Master Thesis for the Attainment of the Degree Master of Science at the TUM School of Management at Technische Universität München

Understanding Car Ownership in Germany: Findings from Using Statistical and Machine Learning Methods

Reviewer:	Prof. Dr. Gunther Friedl Chair for Management Accounting
Advisor:	Lukas Michael Schloter
Study Program:	M.Sc. Management and Technology
Submitted by:	Vanessa Wanner Tegernseer Landstraße 129 81539 München Matriculation number: 03716382
Submitted on:	15th of January 2021

Abstract

The ever-rising number of cars in Germany exacerbates environmental problems, demonstrating the need for a thorough understanding of car ownership to foster a decrease in overall car numbers. This paper aims to understand the general factors determining car ownership decisions in German households and to examine possible regional differences. Moreover, it analyzes the prediction capabilities of a statistical and a machine learning model regarding car ownership levels. A multinomial logit model featuring different car ownership levels is estimated based on data from the "Mobilität in Deutschland" survey in the year 2017. To analyze the differences in impacts among urban, suburban and rural regions, interaction effects are introduced into the model. Lastly, the prediction performances of a random forest and the multinomial logit model are compared. Our findings suggest that: (1) The number of licensed drivers and household income are the strongest determinants of car ownership. (2) The promotion of carsharing, as well as a better quality of public services may offer higher potential for lowering car ownership than the quality of public transport, especially in urban areas. (3) A random forest algorithm outperforms the statistical model in terms of predictive capability, confirming the usefulness of machine learning.

Die nach wie vor steigende Anzahl von Autos in Deutschland verschärft Umweltprobleme und zeigt, dass ein tiefergehendes Verständnis von Autobesitz notwendig ist, um eine Reduzierung der Zulassungszahlen zu fördern. Diese Studie versucht generelle Faktoren zu identifizieren, die auf die Entscheidung ein Auto zu besitzen Einfluss nehmen. Dabei sollen insbesondere regionale Unterschiede berücksichtigt werden. Des Weiteren werden die Prognosefähigkeiten von statistischen Modellen und Algorithmen, die maschinelles Lernen verwenden, hinsichtlich Autobesitz analysiert. So wird auf Basis der Studie "Mobilität in Deutschland" aus dem Jahr 2017 ein multinomiales logistisches Modell mit mehreren Kategorien hinsichtlich der Anzahl von Autos je Haushalt bestimmt. Um die Unterschiede der Einflüsse in städtischen, vorstädtischen und ländlichen Gebieten zu bestimmen, werden Interaktionseffekte verwendet. Außerdem wird die Forecasting-Performance eines Random Forest Algorithmus mit der eines multinomialen logistischen Modells verglichen. Die Resultate deuten darauf hin, dass: (1) die Anzahl von Führerscheinbesitzern und Haushaltseinkommen die stärksten Einflussfaktoren für Autobesitz darstellen, (2) die Förderung von Carsharing, ebenso wie die Qualität der Nahversorgung einen größeren Hebel zur Reduktion von Autobesitz bieten können als die Qualität des öffentlichen Nahverkehrs, vor allem in Städten, (3) die Prognosefähigkeit des Random Forest die des statistischen Modells übertrifft und damit die Verwendung von maschinellem Lernen bekräftigt.

Table of contents

List of Figures **III**

List of Tables **V**

List of Appendices **VII**

List of Abbreviations **VIII**

1 Introduction **1**

2 Car Ownership Research **3**

2.1 Overview of Car Ownership Studies Over Time 3

2.2 Findings from Disaggregate Models of Household Car Ownership 4

2.2.1 Car Ownership in Developed Countries 4

2.2.2 Car Ownership in Emerging Countries 11

3 Transportation Research in Germany **14**

4 Artifical Intelligence in Transportation Research **18**

5 Data and Methodology **20**

5.1 Data Assembly 20

5.2 Evaluating Feature Effects on Car Ownership 21

5.2.1 Model Selection and Further Considerations 22

5.2.2 Theoretical Background of the Multinomial Logit Model 23

5.2.3 Definition of Explanatory Variables 26

5.3 Evaluating the Influence of Regional Environment on Feature Effects 32

5.4 Prediction of Car Ownership 34

5.4.1 Generating Predictions Using Different Classifiers 34

5.4.2 Tuning Process and Measurement of Prediction Performance 35

5.4.3 Determining Variable Importance 37

6 Results and Interpretation **39**

6.1 Analysis of Car Ownership Levels in Germany 39

6.1.1 Model Validation 40

6.1.2 Model Results 41

6.2 Variance in Feature Impact in Different Regional Environments 51

6.3 Comparing the Prediction Capability of Statistical and Machine Learning Classifiers 59
6.3.1 Tuning Process 60
6.3.2 Predicition Capability and Variable Importance 61

