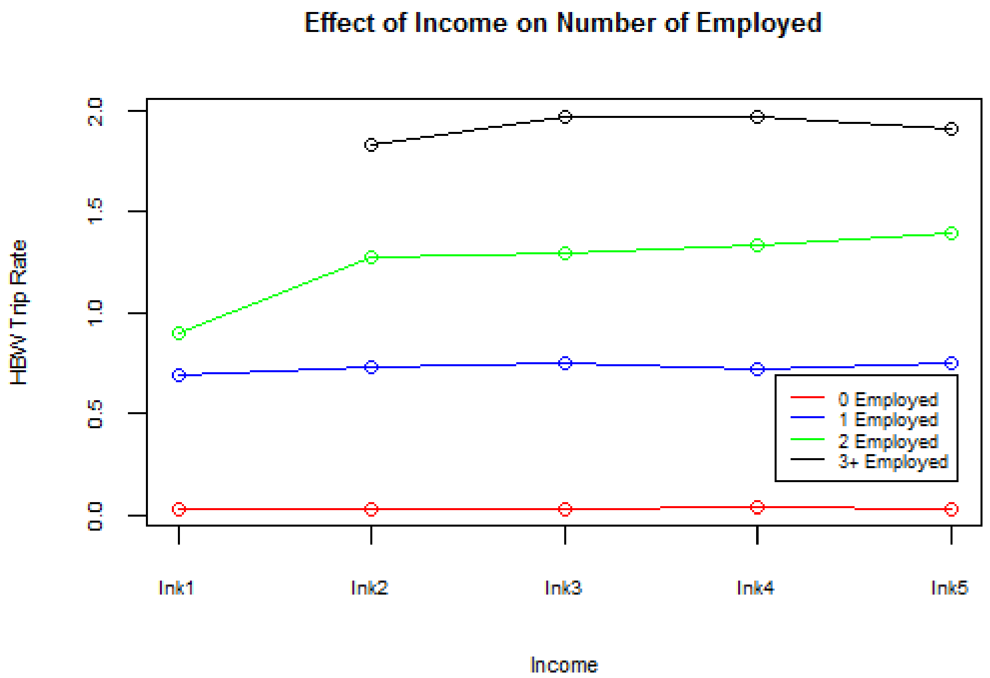


Fig. 6.2 Effect of income on number of employed (HBW trips)

The HBW trips per household predicted from the model and the observed values are illustrated in Fig. 6.3. The observed HBW trips per household are as high as almost eight while the highest predicted trips are about two.

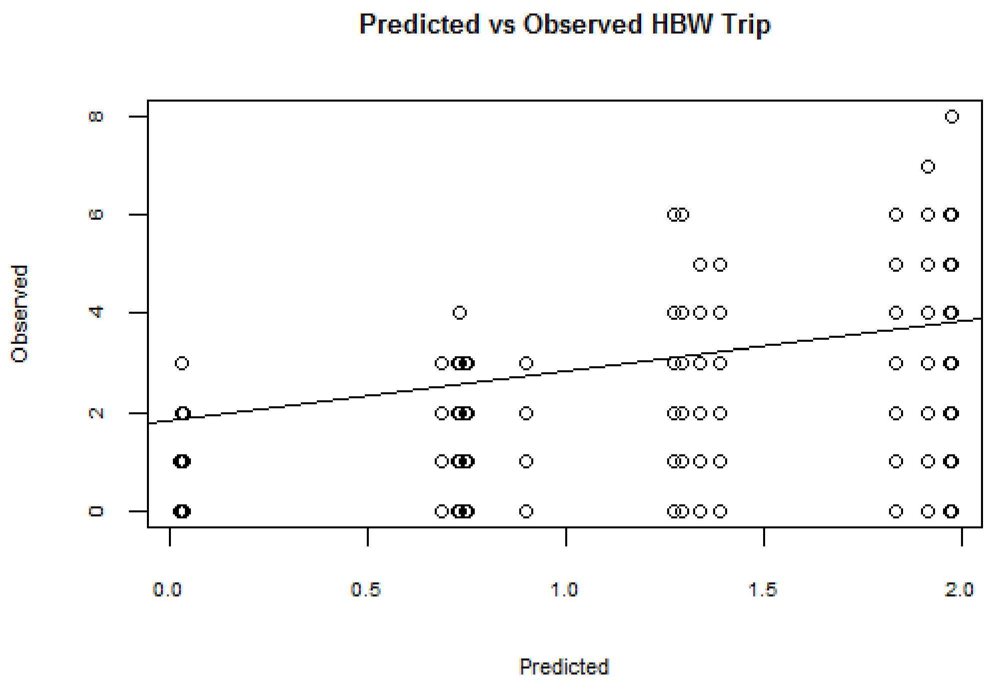


Fig. 6.3 Predicted vs observed HBW trips per household

6.1.2 Home-based Education Trips

Since younger individuals in the household are more likely to make HBE trips, individuals aged below 18 categorised as zero children, one child, two children and three or more children was cross-classified with income and cars separately along with two other cross-classification of household size once with income and once with car ownership. Four cross-classifications with their respective RMSE and NRMSE values are shown in Tab. 6.3.

Tab. 6.3 HBE cross-classification stratifications

Cross-classification Stratifications	RMSE	% RMSE	R^2
Children (0,1,2 and 3+) and income (5 groups)	0.540	72.2	0.479
Children (0,1,2 and 3+) and cars (0,1,2 and 3+)	0.524	70.2	0.507
Household size (1,2,3,4 and 5+) and income (5 groups)	0.562	75.1	0.435
Household size (1,2,3,4 and 5+) and cars (0,1,2 and 3+)	0.554	74.3	0.448

As can be observed, the stratification of children and car produced the highest R^2 value and the lowest RMSE and NRMSE values. Thus, it was selected for the HBE trip production analysis whose results are depicted in Tab. 6.4. The cell with households having more than three children and owning zero cars whose observations was less than 30, was aggregated with the nearest cell of households having more than three children and owning one car. Likewise, for the same reason, the cell with household having more than three children and possessing more than three was aggregated with the cell of households having two cars and more than three children.

The HBE trip production rates varied from a low value of 0.047 for households having zero student and zero cars to a high value of 2.456 for households with more than three children and more and two or more than two cars. In general, trip progression looks logical with the exception in two cells which are marked * in the table. The HBE trip production appears to increase with respect to increase in the number of cars and children per household. However, the increase with respect to the number of children is substantial and overshadows the another effect. Fig. 6.4 presents the results in graphical form. Other three graph of cross-classification results can be viewed from Appendix D. Due to the marginal trip increment rate, a graph of the effect of car ownership on student appears almost flat except households with three or more students.

Tab. 6.4 HBE trips per household

Cars per Household	**Children (<18)**			
	Zero	**One**	**Two**	**Three+**
Zero	0.047 (1477)	0.576 (91)	1.306 (49)	- (20)
One	0.028* (5840)	682 (769)	1.355 (605)	1.935 (190+20)
Two	0.084 (2793)	0.707 (856)	1.389 (758)	2.456 (179+24)
Three+	0.242 (725)	0.804 (187)	1.239* (70)	- (24)
Number of observations=14,633, R^2=0.507, RMSE=0.524, NRMSE=70.2				

Note: The figures in the parentheses indicate the number of individuals in each category.

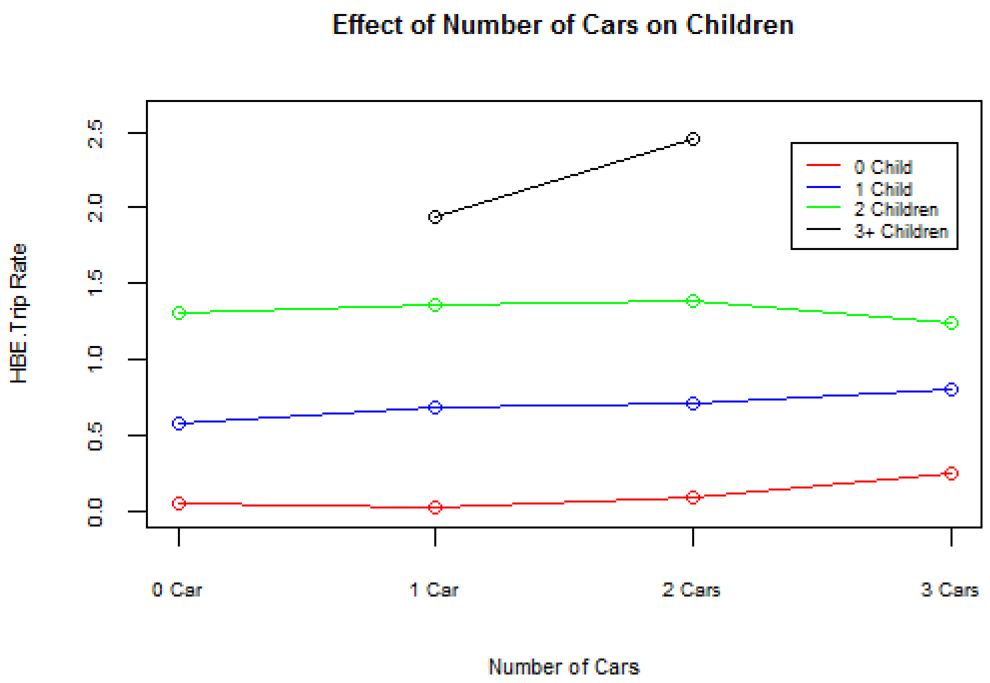


Fig. 6.4 Effect of Cars on number of children (HBE trips)

Fig. 6.5 below illustrates the plot of HBE trips per household predicted by the chosen cross-classification model versus observed HBE trips per household. The highest predicted value is about six while the highest observed value is just below 2.5.

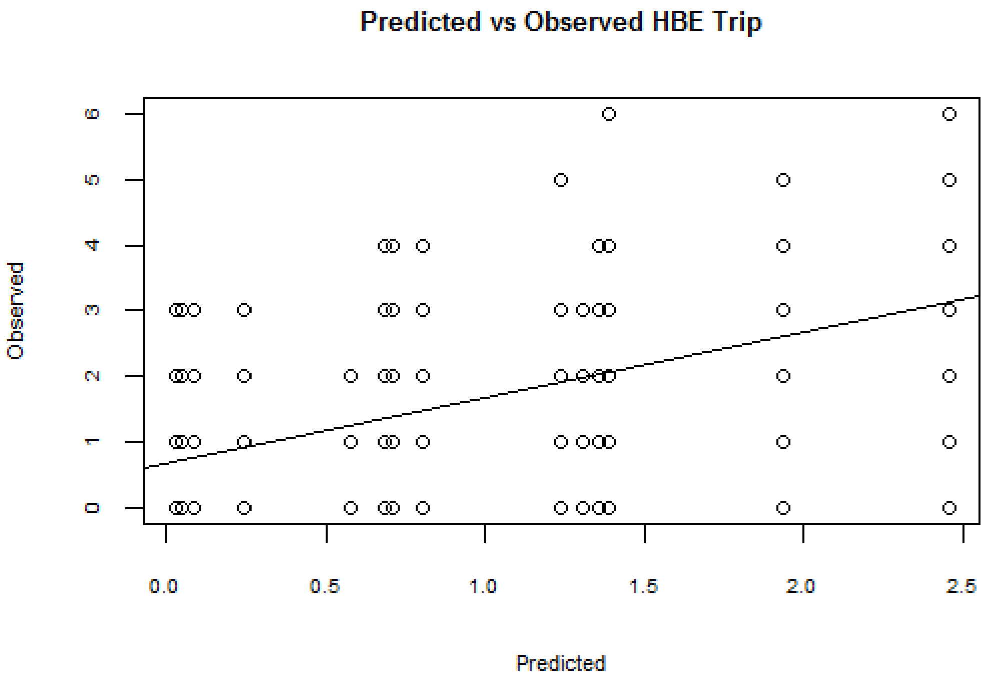


Fig. 6.5 Predicted vs observed HBE trips per household

6.1.3 Home-based Shopping Trips

During regression analysis of HBS trip production, the number of unemployed members appeared to have a higher tendency of making HBS trips in the household. Thus, the number of unemployed members with categories zero unemployed, one unemployed, two unemployed and three or more unemployed was cross-classified with income and cars separately along with two other cross-classification of household size once with income and once with car ownership. All these cross-classifications with their respective RMSE and NRMSE values are shown in Tab. 6.5.

Tab. 6.5 HBS cross-classification stratifications

Cross-classification Stratifications	**RMSE**	**% RMSE**	**R^2**
Unemployed (0,1,2 and 3+) and income (5 groups)	0.855	98.1	0.038
Unemployed (0,1,2 and 3+) and cars (0,1,2 and 3+)	0.853	97.9	0.041
Household size (1,2,3,4 and 5+) and income (5 groups)	0.863	99.0	0.02
Household size (1,2,3,4 and 5+) and cars (0,1,2 and 3+)	0.863	99.0	0.02

The cross-classification stratified by unemployed and cars owned by the households was chosen, as its performance is marginally better than the rest in terms of R^2, RMSE, and NRMSE values. Furthermore, the graph of the results in the case of cross-classification of unemployed and cars owned by the household looks more convincing compared to rest. The results of the selected stratification are graphically presented in Fig. 6.6 and rest graphs can be referred from Appendix D. Tab. 6.6 shows the HBS trip production rates. As can be seen, HBW trip rates varied from a low value of 0.291 for zero unemployed members and zero cars owned by households to 1.013 for more than three unemployed members and more than three household cars. HBS trip production rates seem to increase with the increase in unemployed members and the number of cars owned by the household with the exception in two cells which are marked *. It shows that the relationship between HBS and car ownership is non-linear.

Tab. 6.6 HBS trips per household

Cars per Household	Unemployed Members			
	Zero	**One**	**Two**	**Three+**
Zero	0.291 (284)	0.548 (871)	0.877 (415)	0.723 (67)
One	0.343 (1502)	0.572 (2096)	0.889 (3237)	0.764 (569)
Two	0.455 (1136)	0.590 (1309)	0.710* (1547)	0.787 (594)
Three+	0.462 (283)	0.724 (336)	0.703* (261)	1.013 (126)
Number of observations=14,633, R^2=0.041, RMSE=0.853, NRMSE=97.9				

Note: The figures in the parentheses indicate the number of households in each category.

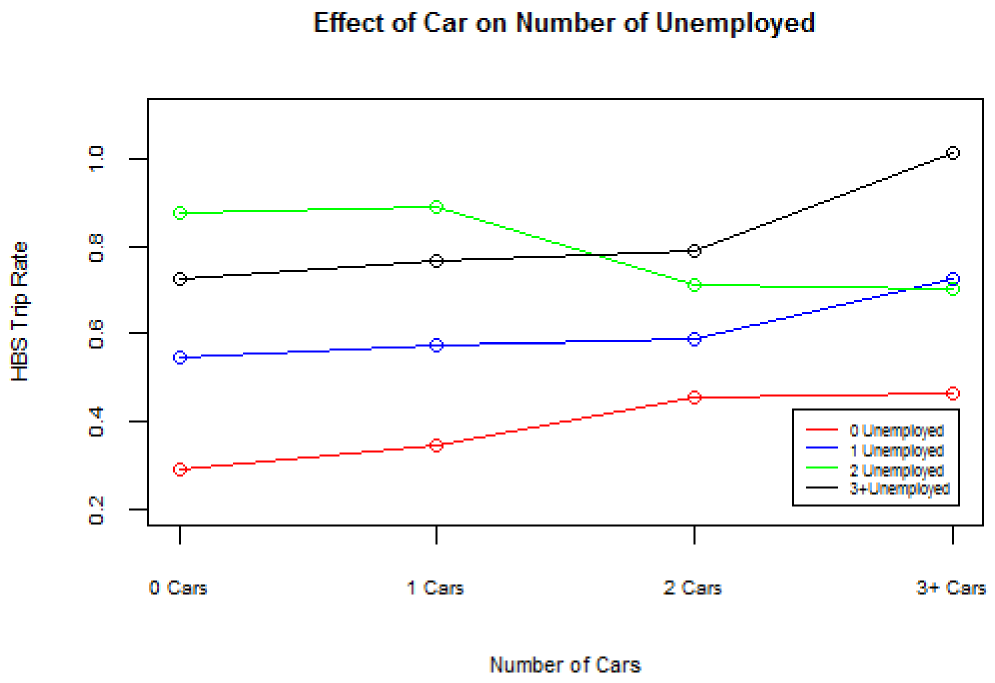


Fig. 6.6 Effect of cars on number of unemployed (HBS trips)

Fig. 6.7 demonstrates the plot of predicted versus observed HBS trips per household. As can be seen, the highest observed value is about eight while the highest predicted value is just above one.