7 Conclusion **63**

8 Limitations and Outlook **65**

Appendix **66**

Bibliography **93**

List of Figures

1 Figure 1: Operation mode of a random forest 36
2 Figure 2: Influence of the number of license holders and monthly income on predicted probabilities 44
3 Figure 3: Influence of average trip length and the average number of trips on predicted probabilities 45
4 Figure 4: Influence of the number of full-time and part-time workers on predicted probabilities 46
5 Figure 5: Influence of children and the prevailing generation on predicted probabilities 47
6 Figure 6: Influence of carsharing subscriptions on predicted probabilities 48
7 Figure 7: Influence of the number of bikes and motorbikes on predicted probabilities 49
8 Figure 8: Influence of housing type, garage availability and region on predicted probabilities 49
9 Figure 9: Influence of the quality of public transport and public services on predicted probabilities 50
10 Figure 10: Influence of the distance to public transport stations on predicted probabilities 51
11 Figure 11: Influence of the interaction between the variables ”licenses” and ”region” on predicted probabilities 53
12 Figure 12: Influence of the interaction between the variables ”full-time” and ”region” on predicted probabilities 53
13 Figure 13: Influence of the interaction between the variables ”part-time” and ”region” on predicted probabilities 54
14 Figure 14: Influence of the interaction between the variables ”children” and ”region” on predicted probabilities 55
15 Figure 15: Influence of the interaction between the variables ”generation” and ”region” on predicted probabilities 56
16 Figure 16: Influence of the interaction between the variables ”carsharing” and ”region” on predicted probabilities 56
17 Figure 17: Influence of the interaction between the variables ”bikes” and ”region” on predicted probabilities 57

18 Figure 18: Influence of the interaction between the variables ”housing type” and ”region” on predicted probabilities 57
19 Figure 19:Influence of the interaction between the variables ”garage” and ”region” on predicted probabilities 58
20 Figure 20: Influence of the interaction between the variables ”quality of public services” and ”region” on predicted probabilities 59
21 Figure 21: Out-of-bag error for different numbers of trees in the random forest algorithm 60
22 Figure 22: Mean balanced accuracy for different nodesizes and number of splitting variables 61

List of Tables

1 Table 1: Variables description and statistics 29
2 Table 2: Model validation measures for preliminary and final fit 32
3 Table 3: Overall model evaluation 40
4 Table 4: Coefficients of the MNL Model 42
5 Table 5: Average marginal effects and standardized coefficients 43
6 Table 6: Comparison of car ownership predictions 61
7 Table 7: Comparison of variable importance 63

List of Appendices

1 Appendix 1: Overview of spearman correlation coefficients between explanatory variables 66
2 Appendix 2: Overview of variance inflation factors for predictor variables 67
3 Appendix 3: Forward selection process of the preliminary model 67
4 Appendix 4: Overview of coefficients for separate logits and the preliminary model 68
5 Appendix 5: Logit(One Car) - Relation between explanatory variables and the logit before transformations 69
6 Appendix 6: Logit(One Car) - Relation between explanatory variables and the logit after transformations 69
7 Appendix 7: Logit(Two Cars) - Relation between explanatory variables and the logit before transformations 70
8 Appendix 8: Logit(Two Cars) - Relation between explanatory variables and the logit after transformations 70
9 Appendix 9: Logit(Three Cars) - Relation between explanatory variables and the logit before transformations 71
10 Appendix 10: Logit(Three Cars) - Relation between explanatory variables and the logit after transformations 71
11 Appendix 11: Logit(One car) - Binned Plots 72
12 Appendix 12: Logit(Two cars) - Binned Plots 72
13 Appendix 13: Logit(Three cars) - Binned Plots 72
14 Appendix 14: Forward selection process of the final model 73
15 Appendix 15: Log-odds of the final MNL model 74
16 Appendix 16: Model validation measures of the model including interactions . . 75
17 Appendix 17: Coefficients of the MNL model including interactions 76
18 Appendix 18: AMEs for interactions between independent variables and the variable ”region” 78
19 Appendix 19: Complete effect plots for the interaction between the variables ”children” and ”region” 79
20 Appendix 20: Complete effect plots for the interaction between the variables ”generation” and ”region” 79
21 Appendix 21: Complete effect plots for the interaction between the variables ”carsharing” and ”region” 80