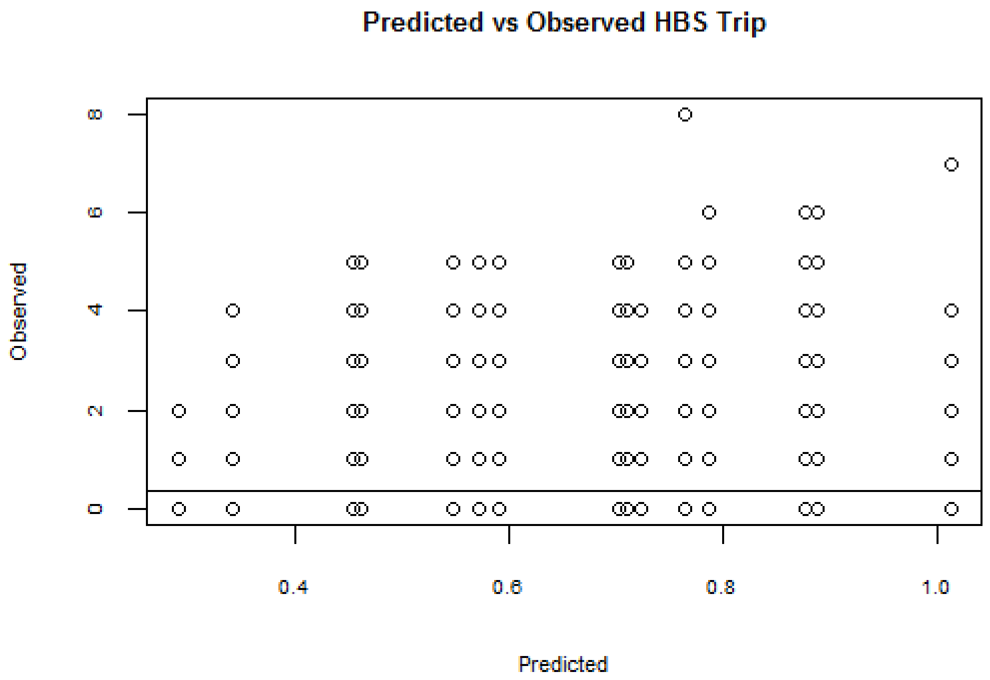


Fig. 6.7 Predicted vs observed HBS trips per household

6.1.4 Home-based Other trips

For regression analysis of HBO trip production, the number of unemployed members seemed to influence more on HBO trips in the household. Thus, the number of unemployed members with categories zero unemployed, one unemployed, two unemployed and three or more unemployed was cross-classified with income and cars separately along with two other cross-classification of household size once with income and once with car ownership. All these cross-classifications with their respective RMSE, NRMSE, and R^2 values are shown in Tab. 6.7.

Tab. 6.7 HBO cross-classification stratifications

Cross-classification Stratifications	**RMSE**	**% RMSE**	**R^2**
Unemployed (0,1,2 and 3+) and income (5 groups)	1.594	90.6	0.179
Unemployed (0,1,2 and 3+) and cars (0,1,2 and 3+)	1.605	90.6	0.167
Household size (1,2,3,4 and 5+) and income (5 groups)	1.604	91.2	0.169
Household size (1,2,3,4 and 5+) and cars (0,1,2 and 3+)	1.616	91.3	0.167

On examining the NRMSE and RMSE value of all the stratifications, it was found that the cross-classification stratified by unemployed members and household income was marginally better compared to rest. In addition, the graph of the results yielded by cross-classification stratified by unemployed members and household income shows comparatively more heterogeneity compared to rest. The graph of the selected combination is illustrated in Fig. 6.8 and graph of rest of the combinations can be seen at Appendix D. HBO trip production rates for this combination is demonstrated in Tab. 6.8. The least HBO trips per household was 0.555 which was found in households with zero unemployed members and income below € 1,500. Likewise, with 3.91 HBO trips per household, households with more than three unemployed members and household income greater than € 5,000 produced highest HBO trips. HBO appears to increase with the higher income of the household with the exception in one cell marked *.

These results are depicted in Fig. 6.8 which shows the relationship between HBO trips and monthly income of the household is approximately linear in all cases.

Tab. 6.8 HBO trips per household

Monthly Income	**Unemployed Members**			
	Zero	**One**	**Two**	**Three+**
Ink1 (<€ 1,500)	0.555 (454)	1.118 (1213)	1.746 (715)	2.591 (104)
Ink2 (€ 1,500 - € 2,600)	0.728 (924)	1.365 (1302)	1.979 (2118)	3.287 (349)
Ink3 (€ 2,600 - € 3,600)	0.904 (748)	1.809 (809)	2.252 (1024)	3.493 (337)
Ink4 (€ 3,600 - € 5,000)	1.153 (440)	2.062 (452)	2.592 (527)	3.823 (226)
Ink5 (>€ 5,000)	1.199 (234)	1.911* (234)	2.746 (248)	3.912 (113)
Number of observations=12,577, R^2=0.179, RMSE=1.594 NRMSE=90.6				

Note: The figures in the parentheses indicate the number of households in each category.

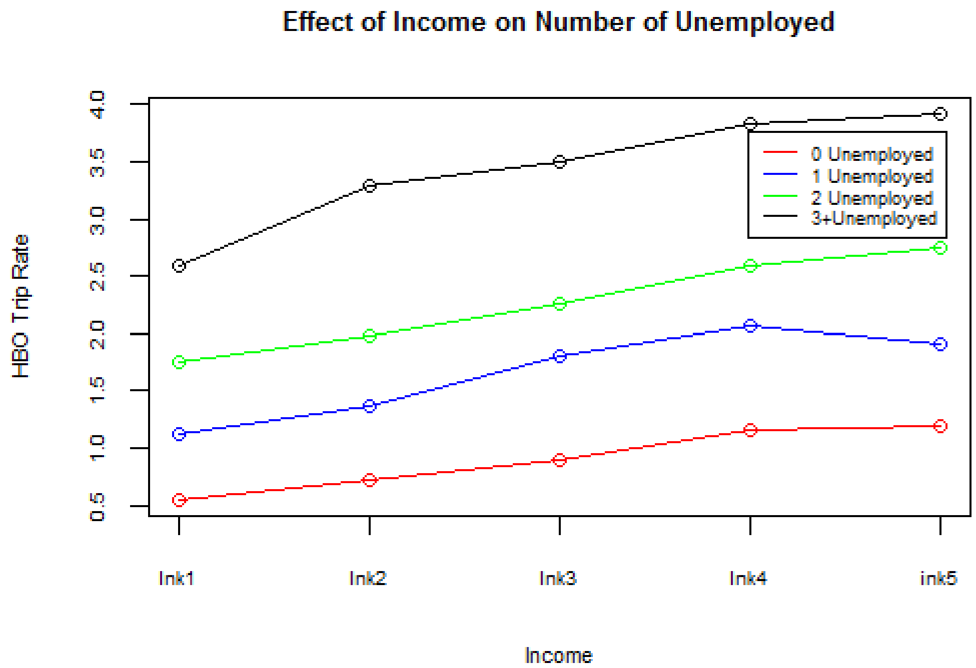


Fig. 6.8 Effect of Income on number of unemployed (HBO trips)

Fig. 6.8 presents the plot of predicted versus observed HBO trips per household. It is apparent from the figure that observed value is as high as about 14 while predicted value is just below four. The line of best fit is parallel to the baseline.

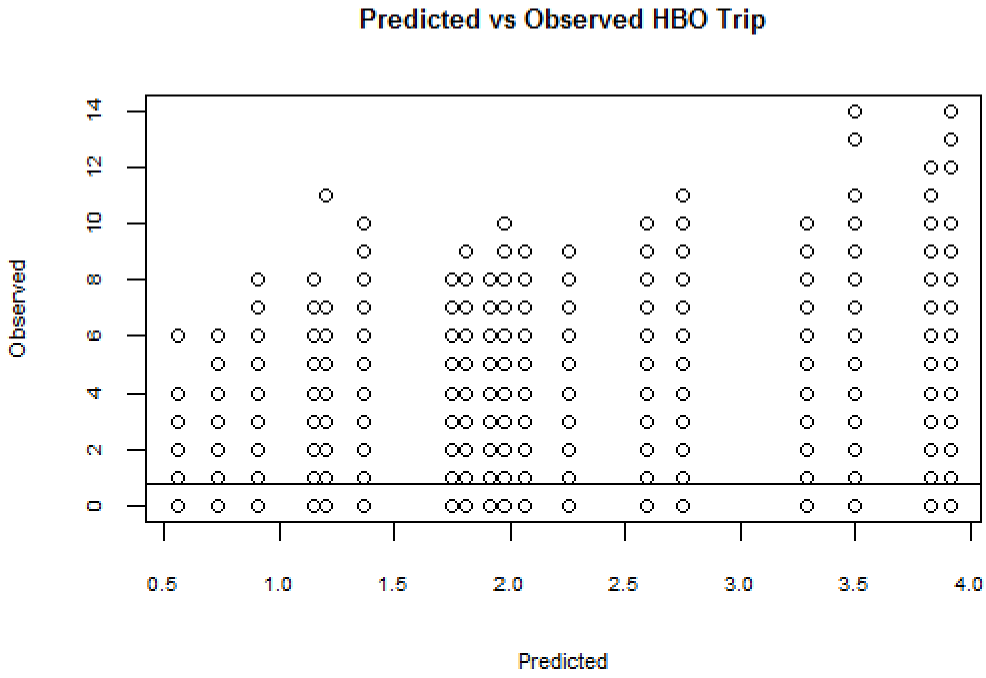


Fig. 6.9 Predicted vs observed HBO trips per household

6.1.5 Non-Home-Based Trips

Regression analysis of NHB trip production revealed the contribution of unemployed members in producing NHB trip. Thus, the number of the unemployed member was categorised as zero unemployed, one unemployed, two unemployed and three or more unemployed was cross-classified once with income and once with cars along with two other cross-classification of household size once with income and once with car ownership. All these cross-classifications with their respective RMSE and NRMSE values are demonstrated in Tab. 6.9.

Tab. 6.9 NHB cross-classification stratifications

Cross-classification Stratifications	**RMSE**	% **RMSE**	R^2
Unemployed (0,1,2 and 3+) and income (5 groups)	1.939	97.0	0.058
Unemployed (0,1,2 and 3+) and cars (0,1,2 and 3+)	1.937	97.9	0.041
Household size (1,2,3,4 and 5+) and income (5 groups)	1.986	99.4	0.012
Household size (1,2,3,4 and 5+) and cars (0,1,2 and 3+)	1.922	97.2	0.056

From the above stratifications and their corresponding RMSE, NRMSE and R^2 values, cross-classification stratified by household size and cars owned by the household and cross-classification stratified by unemployed members and household income appear to show similar performances in terms of R^2, RMSE, and NRMSE values. However, for the analysis commonly used household size and cars were preferred, whose results are demonstrated in Tab. 6.10 below.

Expectedly, the household with one member having three or more cars is short of observations. Thus, it was aggregated with 62 observations from the households with one member and two cars. Similarly, the households with four or more than four members and having car zero also fell short of 30 records. Thus, it was merged with the households of type four or more members and one car.

The NHB trip production rates varied from a low value of 0.60 to a high value of 2.719. The trip rates seem fairly logical in terms of increment of trips with respect to increasing number of household members but had some exception with respect to increase in car ownership. The exceptions are marked * in the table. It is apparent from the table and graph in Fig. 6.10 that as the household size increase, the impact of the number of cars on the non-home-based trip seems irregular which is supported by the line graph of household with three members and four or more members.

Tab. 6.10 NHB trips per household

Cars per Household	**Household Size**				
	One	**Two**	**Three**	**Four**	**Four+**
Zero	0.600 (934)	0.844 (535)	1.539 (100)	1.699 (49)	- (19)
One	0.827 (1502)	1.089 (4100)	1.494* (873)	2.117 (726)	2.309 (255+19)
Two	0.895 (62+9)	1.221 (1842)	1.517 (1210)	1.901* (1140)	2.719 (332)
Three+	- (9)	1.243 (93)	1.490* (348)	1.510* (390)	2.131* (166)
Number of observations=14,632, R^2=0.056, RMSE=1.922, NRMSE=97.2					

Note: The figures in the parentheses indicate the number of households in each category.

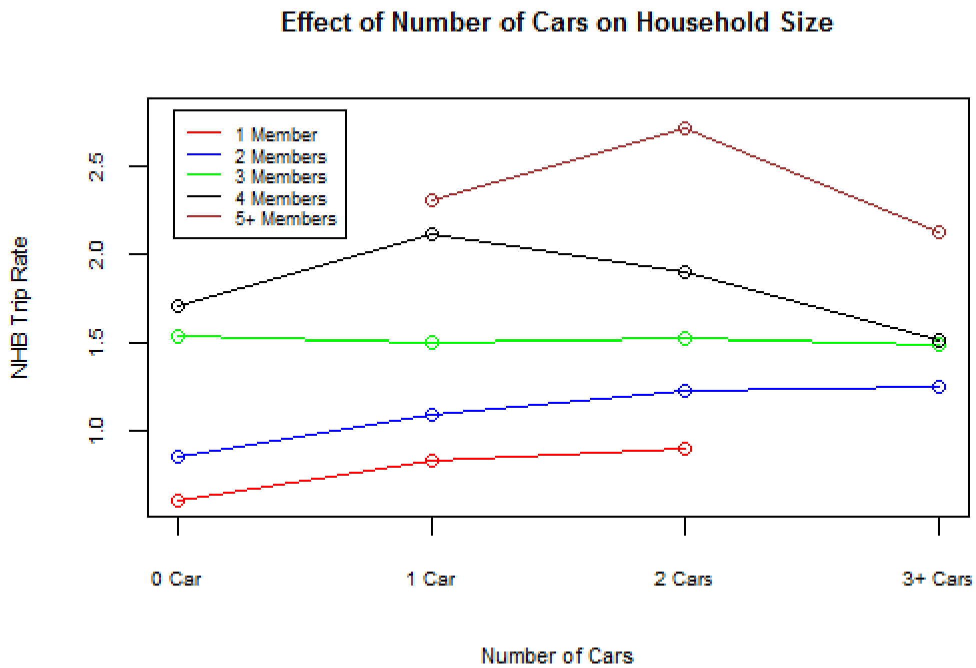


Fig. 6.10 Effect of income on number of unemployed (NHB trips)

Fig. 6.11 presents the plot of predicted versus observed NHB trips per household. It can be observed that highest observed value is about 30 which is very high compared to the high predicted value of below three. The higher values are outliers as they are drifting the line of best fit towards up.

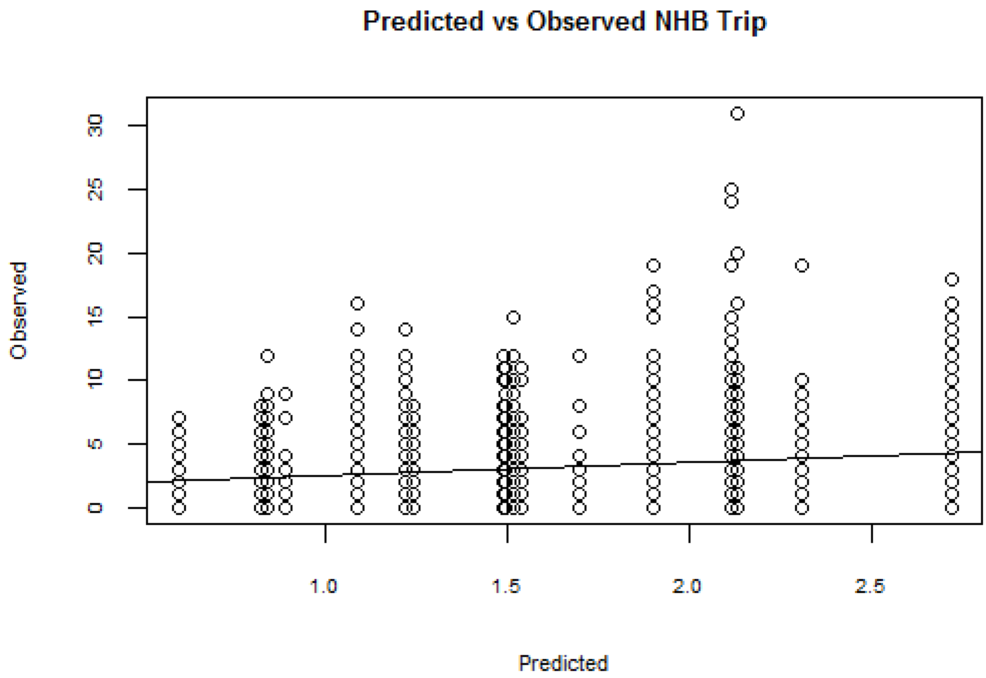


Fig. 6.11 Predicted vs observed NHB trips per household

6.1.6 Summary

Four cross-classification combinations were assessed for each of the five household-category models. The number of cars owned by the household was the most repeated variable followed by income and unemployed members. However, besides, the impact of cars on HBO trips, the impact of cars and income on trip making was not significant. Further, RMSE value of the model shows HBE had the least errors followed by HBW, HBS, HBO, and NHB.

6.2 Person-Category

As witnessed from the trip production analysis by regression, trip production rates per person are predominantly a function of personal attributes. Thus, (Supernak, Talvitie et al. 1983) and regression results were referred for selecting cross-classification stratification. In all models tested, age group was retained as one dimension of categorisation; the other dimensions were sex, employment status, student status and ownership of the license. Depending upon the trip purposes, five to six categories of ages were employed. Since younger individuals are unlikely to be employed, all models cross-classified by employment status and age group considered five age groups (age less than 18, age 18 to 24, age 25 to 44, age 45 to 59 and age more than 59). In the same way, models cross-classified by age against student status and sex considered six age groups (age less than 10, age 10 to 17, age 18 to 24, age 25 to 44, age 45 to 59 and age more than 59).

6.2.1 Home-based Work Trips

Home-based work (HBW) trip production rate was cross-classified by employment status (full-time worker, other worker and non-worker) and age (five age groups). Result is presented in Tab. 6.11. Other workers comprised part-time workers, trainees, and workers whose scope of work was undeclared. It is not surprising that full-time workers below 18 years of age had low observations. Due to just eleven observations falling in that category which is short of at least 30 observations needed in a cell, it was aggregated with full-time working individuals of age 18 to 24.

Tab. 6.11 HBW trips per person

Employment Status	**Age Groups**				
	Ag<18	Ag18_24	Ag25_44	Ag45_59	Ag>59
Full-time worker	- (11)	0.764 (552+11)	0.764 (3467)	0.802 (4972)	0.718 (544)
Other worker	0.386 (172)	0.453 (1362)	0.541 (1541)	0.649 (2064)	0.629 (318)
Non-worker	0.004 (5298)	0.123 (1263)	0.065 (1095)	0.078 (1691)	0.012 (7692)
Number of observations=31,407, R^2=0.444, RMSE=0.388, NRMSE=74.6					

Note: The figures in the parentheses indicate the number of individuals in each category.

HBW trip production rates varied from almost zero for non-workers below 18 years old to 0.802 trips for workers of age group 45 to 59. For a clear overview of variation of trips, Fig. 6.12 can be referred. It shows that full-time workers produced the highest number of HBW trips and it seems to increase linearly with increase in age until 59 and starts decreasing after that. Other employed individuals appear to produce lesser HBW trips than that of full-time workers. Further, as age increases, trips produced by these workers also increases until the age of the individuals is above 59.

Needless to say, individuals who are unemployed make fewer work trips which were reflected in this cross-classification model. However, among unemployed ones, individuals of age group 18 to 24 produced comparatively more HBW trips. This could be because they might be making job search trips. In all three cases, the decline of HBW trips after the age of 59 is common which is not startling. The RMSE and NRMSE value of the above cross-classification is 0.388 and 74.6 respectively.

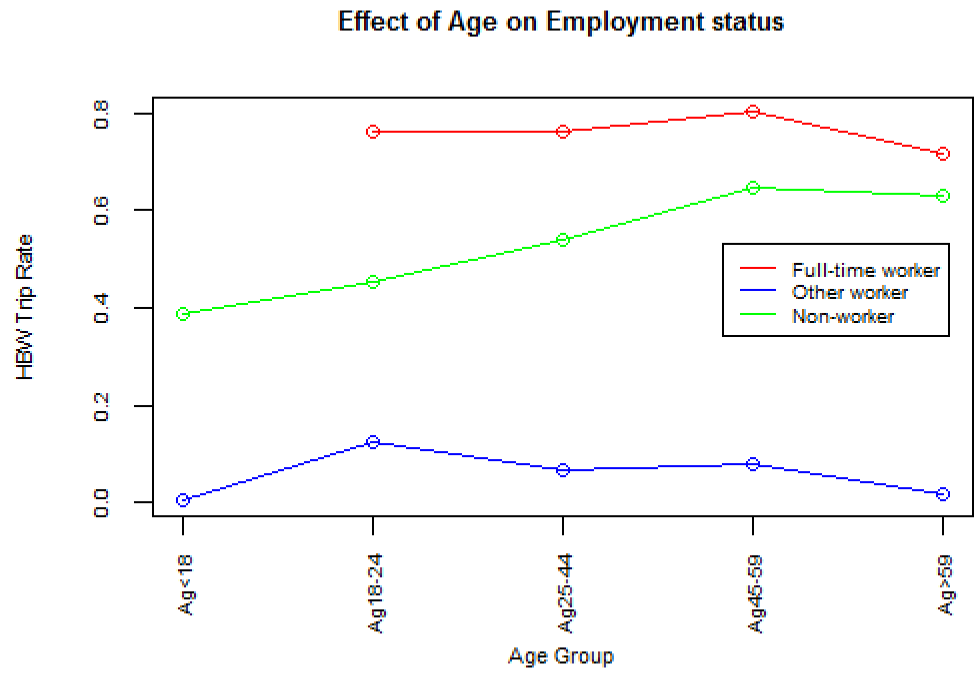


Fig. 6.12 Effect of age on employment status (HBW trips)

Fig. 6.13 demonstrates the plot of predicted versus observed HBW trips per person. It is clear from the figure that observed values which are as high as five compared to the high observed value of less than 0.8 are pulling the line of best fit upwards.

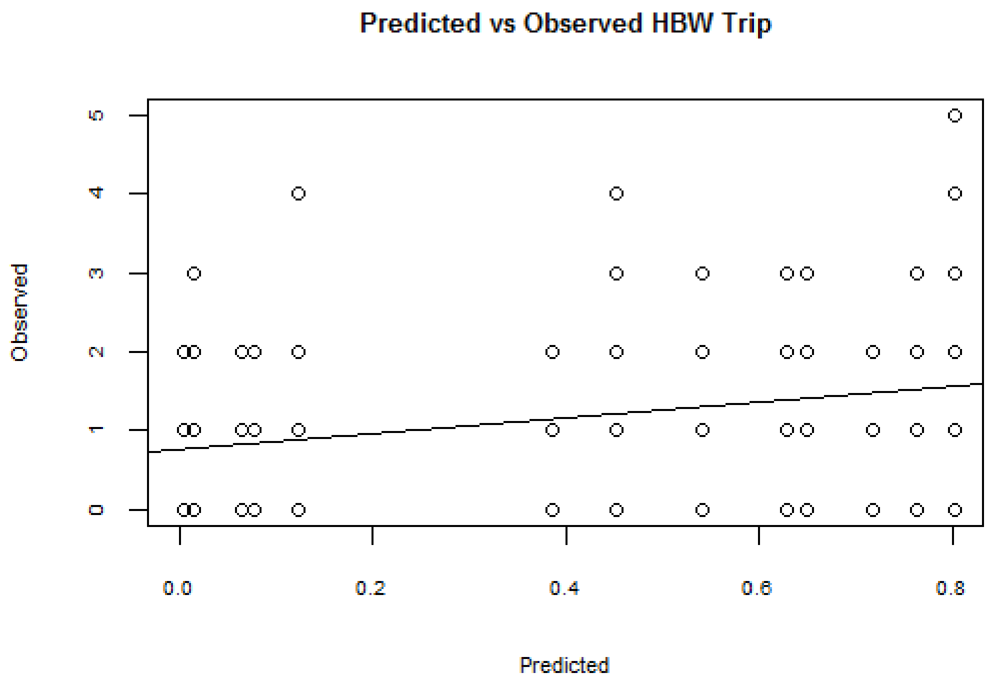


Fig. 6.13 Predicted vs observed HBW trips per person

6.2.2 Home-based Education Trips

Home-based education (HBE) trip production per person stratified by age (six groups) and student status (student and non-student) is depicted in Tab. 6.12 below. It is not unexpected that the number of individuals of age group 45 to 59 and age greater than 59 claiming to be student were six and one respectively. Since the record was too low to consider for calculating trip rates, these individuals were aggregated with individuals of age 25 to 44.

Tab. 6.12 HBE trips per person

Student Status	**Age Groups**					
	Ag<10	Ag10_17	Ag18_24	Ag25_44	Ag45_59	Ag>59
Student	0.787 (1000)	0.801 (3146)	0.577 (1042)	0.361 (245)	- (6)	- (1)
Non-student	0.428 (1140)	0.458 (203)	0.176 (1500)	0.015 (5868)	0.004 (8717+6)	0.001 (8553+1)
Number of observations=31,422, R^2=0.186, RMSE=0.257 NRMSE=66.6						

Note: The figures in the parentheses indicate the number of individuals in each category.

The results from the cross-classification is presented graphically in Fig. 6.14. Students below 18 years of old had almost similar HBE trip production rates. However, as the age of the individuals who are student increase beyond 17, the rate appeared to decline linearly as we can see in the figure. On the other hand, even the non-students seemed to make trips. The tendency was high in age group less than ten and age group 10 to 17. The main reason for the individuals below ten despites of not being a student made trips could be that the younger individuals in that category are admitted in nursery and kindergarten. Elder non-students making trips could be owed to trainees. Similar to individuals who are a student, HBE trips produced by non-student seemed to linearly decrease as the age increases from 18 and started flattening after age 25 to 44. The RMSE and NRMSE value of the cross-classification were found to be 0.257 and 66.6 % respectively.

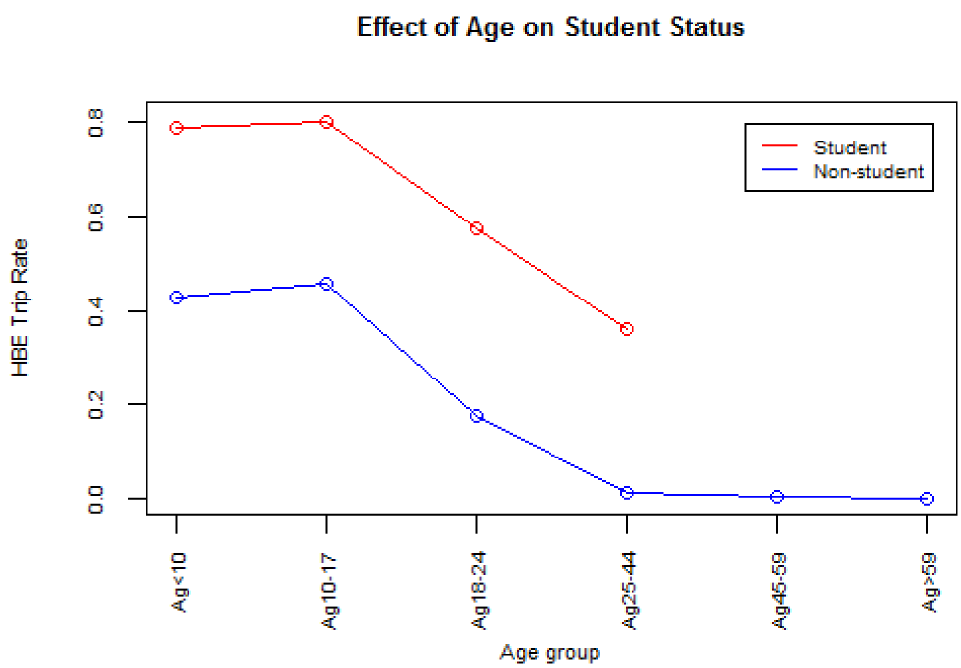


Fig. 6.14 Effect of age on student status (HBE trips)

Fig. 6.15 presents the graph of predicted versus observed HBE trips per person. As can be seen, the graph contains observed values as high as four compared to about high predicted value of about 0.8.

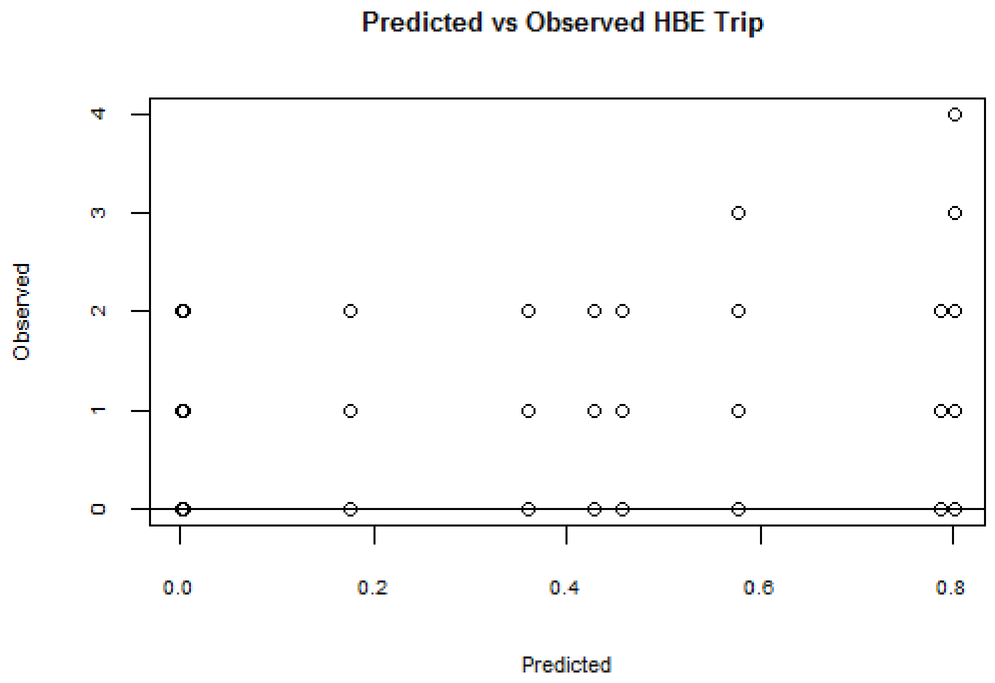


Fig. 6.15 Predicted vs observed HBE trips per person

6.2.3 Home-based Shopping Trips

Home-based shopping trip production rates per person were estimated through cross-classification of individuals into ten groups, stratified by age (five groups) and employment status (worker and non-worker). The results of the cross-classification are presented in Tab. 6.13.

Tab. 6.13 HBS trips per person

Employment Status	**Age Groups**				
	Ag<18	Ag18_24	Ag25_44	Ag45_59	Ag>59
Worker	0.074 (143)	0.166 (993)	0.230 (4998)	0.265 (7035)	0.255 (862)
Non-worker	0.073 (5227)	0.191 (1169)	0.410 (1102)	0.511 (1688)	0.566 (7663)
Number of observations= R^2=0.113, RMSE=0.505, NRMSE=94.2					

Note: The figures in the parentheses indicate the number of individuals in each category.

The tabular results are presented graphically in Fig. 6.16. As we can see in the graph, HBS trips per person varied from 0.074 for employed individuals below age 18 to 0.566 for unemployed individuals above 59 years. The trip rate remained almost similar until the age of 18 to 24 for both cases. However, as the age increases beyond 24 the rate of increase is significant for unemployed individuals compared to employed individuals. For both cases, the increment appears to be approximately linear. The results look logical because adults, particularly unemployed ones are likely to make more shopping trips. RMSE and NRMSE values of the cross-classification was 0.505 and 94.2 respectively.

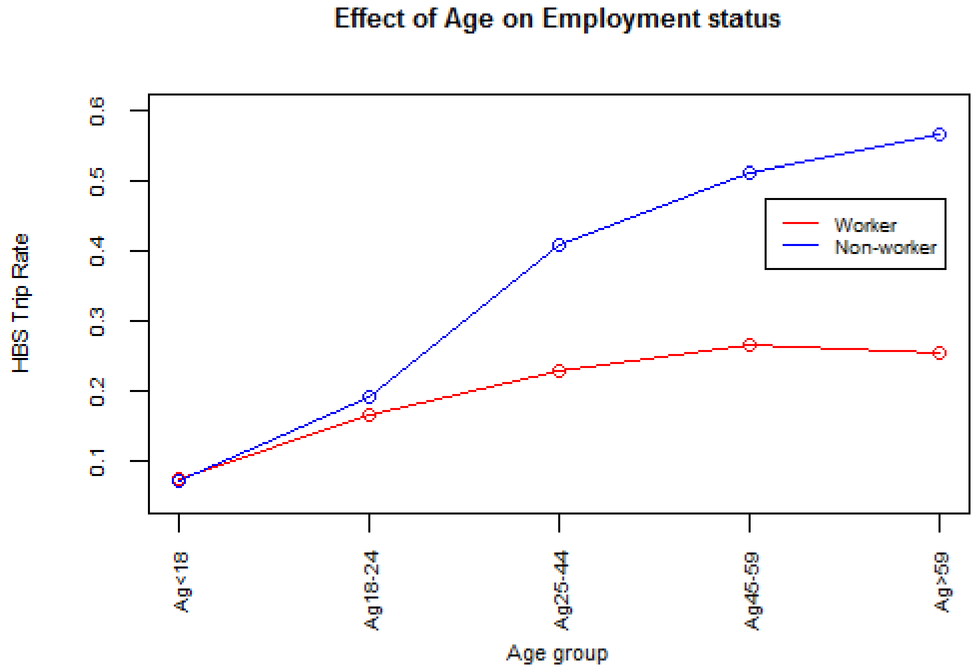


Fig. 6.16 Effect of age on employment status (HBS trips)

The plot of predicted versus observed HBS trips per person is presented in Fig. 6.17**.** It is apparent from the figure that observed values are as high as four compared to far less predicted value of about 0.6.

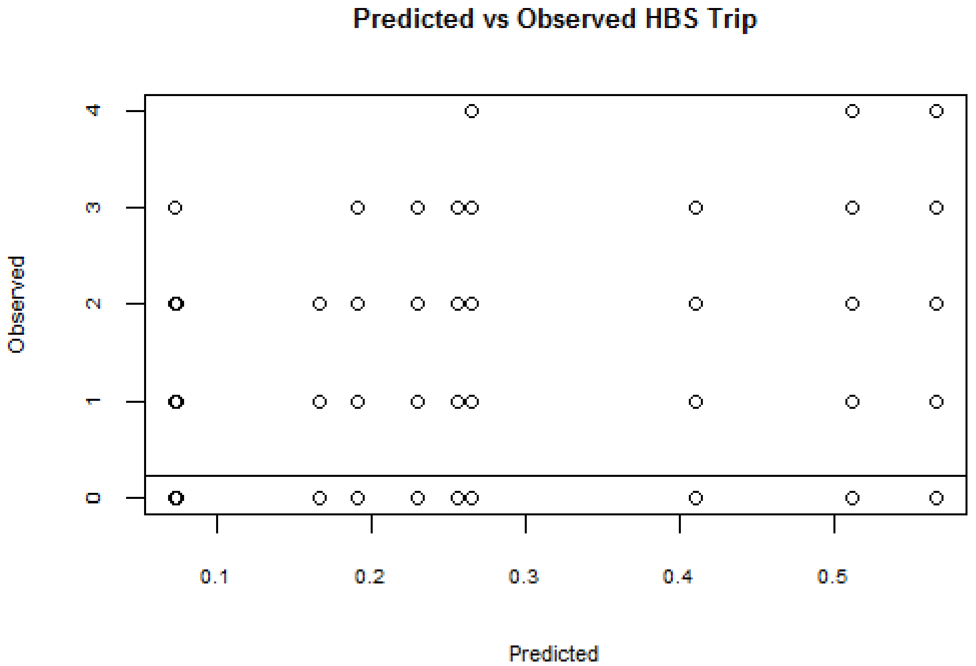


Fig. 6.17 Predicted vs observed HBS trips per person

6.2.4 Home-based Other Trips

Individuals were cross-classified by age (five groups) and employment status into ten categories to estimate home-based other (HBO) trip production rates per person. The results of cross-classification are shown in Tab. 6.14 below.

Tab. 6.14 HBO trips per person

Employment Status	**Age Groups**				
	Ag<18	Ag18_24	Ag25_44	Ag45_59	Ag>59
Worker	0.578 (143)	0.574 (1169)	0.661 (4998)	0.602 (7035)	0.674 (862)
Non-worker	0.788 (5227)	0.719 (1169)	1.130 (1102)	1.177 (1688)	1.094 (7663)
Number of observations=31,249, R^2=0.056, RMSE=0.874, NRMSE=97.2					

Note: The figures in the parentheses indicate the number of individuals in each category.

HBO trip production rates per person varied from low value of 0.580 for employed individuals of age less than 18 to 1.171 for unemployed individuals of age 45 to 59. The results are graphically illustrated in Fig. 6.18. It is apparent from the results that unemployed individuals tend to make more HBO trips. Among unemployed ones, trip rates seems to increase with the age of the individuals. However, the increment does not look linear. Individuals of age 45 to 59 made the most HBO trips. In case of employed individuals there is no significant increase of HBO trip rates as the age increase. RMSE and NRMSE value of the cross-classification was found to be 0.875 and 97.2% respectively.

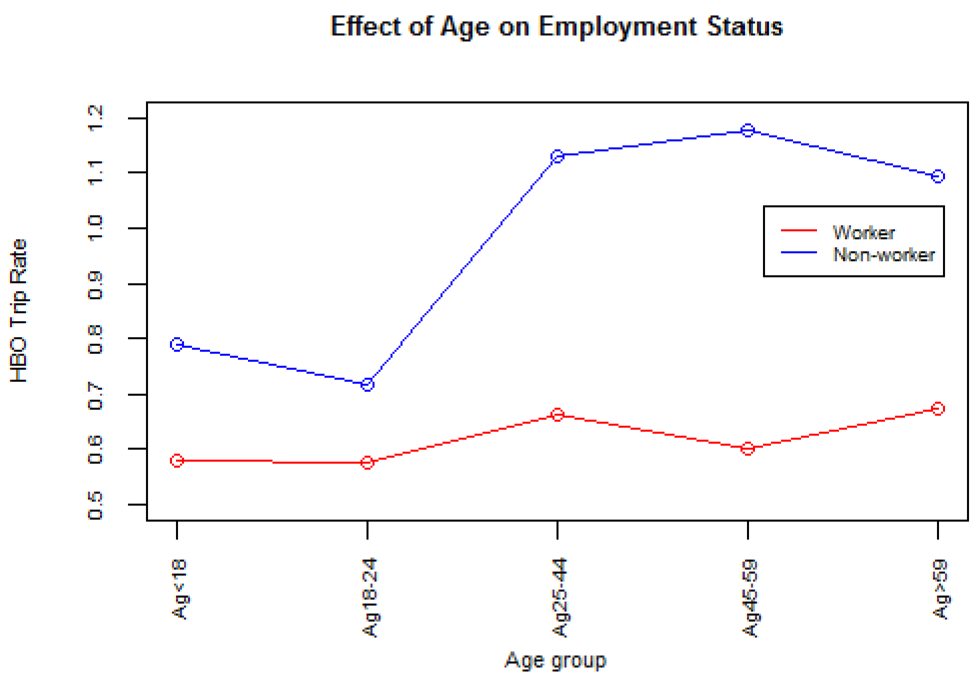


Fig. 6.18 Effect of age on employment status (HBO trips)

Fig. 6.19 presents the plot of predicted vs observed HBO trips per person. The graph contains observed values as high as about eight and predicted value as high as about 1.2.

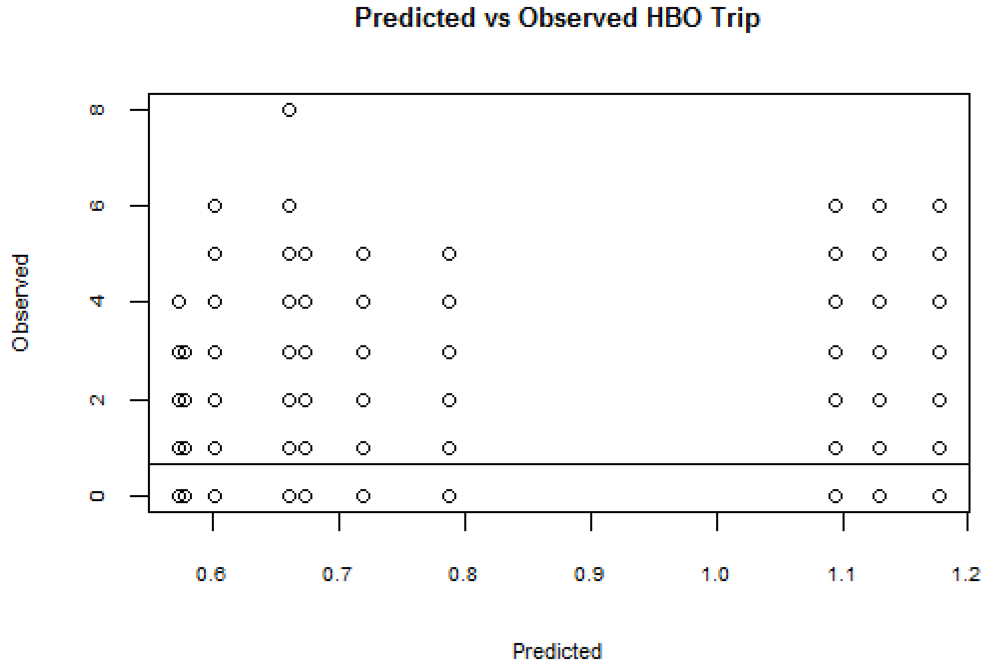


Fig. 6.19 Predicted vs observed HBO trips per person

6.2.5 Non-home-based Trips

Cross-classification of non-home-based (NHB) trip production rates per person was performed by stratifying individuals on the basis of age (six groups) and sex of the individuals into twelve groups. The results from the cross-classification are depicted in Tab. 6.15.

Tab. 6.15 NHB trips per person

Sex	Age Groups					
	Ag<10	Ag10_17	Ag18_24	Ag25_44	Ag45_59	Ag>59
Male	0.647 (1103)	0.375 (1740)	0.537 (1347)	0.706 (2857)	0.563 (4113)	0.622 (4406)
Female	0.628 (1037)	0.400 (1609)	0.635 (1194)	0.878 (3249)	0.665 (4615)	0.616 (4149)
Number of observations=31,419, R^2=0.011, RMSE=1.046, NRMSE=99.5						

Note: The figures in the parentheses indicate the number of individuals in each category.

Graphically, the results of the cross-classification are presented in Fig. 6.20. As shown in the figure, NHB trip production rates is almost identical at age below 18 for both male and female. The rates declined sharply as age ascended. However, as the age of the individuals' increase, the disparity between males and females also increase until the age 25 to 44, and start decreasing from there and finally seems to share virtually the same rate at age over 59. Although, the overall variation of the trips is irregular, the trip rates from age 18 to 24 to age 25 to 44 appears to increase linearly for both sex. The higher NHB trip production rates for individuals below ten could be because of the higher tendency of younger kids accompanying their parents. RMSE and NRMSE value of the cross-classification were 1.046 and 99.5% respectively.

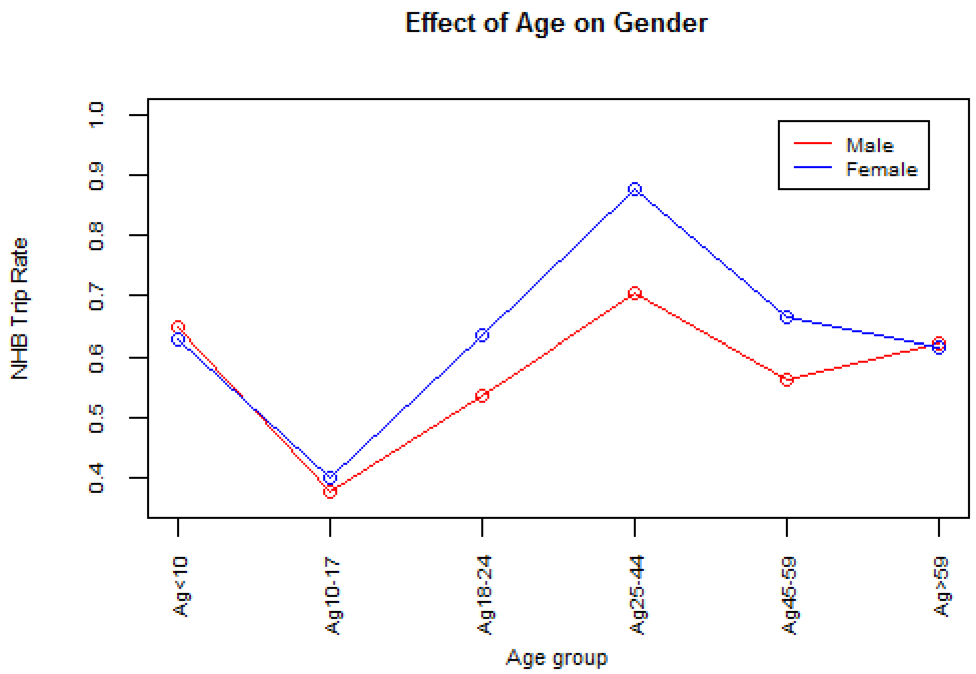


Fig. 6.20 Effect of age on sex (NHB trips)

Fig. 6.21 shows the plot of predicted versus and observed NHB trips per person. The highest observed value is about ten while highest predicted values are just about 0.9.

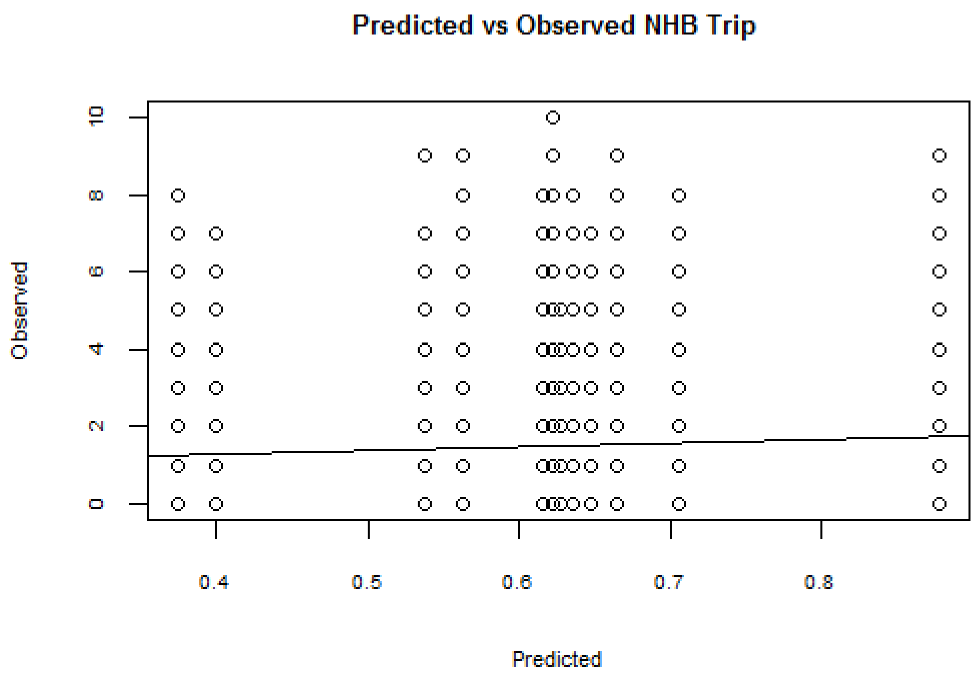


Fig. 6.21 Predicted vs observed NHB trips per person

6.2.6 Summary

In person-category cross-classification of trips, age variable was present in every model. In the same way, another most repeated variable was employment status (three times). Further, age and student status were used only once. In summary, person-category trip production by cross-classification results shows that HBE trip production model was the best performing model with least error. As shown by RMSE values, error varied from 0.257 for HBE trip to 1.046 for NHB trip. Higher errors were seen for trip categories with higher person trip rates.

6.3 Comparison of Household-Category and Person-Category Trip Production Models

The performance comparison of the trip production models at two disaggregate levels were conducted based on the coefficient of determination R^2 and two model validation measures RMSE and NRMSE. Tab. 6.16 summarises the statistical values obtained for different trip purposes at both household and personal level of trip production analysis using multiple regression and cross-classification.

As can be observed, HBW and HBE trip production models have a fairly good performance at both household and personal level. R^2 value of person-category HBW trip production model is 0.450 compared to the marginally higher value of 0.483 yielded at the household level. In the same way, person-category HBE trip production model yielded R^2 value of 0.532, which is again marginally lower than that of 0.579 produced at the household level. HBS trip production models' prediction performance was found to be poor for both levels. Likewise, the only person-category trip production model faring well compared to household-category in terms of R^2 value, was in the case of HBS trips, as persona-category model yielded R^2 value of 0.091 while household-category yielded 0.052. HBO trip production model at household level yielded 0.259 which is much higher than at personal level which yielded mere 0.061. Lastly, in the case of NHB trips, again household-category model with the R^2 value of 0.08 outperformed poorly fit the person-category model that yielded just 0.016 R^2 value. This trend appears to be followed by trip production models by cross-classification technique as well, as it is apparent from the R^2 value comparison that besides HBS trip production model, all other models performed better at a household level compared to personal level.

In contrast to the comparison based on R^2 values, the investigation of RMSE and NRMSE produced different results. Among household-category models, RMSE obtained from HBE trip production model was the lowest followed by HBW trips, HBS trips, HBO trips and NHB trips. Likewise, person-category models also appear to follow the same order. Further, it is apparent from Tab. 6.16 that all five person-category trip production models by both multiple regression and cross-classification produced lower RMSE than that of the household. RMSE value shows that household trip rates were off by as low as 0.592 trips in HBE trips to as high as 1.922 trips in NHB trips. A Higher value is likely to have a higher error as can be observed, the trip categories with higher trip rates had a higher error. On the other hand, error in case of person-category person trip rates was off by the low value of 0.254 trips again in HBE trips and high value of 1.055 trips in NHB trips. As household trip rate is obviously higher than personal trip rate, errors were high in case of household trips. However, NRMSE which is normalised at the standard deviation of observed trip rate gave a different picture. It defied the trend shown by

RMSE, as person based models managed to yield lower NRMSE only in case of HBE and HBS trips. Another important indication of the results is that RMSE and NRMSE values yielded by multiple regression and cross-classification appears to have very small differences.

Finally, only in case of HBS, person-category trip production models proved to be better on all three measures, against no models in case of household-category. Thus, with these evidence, it would be inappropriate to conclude the superiority of either of the two household or person-category trip production models.

Tab. 6.16 Comparison of household-category and person-category trip production models

Trip Categories	Unit of Analysis	Multiple Regression			Cross-classification		
		R^2	RMSE	NRMSE	R^2	RMSE	NRMSE
HBW	Household	0.483	0.663	74.1	0.451	0.659	74.1
	Personal	0.450	0.388	74.6	0.444	0.388	74.6
HBE	Household	0.579	0.492	66.0	0.507	0.524	70.2
	Personal	0.532	0.254	65.8	0.186	0.257	66.6
HBS	Household	0.052	0.854	98.0	0.041	0.853	97.9
	Personal	0.078	0.512	95.5	0.113	0.505	94.2
HBO	Household	0.259	1.544	87.8	0.179	1.594	90.6
	Personal	0.061	0.872	97.1	0.056	0.874	97.2
NHB	Household	0.0834	1.917	95.9	0.056	1.922	97.2
	Personal	0.0157	1.044	99.4	0.011	1.046	99.5

7 Conclusion

The thesis performed the comparison of the predictive performance of person-category and household-category trip production models. MiD 2008, national household travel survey data from Germany conducted in 2008 was used in the analysis. It is the most recent travel data from Germany which gives a detailed account of the social demography of individuals and households and daily travel behaviour of residents of Germany. Trip production models were developed for five different trip categories using multiple regression and cross-classification at both household and personal level. Only, weekdays data was considered in the analysis.

7.1 Achieving the Research Goals

The first objective of the thesis was to examine the predictive performance of trip production models with household and personal unit of analysis and determine which one best describes the travel behaviour.

Firstly, trip production models of all five trip categories were scrutinised at household and personal level by multiple regression. At the household level, socio-economic and demographic attributes of the households were included as independent variables while both personal and household attributes were included at personal level. However, no household attributes were found to impact person-category models. Finally, coefficient of determination R^2 value and model validation measures, RMSE and NRMSE yielded by the models at the household and personal level were evaluated and compared. At both levels, HBW and HBE performed well as they managed to explain more than 45% of the total variation in the model. However, household–category models performed marginally better in both cases. Other three models were poorly fit. Besides, HBS trip production model, remaining two also yielded higher R^2 value at the household level. On the other hand, RMSE values yielded by all five models were found to be lower at a personal level. Contrary to this, a comparison based on NRMSE values, almost followed the trend shown by R^2 value as only HBS and HBE yielded lesser NRMSE values at a personal level.

Secondly, two-way cross-classification was employed at both levels. In the case of household-category trip production models, along with conventionally used household attributes such as household size, household income, car ownership and the number of workers in the households, other variables such as unemployed and children in the household were also tested. Best cross-classification for a trip category was chosen based on R^2, RMSE, and NRMSE values. On the other hand, person-category models were cross-classified by their personal attributes. In the two-way cross-classification of person-category models, age variable remained constant in all cases while other variable varied from sex, employment status, student status and license ownership status across the models. Surprisingly, comparison of models based on R^2, RMSE and NRMSE exhibited the same trend that multiple regression showed.

Another important indication from the comparison of RMSE revealed that household trip rates were off by higher values compared to person trip rates. Household trip rates were off by the low value of 0.592 at HBE trips to the high value of 1.922 for NHB trips while person trip rates were off by the low value of 0.254 again for HBE trips and high value of 1.055 for NHB trips.

This is not surprising because household trip rates are higher compared to person trip rates and thus resulting in a higher error.

It can be observed that both techniques produced the same trend. To sum up, it can be stated from the comparison of statistical measures that the research has produced no concrete evidence in backing trip production model at any of the two units of analysis. On the other hand, performance of person-category models performed similar to household-category models, relevance of person-category models cannot be dismissed in further research.

The second objective of the thesis was to determine the respective benefits of use of multiple regression and cross-classification approaches in trip production models

Some notable findings were deducted from the use of multiple regression and cross-classification in the analysis. Both models were equally good at estimating trip rates as RMSE and NRMSE values suggested. However, both approaches had positive as well as negative aspects. The ability to use more and more independent variables in the regression analysis was one of the main benefits over cross-classification. On the other hand, use of a higher number of variables in cross-classification results in higher categories and the higher sample size is required to fill the categories. In the thesis, even for categories generated by two-way cross-classification (particularly in household-category), several cells were found short of data. Unlike cross-classification, regression is open to variable experiments such as logarithm of variables, a product of two or more variables. Use of logarithm of household size was tried in the thesis and was found to be significant in one case. Comparatively, person-category regression models required more time and effort. Since most of the independent variables affecting personal trip rates were found to be dummy variables, correlation with the dependent variables had to be tested individually. Further, stepwise regression did not always deliver relevant variables.

Due to the limitation of variables being used in cross-classification (particularly in household-category), the variables like household size, the number of employed members appear to overshadow the effect of other cars and household income. Cars and household income did not deliver consistent results in the thesis. On the other hand, despite using two-way cross-classification, the error yielded by these models was similar in magnitude to that of regression which can be considered as a benefit. In contrast to household-category cross-classification, person-category cross-classification can incorporate many variables. Although, only two-way cross-classification was used in the thesis, due to the binary nature of behavioural variables (such as employment status, sex, license ownership) multiple variables can be used in person-category models. Yet, it can have fewer categories than household-category models.

Regression and cross-classification are different approaches, but both approaches produced comparable results. Thus, since, cross-classification method is simpler and easier to understand than regression models, cross-classification can be recommended for planning. Moreover, possibility to use many variables and requirement of comparatively lesser data, in the case of person-category cross-classification makes it a promising approach. Further studies in improving person-category cross-classification is recommended.

7.2 Limitations of the study:

Some of the noticeable limitations of the study are listed below:

1. Since most recent data was not available, the thesis had to rely on the travel survey data collected in 2008.
2. For simplification, two-way cross-classification was carried out which limited the model with just two variables unlike multiple regression, where multiple variables were involved.
3. Non-home-based trips which constituted about 30% was not broken down into further trip categories. Aggregating all trip purposes on same categories might have resulted in poor regression results.
4. Since monthly household income was available in the range, it was decided to consider to use it as a dummy variable. Had there been definite household income, it could have given the extra option to test in improving the models
5. Effect of policy measures such as parking supply, parking costs, congestion charge was not assessed.
6. Many households and persons had to be removed due to incomplete information.

Above limitations need to be addressed in further research for improving the results.

List of References

Allaman, P. M., et al. (1982). New Approaches to Understanding Travel Behavior, Transportation Research Board, National Research Council.

Badoe, D. A. (2007). "Forecasting Travel Demand with Alternatively Structured Models of Trip Frequency." Transportation Planning and Technology **30**(5): 455-475.

Badoe, D. A. and C.-C. Chen (2004). "Unit of analysis in conventional trip generation modelling: an investigation." Canadian Journal of Civil Engineering **31**(2): 272-280.

Badoe, D. A. and J. L. Mwakalonge (2011). "Estimating Household Trip Rates for Cross-Classification Cells with No Data: Alternative Methods and Their Performance in Prediction of Travel." Journal of Urban Planning and Development **137**(3): 262-271.

Badoe, D. A. and G. N. Steuart (1997). "Urban and travel changes in the greater Toronto area and the transferability of trip-generation models." Transportation Planning and Technology **20**(4): 267-290.

Ben-Akiva, M., et al. (2007). "Towards disaggregate dynamic travel forecasting models." Tsinghua Science and Technology **12**(2): 115-130.

Berechman, J. and R. E. Paaswell (1977). "The impact of car availability on urban transportation behavior." Transportation **6**(2): 121-134.

Boarnet, M. G. and S. Sarmiento (1998). "Can Land-use Policy Really Affect Travel Behaviour? A Study of the Link between Non-work Travel and Land-use Characteristics." Urban Studies **35**(7): 1155-1169.

Bruton, M. J. (1985). Introduction to Transportation Planning, Hutchinson.

Bureau of Public Roads (1965). Traffic Assignment and Distribution for Small Urban Areas. U.S.

Chang, J. S., et al. (2014). "Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area." Transportation Letters **6**(2): 78-88.

Chatterjee, S. and A. S. Hadi (2006). Introduction. Regression Analysis by Example, John Wiley & Sons, Inc.**:** 1-19.

Davies, R. L. (1969). "EFFECTS OF CONSUMER INCOME DIFFERENCES ON SHOPPING MOVEMENT BEHAVIOR." Tijdschrift voor economische en sociale geografie **60**(2): 111-121.

Douglas, A. (1973). "Home-based trip end models—a comparison between category analysis and regression analysis procedures." Transportation **2**(1): 53-70.

Douglas, A. A. and R. J. Lewis (1971). Trip Generation Techniques, Traffic Engineering & Control.

Downes, J. D., et al. (1978). Household and person trip generation models. Crowthorne, Berkshire, Access and Mobility Division, Transport Operations Department, Transport and Road Research Laboratory.

Draper, N. R. and H. Smith (1966). Applied Regresion Analisis, John Wiley & Sons.

Farrar, D. E. and R. R. Glauber (1967). "Multicollinearity in regression analysis: the problem revisited." The Review of Economic and Statistics: 92-107.

German Aerospace Center (DLR) (2012). "Mobility in Germany 2008." Retrieved May 6, 2017, from http://daten.clearingstelle-verkehr.de/223/.

Golding, S. (1972). A CATEGORY ANALYSIS APPROACH TO TRIP GENERATION*. Australian Road Research Board (ARRB) Conference, 6th. Canberra

Goulias, K. G., et al. (1990). "Practical method for the estimation of trip generation and trip chaining." University of California Transportation Center.

Hanson, S. (1982). "THE DETERMINANTS OF DAILY TRAVEL-ACTIVITY PATTERNS: RELATIVE LOCATION AND SOCIODEMOGRAPHIC FACTORS." Urban Geography **3**(3): 179-202.

Hanson, S. and P. Hanson (1980). "Gender and urban activity patterns in Uppsala, Sweden." Geographical Review: 291-299.

Heggie, I. G. and P. M. Jones (1978). "Defining domains for models of travel demand." Transportation **7**(2): 119-125.

Hu, S. (2010). Modelling trip generation/trip accessibility using logit models, Edinburgh Napier University.

infas (Institut für angewandte Sozialwissenschaft) and DLR (Deutsches Zentrum für Luft- und Raumfahrt) (2010). "Mobilität in Deutschland 2008 Ergebnisbericht Struktur – Aufkommen – Emissionen – Trends." Retrieved May 21, 2017, from http://www.mobilitaet-in-deutschland.de/pdf/MiD2008_Abschlussbericht_I.pdf.

Kitamura, R. (1996). Applications of models of activity behavior for activity based demand forecasting. Proceedings of Activity-Based Travel Forecasting Conference.

Kitamura, R. and L. P. Kostyniuk (1986). "Maturing motorization and household travel: The case of nuclear-family households." Transportation Research Part A: General **20**(3): 245-260.

Kockelman, K. (1997). "Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area." Transportation Research Record: Journal of the Transportation Research Board(1607): 116-125.

McCarthy, G. M. (1969). "Multiple-Regression Analysis of Household Trip Generation-A Critique." Highway Research Record(297).

McFadden, D. (1976). The theory and practice of disaggregate demand forecasting for various modes of urban transportation, Dept. of Transportation, Office of the Secretary.

McFadden, D. and F. Reid (1975). "AGGREGATE TRAVEL DEMAND FORECASTING FROM DISAGGREGATED BEHAVIORAL MODELS." Transportation Research Record(534): 24-37.

Moeckel, R., et al. (2017). "From Macro to Microscopic Trip Generation: Representing Heterogeneous Travel Behavior." The Open Transportation Journal **11**(1).

Ortúzar, J. D. and L. G. Willumsen (2011). Modelling Transport.

Paez, A., et al. (2007). "Elderly mobility: demographic and spatial analysis of trip making in the Hamilton CMA, Canada." Urban Studies **44**(1): 123-146.

Pas, E. (1977). "A study of residential trip generation in Cape Town." Civil Engineering= Siviele Ingenieurswese **19**(2): 27-37.

Pas, E. (1978). "An empirical comparison of zonal, household and personal models of home-based trip generation." Traffic Engineering & Control **19**(2).

Pas, E. I. (1974). An analysis of residential trip generation in Cape Town, University of Cape Town.

Pas, E. I. (1986). "Multiday Samples, Parameter Estimation Precision, and Data Collection Costs for Least Squares Regression Trip-Generation Models." Environment and Planning A **18**(1): 73-87.

Pettersson, P. and J.-D. Schmöcker (2010). "Active ageing in developing countries?–trip generation and tour complexity of older people in Metro Manila." Journal of Transport Geography **18**(5): 613-623.

Prevedouros, P. D. and J. L. Schofer (1991). "Trip characteristics and travel patterns of suburban residents." Transportation Research Record(1328).

Rawlings, J. O., et al. (2001). Applied regression analysis: a research tool, Springer Science & Business Media.

Rengaraju, V. R. and M. Satyakumar (1994). "Structuring Category Analysis Using Statistical Technique." Journal of Transportation Engineering **120**(6).

Ross, C. L., et al. (1995). "Analysis of travel behavior using three-parameter data loggers." Journal of Transportation Engineering **121**(4): 338-344.

Scheiner, J., et al. (2011). "Gendered activity spaces: trends over three decades in Germany." Erdkunde: 371-387.

Shuldiner, P. W. and N. U. T. Institute (1962). Trip Generation and the Home, Northwestern University, Technological Institute.

Silvey, S. D. (1969). "Multicollinearity and Imprecise Estimation." Journal of the Royal Statistical Society. Series B (Methodological) **31**(3): 539-552.

Supernak, J. (1979). A behavioural approach to trip generation modelling. Proceedings 7th PTRC Summer Annual Meeting, University of Warwick, England.

Supernak, J., et al. (1983). Person-category trip-generation model.

Takyi, I. K. (1990). "Trip generation analysis in a developing country context." Transportation Research Record **1285**: 9-21.

The R Foundation (2017). "What is R?". Retrieved May 7, 2017, from https://www.r-project.org/about.html.

United States Federal Highway Administration (1975). Trip generation analysis, Dept. of Transportation, Federal Highway Administration [Office of Highway Planning], Urban Planning Division.

Wooton, H. J. and G. W. Pick (1967). "A Model for Trips Generated by Households " Journal of Transport, Economics and Policy **1**: 137-153.

List of Abbreviations

HBW: Home-Based Work

HBE: Home-Based Education

HBS: Home-Based Shopping

HBO: Home-Based Other

HBPB: Home-Based Personal Business

HBSR: Home-Based Social and Recreational

THB: Total Home-Based

NHB: Non-Home-Based

RMSE: Root Mean Square Error

NRMSE: Normalized Root Mean Square Error

List of Figures

List of Tables

Appendix A: Summary of Previous Studies

Study	(Goulias, Pendyala et al. 1990) Practical Method for the Estimation of Trip Generation and Trip Chaining	(Prevedouros and Schofer 1991) Trip Characteristics and Travel Patterns of Suburban Residents	(Badoe and Steuart 1997) Urban and Travel Changes in the Greater Toronto Area and the transferability of trip generation models	(Hu 2010) Modeling trip generation/trip accessibility using logit models	(Wooton and Pick 1967) A Model for Trips Generated by Households	(Supernak, Talvitie et al. 1983) Person-Category Trip-Generation Model
Place	USA	USA	Canada	UK	UK	USA
Level	Household	Person	Household	Household and Person	Household	Person and Household
Trip categories	Work, School, shopping, social, personal business and serve-passenger	Total, work and non-work	THB, HBW, HBS, HBSR, HBPB	Work (household) Shopping (Person)	Total	Work, education, Business, shopping, social recreational and other (HBO, HBD and NHB)
Independent Variables	Household demographics, household life cycle stage, household head characteristics, household income, residence county and area type, car availability and car ownership	Employment status, gender, resident location, car availability, transit, age,	Household size, Number of workers, Number of licensed person, number of vehicles	(Household) Worker (full-time, part-time), company car, cars owned, income, child (Person) Car, age, social status, location, gender, parking cost	Income, car, household size	(Person) Age, employment status, gender, race, income, family type, car availability (Household) Household size, car, workers
Modelling Technique	Multiple regression	Multiple regression	Multiple regression	Multiple regression	Cross-classification	Cross-classification
R^2	0.3125 (Work) 0.5453 (Education) 0.0787 (Shopping 0.0869 (Social) 0.114 (Personal business) 0.050 (Serve-passenger)	0.08 (Total) 0.28 (Non-work) 0.33 (Work)	0.348(THB) 0.584 (HBW) 0.045 (HBS) 0.0345 (HBSR) 0.025 (HBPB)	0.326 (Work) 0.135 (Shopping)	-	-

Appendix B: Data Sample

hhid	hwtrip	hetrip	hstrip	hotrip	nhbtrip	wt	HSIZE	INK	LICN	CARS	CHILD	FTIME	PTIME	TRAINEE	UNEMP	SCHstud	UNIstud	Wptime	CHILD6_17	CHILD5
200814	1	1	0	0	1	0.75191286	3	high	2	1	1	2	0	1	0	0	0	0	1	0
200816	4	1	3	1	0	1.03889636	4	high	3	2	1	1	1	1	1	1	0	1	1	0
200828	1	1	1	2	4	0.89642155	3	high	2	2	1	1	1	0	1	1	0	1	1	0
200856	0	0	1	1	1	0.97907729	2	NA	2	1	0	0	0	0	2	0	0	0	0	0
200870	3	0	0	2	1	0.36060390	2	low	2	0	0	0	1	0	1	0	0	0	0	0
200914	2	0	0	2	0	1.24761377	3	high	3	1	0	1	1	0	1	0	1	1	0	0
200916	1	0	0	1	1	2.75536821	1	low	1	1	0	1	0	0	0	0	0	0	0	0
200930	1	0	0	0	0	0.47999723	2	high	2	1	0	1	0	0	1	0	0	0	0	0
200951	0	0	0	2	0	0.34539478	2	low	2	2	0	0	0	0	2	0	0	0	0	0
200954	1	0	0	1	3	2.70363921	1	high	1	2	0	1	0	0	0	0	0	0	0	0
200961	0	0	0	2	2	1.43964751	2	high	2	2	0	0	0	0	2	0	0	0	0	0
200969	0	0	1	0	8	3.98504597	1	high	1	1	0	1	0	0	0	0	0	0	0	0
200985	1	0	0	1	0	0.63549571	2	low	1	0	1	0	1	0	1	1	0	1	1	0
200991	1	3	2	8	3	0.51644142	5	high	2	2	3	0	2	0	3	3	0	1	3	0
200995	2	3	1	1	0	0.63544025	5	high	2	2	2	1	1	0	3	2	0	1	2	0
200998	0	0	1	0	0	6.20610725	1	high	1	0	0	0	0	0	1	0	0	0	0	0
201026	0	0	1	3	0	2.25785122	1	low	1	1	0	1	0	0	0	0	0	0	0	0

Appendix B: Sample of Regression Summary

Household-Category

```
Call:
lm(formula = hwtrip ~ FTIME + PTIME + TRAINEE + CARS, data = my.data,
    weights = wt, na.action = na.exclude)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-3.1679 -0.2262 -0.0449  0.2852  5.0376

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.033757   0.008833   3.822 0.000133 ***
FTIME       0.655252   0.007873  83.229  < 2e-16 ***
PTIME       0.546828   0.011476  47.649  < 2e-16 ***
TRAINEE     0.370947   0.021094  17.586  < 2e-16 ***
CARS        0.022857   0.007297   3.132 0.001738 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5933 on 14628 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.4832,    Adjusted R-squared:  0.4831
F-statistic:  3420 on 4 and 14628 DF,  p-value: < 2.2e-16
```

Person-Category

```
Call:
lm(formula = pwtrip ~ FTIME + PTIME + TRAINEE + Ag18_24 + Ag25_44 +
    Ag45_59 + SEX_, data = my.pdata, weights = wt, na.action = na.exclude)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-2.7045 -0.0491 -0.0073  0.1596  3.9781

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.021610   0.004148   5.210 1.90e-07 ***
FTIME        0.699870   0.006917 101.178  < 2e-16 ***
PTIME        0.535351   0.008553  62.594  < 2e-16 ***
TRAINEE      0.334107   0.014789  22.592  < 2e-16 ***
Ag18_24      0.084351   0.009074   9.296  < 2e-16 ***
Ag25_44      0.041799   0.007270   5.749 9.04e-09 ***
Ag45_59      0.087699   0.007533  11.642  < 2e-16 ***
SEX_FEMALE  -0.015821   0.004629  -3.418 0.000632 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3853 on 31396 degrees of freedom
  (75 observations deleted due to missingness)
Multiple R-squared:  0.4503,    Adjusted R-squared:  0.4501
F-statistic:  3673 on 7 and 31396 DF,  p-value: < 2.2e-16
```

Appendix C: Sample of Cross-Classification Summary

Household-Category (HBW trips)

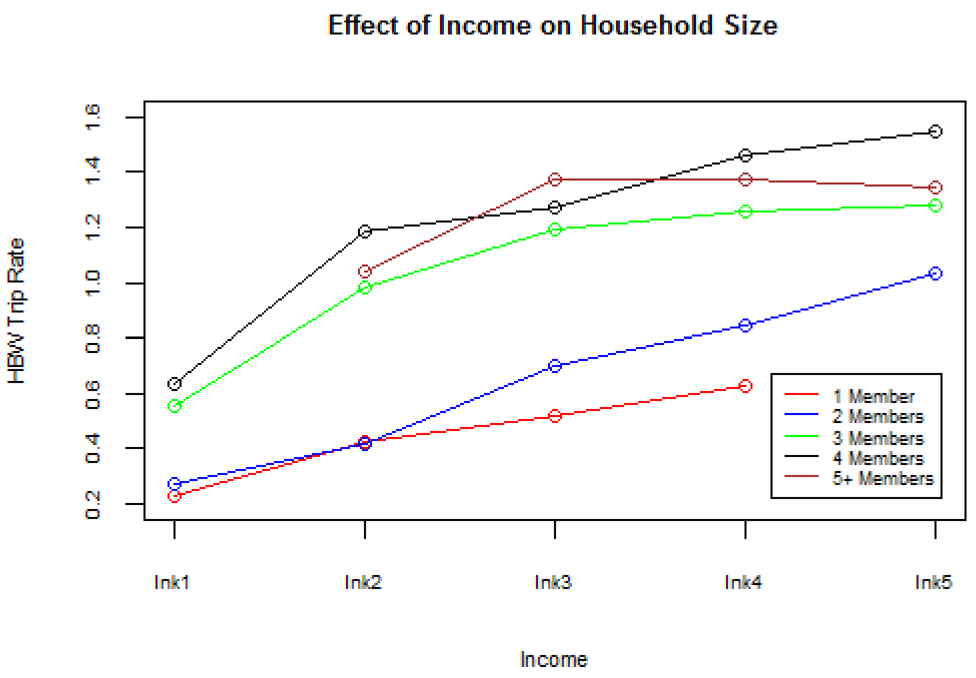
	Income	Employed	Record	W.trips	W.rec	Wtr.rate
1	ink1	Zero	1609	81.077827	2523.92686	0.032
2	ink2	Zero	1957	59.954381	1864.37510	0.032
3	ink3	Zero	564	14.832228	485.25110	0.031
4	ink4	Zero	186	5.009085	145.15147	0.035
5	ink5	Zero	63	1.226442	49.74686	0.025
6	ink1	One	754	795.120501	1154.50334	0.689
7	ink2	One	1532	1334.679106	1817.95002	0.734
8	ink3	One	777	550.746014	733.66020	0.751
9	ink4	One	366	213.282039	293.97888	0.726
10	ink5	One	154	116.611064	155.71134	0.749
11	ink1	Two	110	67.681685	75.02253	0.902
12	ink2	Two	1058	975.810683	767.02823	1.272
13	ink3	Two	1291	1234.470321	953.62643	1.295
14	ink4	Two	855	837.466729	624.98154	1.340
15	ink5	Two	465	514.992433	370.23434	1.391
17	ink12	Three+	159	192.921462	105.38688	1.831
18	ink3	Three+	286	389.867198	197.94596	1.970
19	ink4	Three+	244	342.310279	173.46961	1.973
20	ink5	Three+	147	193.786792	101.11803	1.916

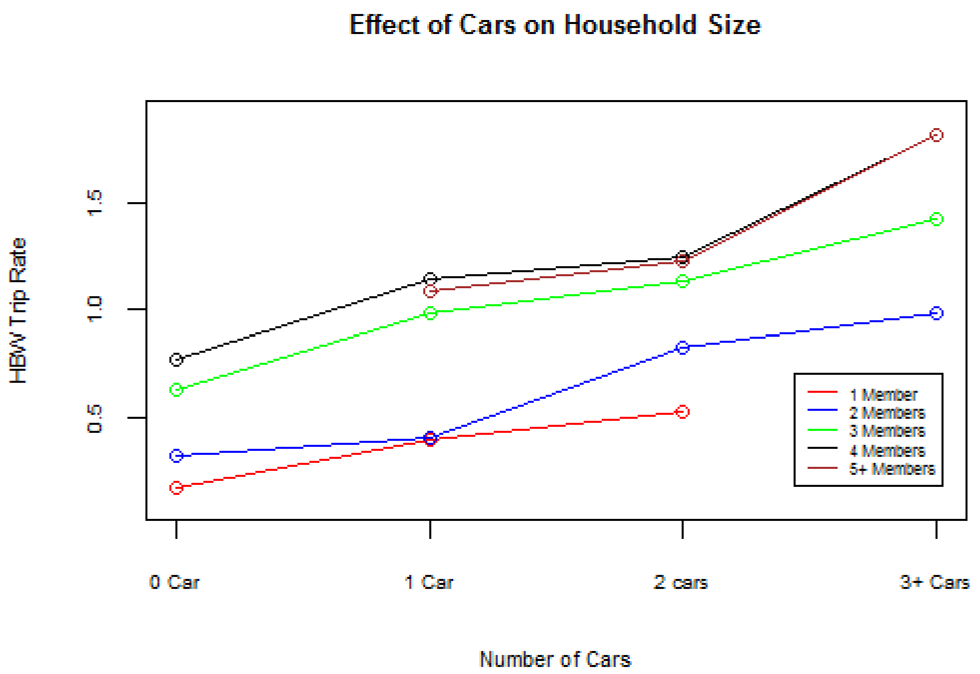
Person-Category (HBW)

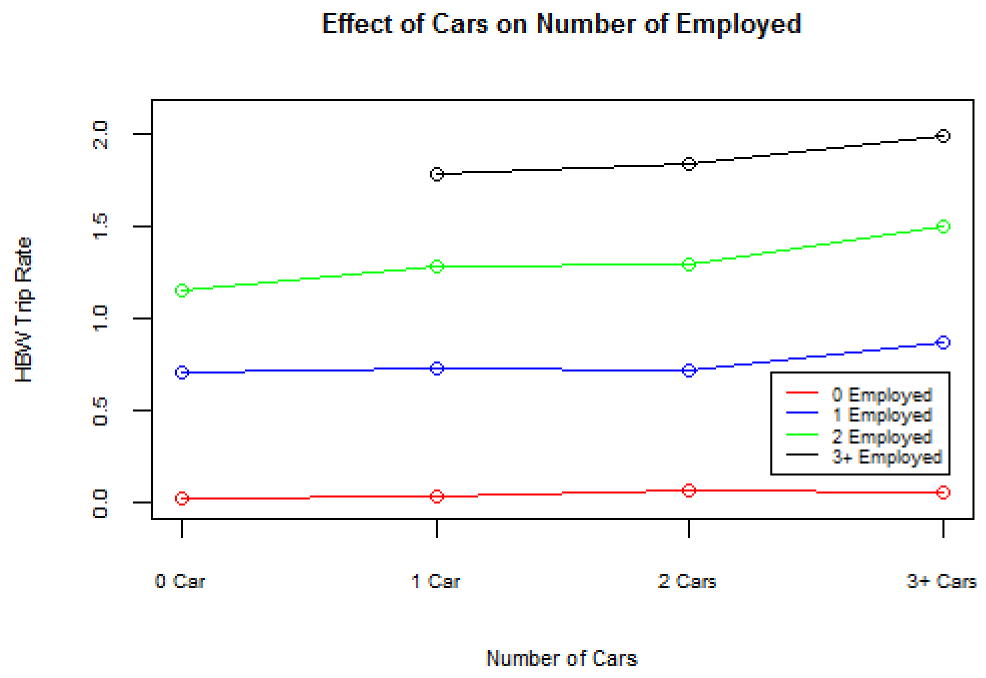
	AgeGr	Emp	Record	Pwtrips	W.rec	Pwtr
2	Ag<18	Nonworker	5298	21.03897	5080.4525	0.004
3	Ag<18	Otherworker	172	53.82475	139.4024	0.386
4	Ag<25	Ftworker	563	507.56985	664.7383	0.764
5	Ag18-24	Nonworker	1263	154.70239	1260.3936	0.123
6	Ag18-24	Otherworker	727	314.74920	695.5121	0.453
7	Ag25-44	Ftworker	3467	4107.62271	5375.8772	0.764
8	Ag25-44	Nonworker	1095	99.42929	1520.9527	0.065
9	Ag25-44	Otherworker	1541	990.10156	1830.1654	0.541
10	Ag45-59	Ftworker	4972	3128.11620	3902.4484	0.802
11	Ag45-59	Nonworker	1691	105.45680	1346.5101	0.078
12	Ag45-59	Otherworker	2064	1016.12180	1566.2062	0.649
13	Ag>59	Ftworker	544	308.85924	430.4314	0.718
14	Ag>59	Nonworker	7692	113.10808	7143.1149	0.016
15	Ag>59	Otherworker	318	167.52347	266.3718	0.629

Appendix D: Household-Category Cross-Classification Graphs

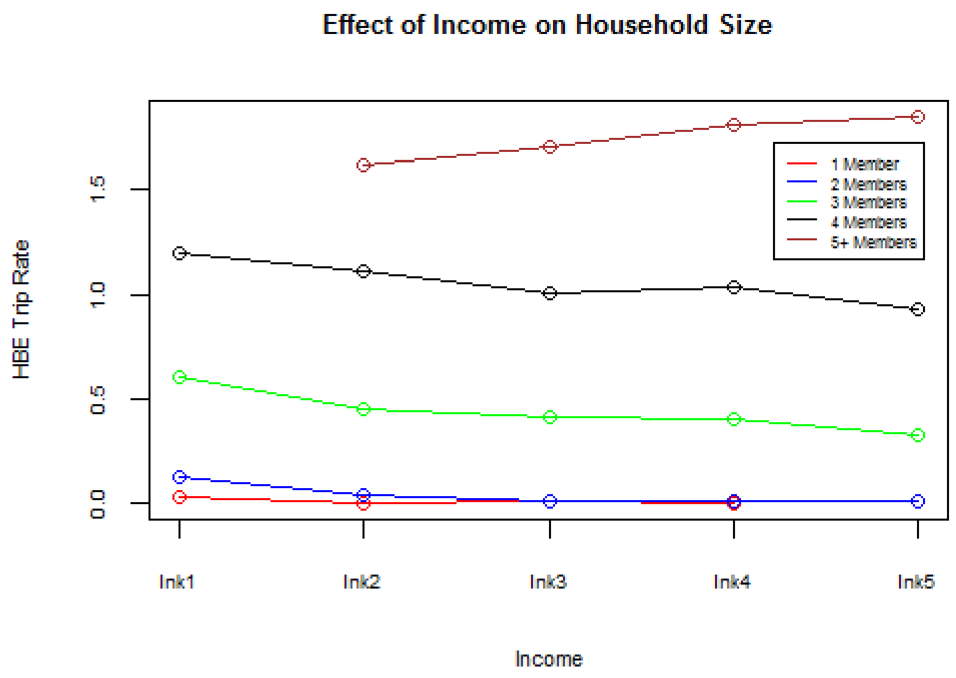
Home-Based Work Trip

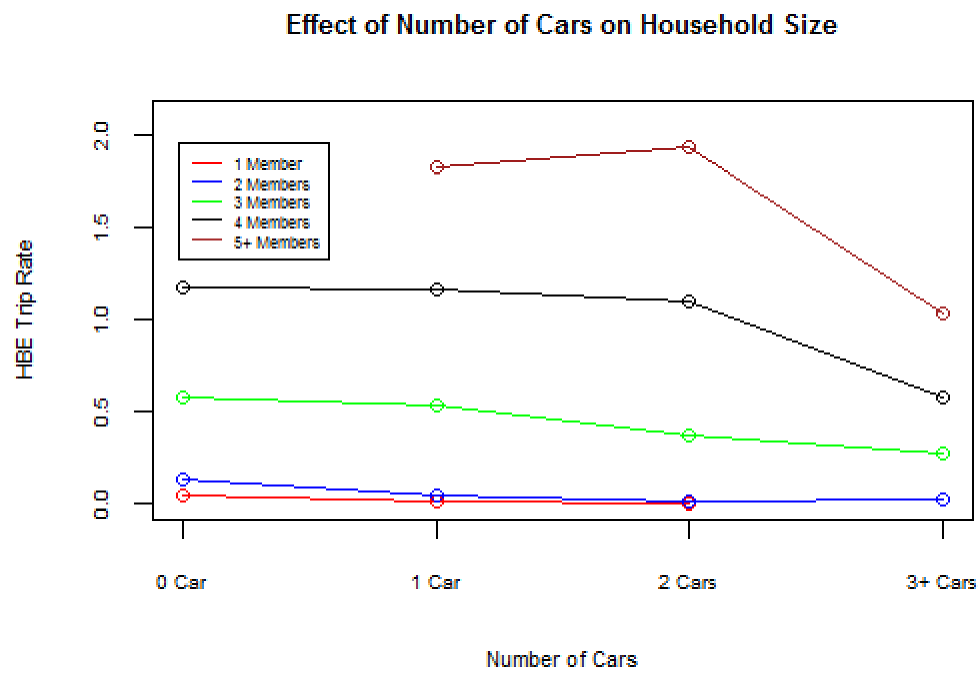


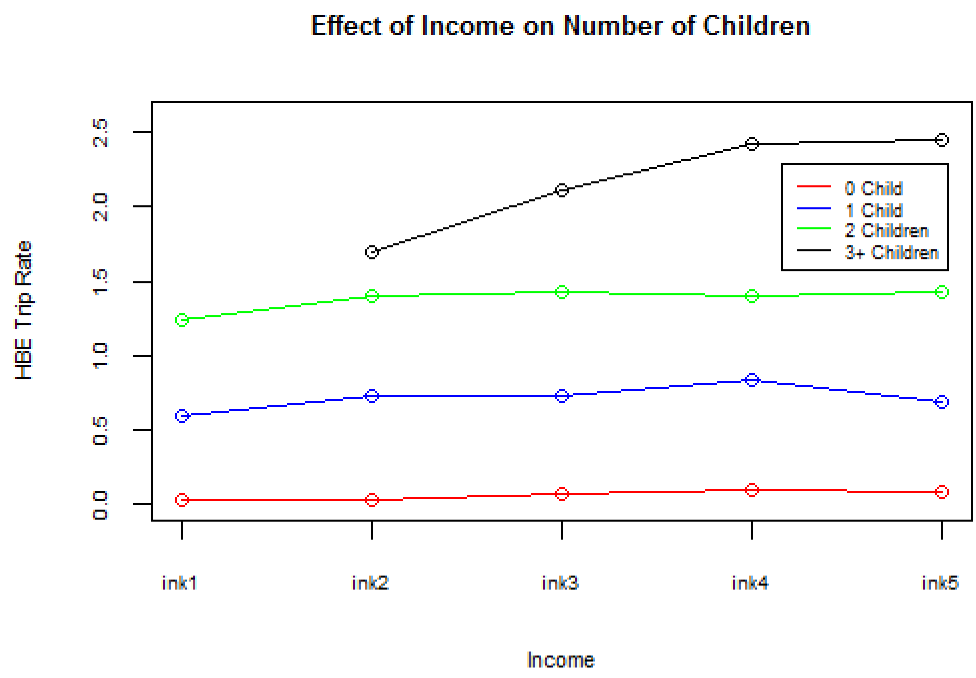




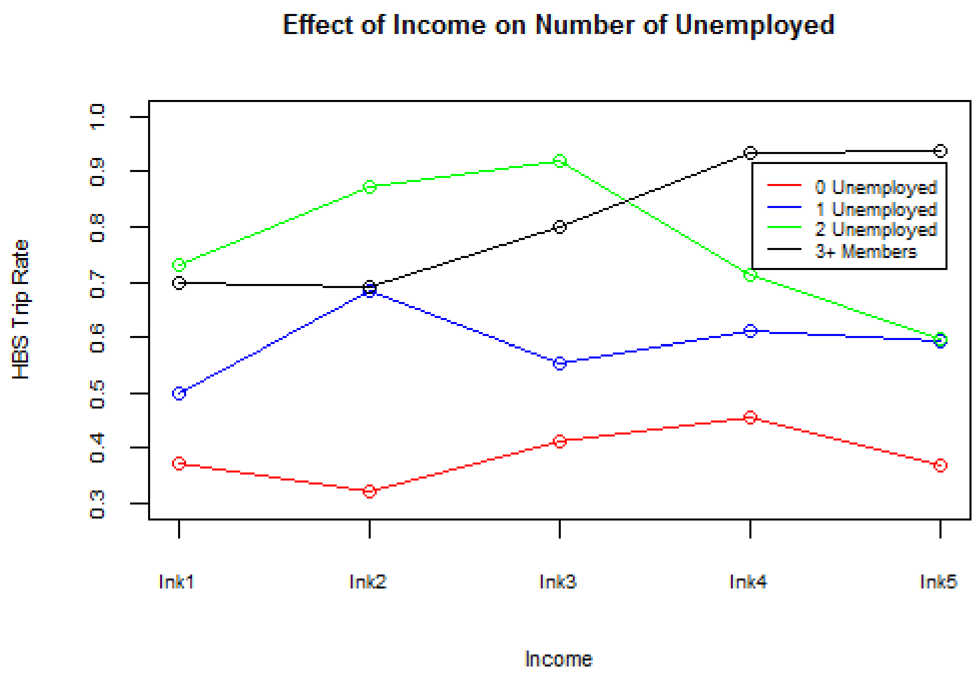
Home-Based Education Trips

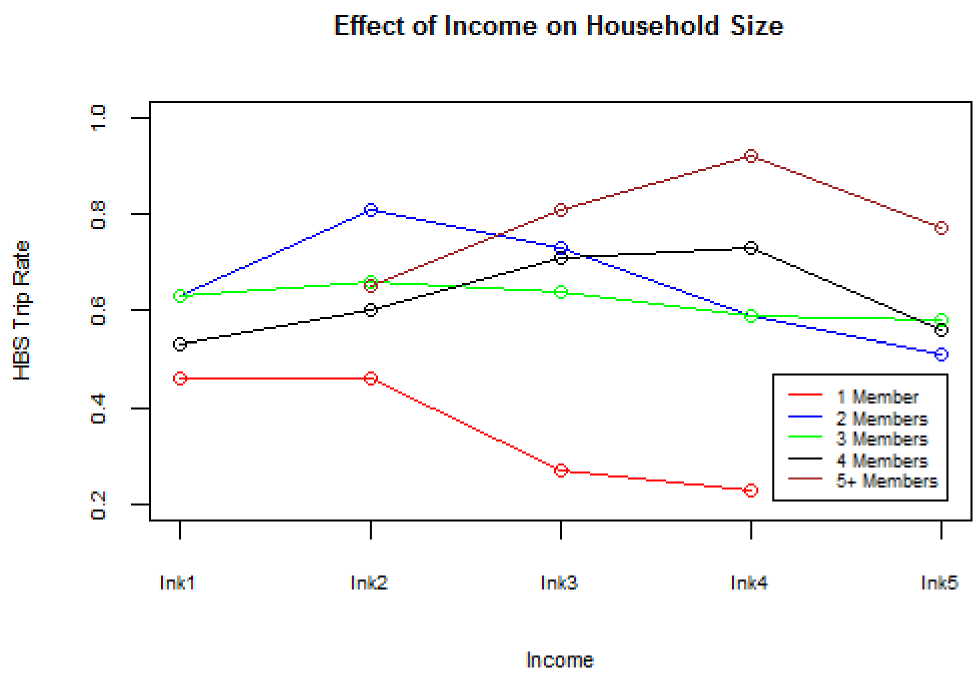


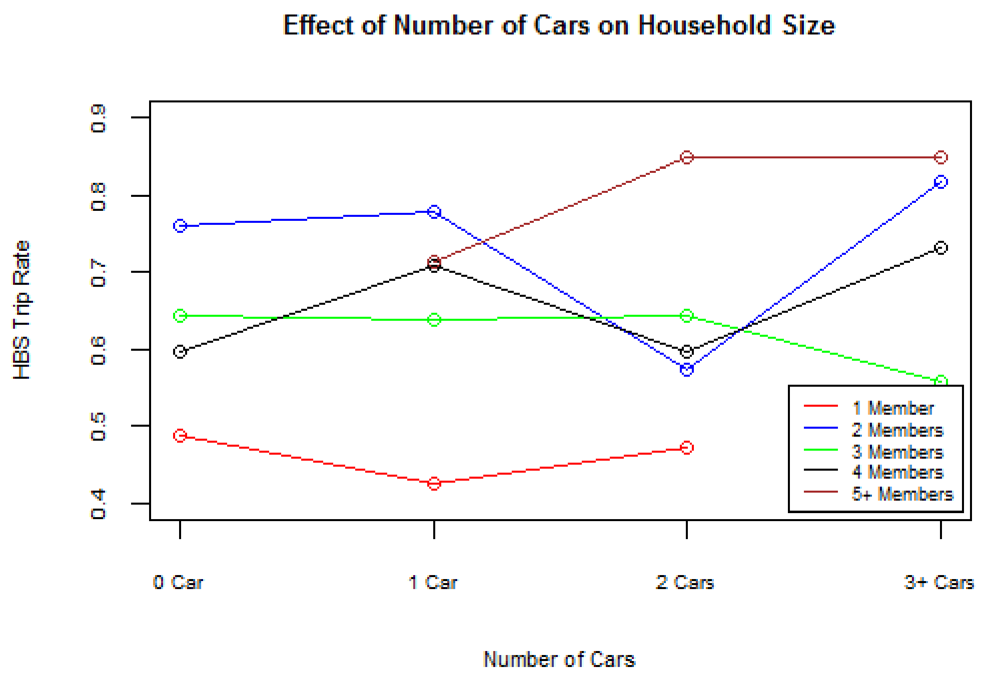

Effect of Number of Cars on Household Size
HBE Trip Rate
0.0
0.5
1.0
1.5
2.0
1 Member
2 Members
3 Members
4 Members
5+ Members
0 Car
1 Car
2 Cars
3+ Cars
Number of Cars


Effect of Income on Number of Children
HBE Trip Rate
0.0
0.5
1.0
1.5
2.0
2.5
0 Child
1 Child
2 Children
3+ Children
ink1
ink2
ink3
ink4
ink5
Income

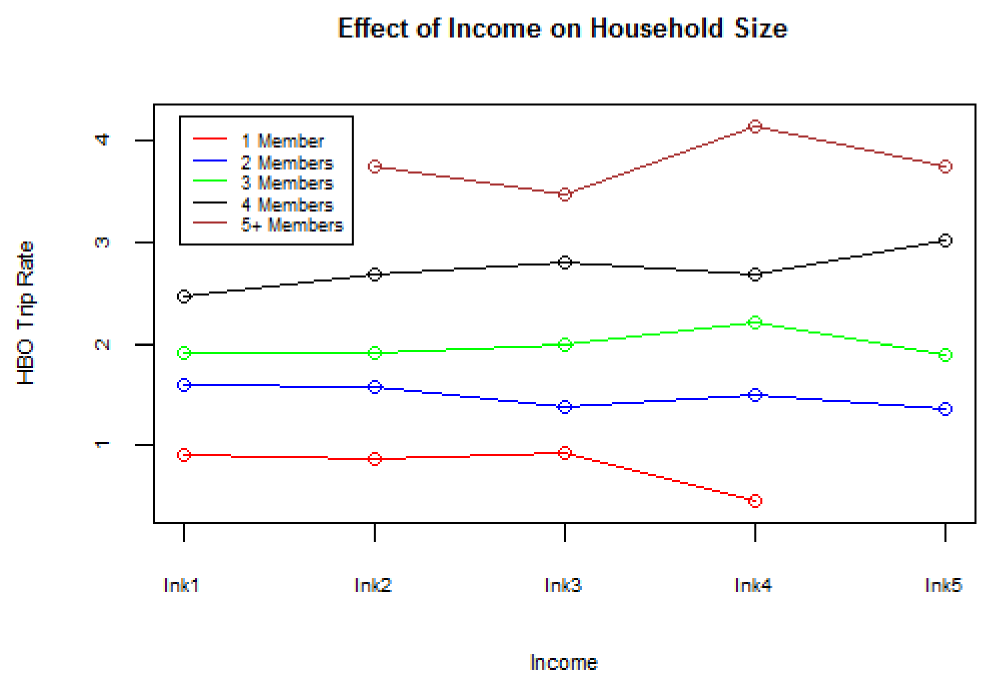
Home-Based Shopping Trips

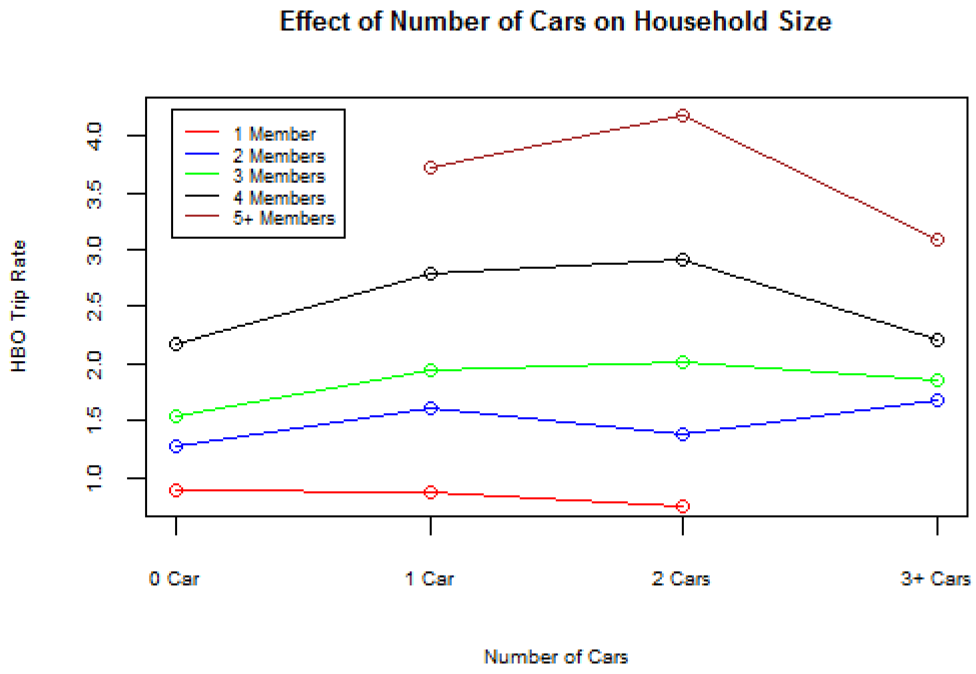


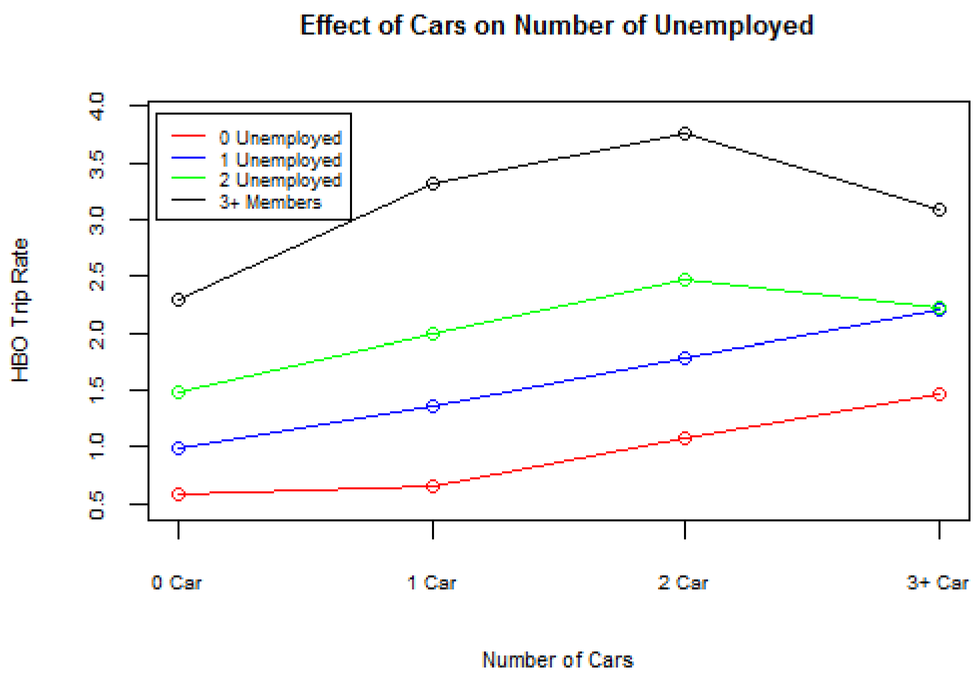




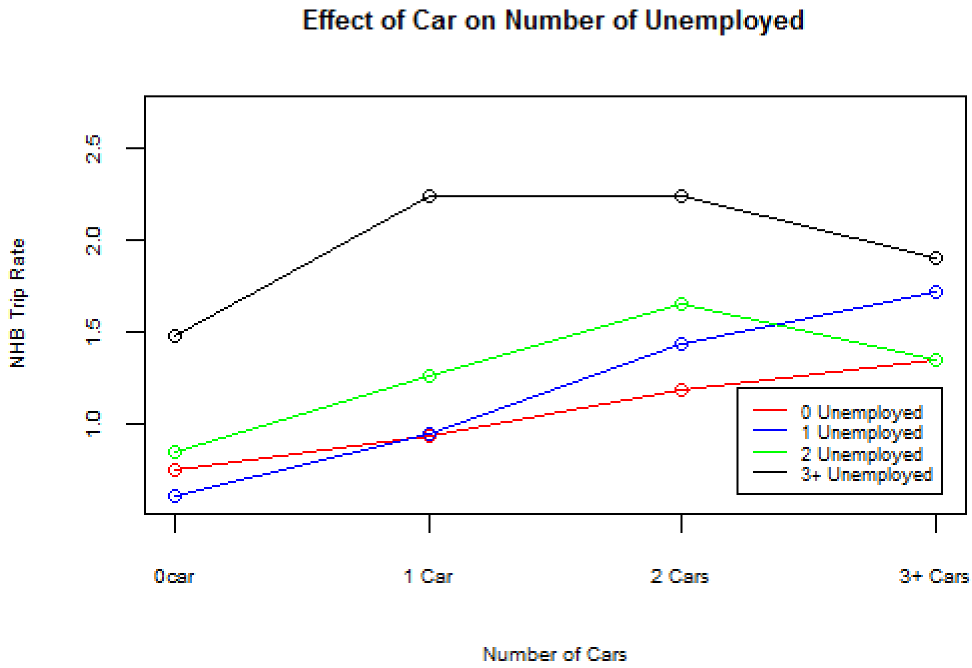
Home-Based Other Trips

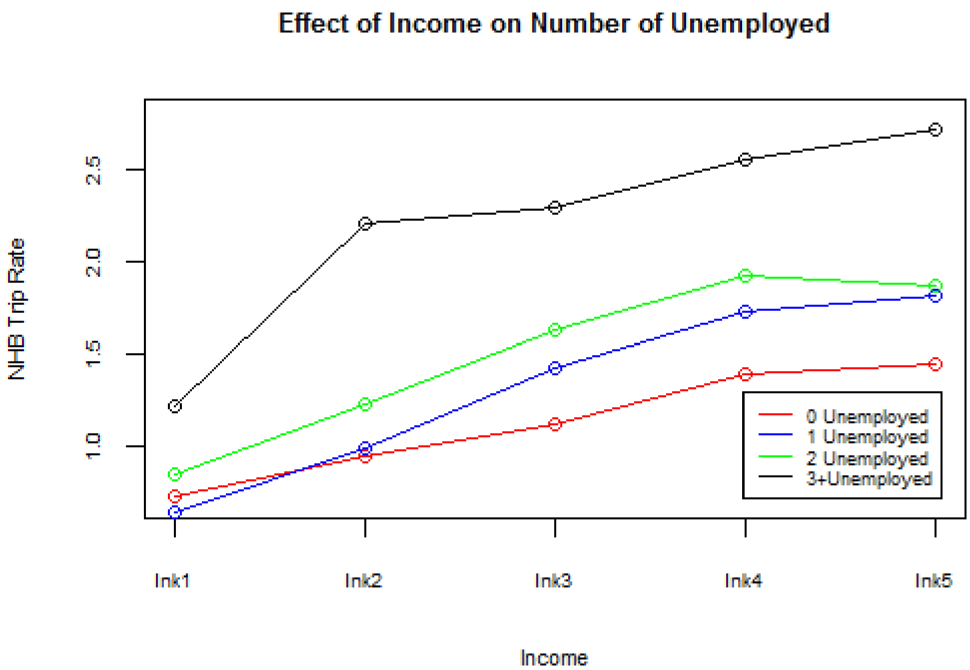


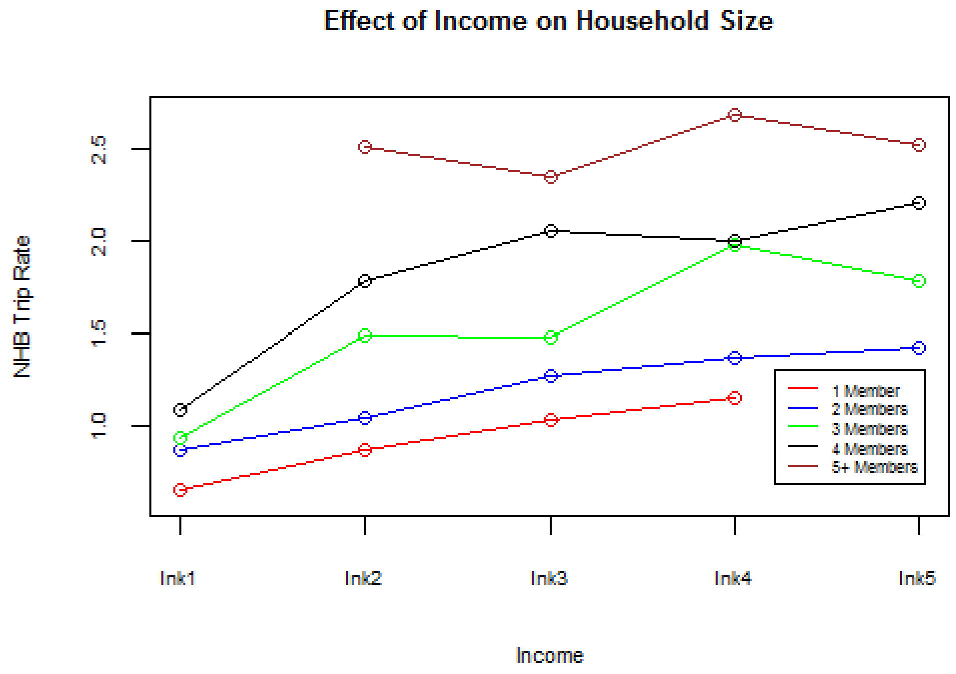

Effect of Number of Cars on Household Size
1 Member
2 Members
3 Members
4 Members
5+ Members
HBO Trip Rate
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0 Car
1 Car
2 Cars
3+ Cars
Number of Cars


Effect of Cars on Number of Unemployed
0 Unemployed
1 Unemployed
2 Unemployed
3+ Members
HBO Trip Rate
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0 Car
1 Car
2 Car
3+ Car
Number of Cars

Non-home-based trips






Effect of Income on Household Size
NHB Trip Rate
1.0
1.5
2.0
2.5
Ink1
Ink2
Ink3
Ink4
Ink5
Income
1 Member
2 Members
3 Members
4 Members
5+ Members

Declaration concerning the Master’s Thesis

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, June 21st, 2017

Aashish Pokhrel