22 Appendix 22: Complete effect plots for the interaction between the variables ”bikes” and ”region” 80
23 Appendix 23: Complete effect plots for the interaction between the variables ”housing” and ”region” 81
24 Appendix 24: Complete effect plots for the interaction between the variables ”garage” and ”region” 81
25 Appendix 25: Complete effect plots for the interaction between the variables ”income” and ”region” 82
26 Appendix 26: Complete effect plots for the interaction between the variables ”triplength.av” and ”region” 82
27 Appendix 27: Complete effect plots for the interaction between the variables ”trips.av” and ”region” 83
28 Appendix 28: Complete effect plots for the interaction between the variables ”motorbikes” and ”region” 83
29 Appendix 29: Complete effect plots for the interaction between the variables ”train” and ”region” 84

List of Abbreviations

AIC	Akaike information criterion
AME	Average marginal effect
GVIF	Generalized variance inflation factor
H-L	Hosmer and Lemeshow
MDA	Mean decrease accuracy
MDG	Mean decrease gini
MID	Mobilität in Deutschland
ML	Machine learning
ML	Multinomial logit
NN	Neural network
RF	Random forest
SVM	Support vector machines
U.K.	United Kingdom
U.S.	United States
VIF	Variance inflation factor

1 Introduction

In recent years, in light of growing environmental pollution and the resulting climate change, sustainable transportation development has increased in priority worldwide. One transportation-related factor that has been numerously confirmed to harm the environment and society is rising car traffic.[1] Problems caused by an increasing number of vehicles such as noise and greenhouse gas emissions, traffic congestions, accidents and energy consumption present challenges for governments and cities around the world and demonstrate the need for a change from high private car ownership and usage to more sustainable forms of mobility.[2] As a result, worldwide transport development increasingly places more weight on sustainability.[3]

Since 1970, the German government has been trying to reduce the amount of cars by the implementation of policies and urban development, such as higher petrol taxes and increased investment in public transport.[4] However, despite these measures, the number of cars per 1000 inhabitants has massively increased over the last decades.[5] The demand for cars in Germany grew by 32 % between 1995 and 2005,[6] while it grew by another 12 % in the last decade only.[7] Especially German cities are affected by high growth in car ownership rates and the associated problems, with annual growth rates of 2 %.[8] When looking at the number of cars per household in Germany, the share of carless households and households with one car has decreased by 1 % since 2008 while the share of households with more than one car increased. Nowadays, every fourth household is in possession of two or more cars.[9] To foster sustainable transport policies and lower car ownership rates in the future, a thorough understanding of present car ownership is needed. In spite of its high degree of mobilization, Germany has rarely been a focus of analysis regarding car ownership and its influence factors.

The objective of this research is threefold. First, this study explores the overall drivers of household car ownership in Germany on a national level. During the last decades, Germany experienced a population shift from rural regions towards urban areas, which is expected to

1 Cf. Schiller et al. (2010), pp. 8-12; cf. Nieuwenhuijsen / Khreis (2019), pp. 55-57.
2 Cf. Banister (2007), p. 9.
3 Cf. Schiller et al. (2010), p. 73.
4 Cf. Buehler / Pucher (2009), pp. 21-22.
5 The measurement for the number of cars per 1,000 inhabitants conducted by the Kraftfahrtbundesamt has changed in 2007. We therefore report growth in two separate time periods.
6 Cf. European Environment Agency (2011).
7 Cf. Statistisches Bundesamt (2020).
8 Cf. Hägler (2020).
9 Cf. Bundesministerium für Verkehr und Infrastruktur (2019a), pp. 10-11.

continue in the future.[10] In the light of these changes, it is even more important for policymakers to be able to understand the background of car ownership in different regional environments. Hence, secondly, we aim to analyze possible variations of certain influence factors on car ownership in different regional settings. Lastly, even though the application of machine learning (ML) algorithms has proved its performance in fields other than information technology over an extended period, the use in car ownership studies is scarce, and the utilization of traditional statistical methods prevails. Thus, this paper's third objective is to assess the predictive capability of our statistical model regarding car ownership in Germany and the comparison with an application of ML. The corresponding research questions that will be answered are the following: Which socio-economic and built-environment attributes will have the greatest influence on car ownership levels? Will the impact be positive or negative? Are households living in an urban environment affected differently in their car ownership choice by certain attributes than households in suburban and rural regions, or are the effects the same among all regions? How good is the predictive performance of a multinomial logit model (MNL) in comparison with a ML classifier in an equal mobility-related scenario?

To address the research questions presented above, this thesis is structured into three steps. First, drawing on a nationwide mobility study, a MNL model, exploring the effect of different socio-economic and built-environment variables on car ownership levels, is estimated. As we are particularly interested in the influence of urban form, the differences in the effects of influencing variables in urban, suburban and rural regions are analyzed in a second step using interaction effects. Following the regional analysis, thirdly, we investigate the model's predictive capability in comparison with a random forest (RF) algorithm. Together, the results can offer guidance for policymakers and practitioners in infrastructure and mobility regards.

The paper will proceed as follows. Chapters 2 to 4 will identify and summarize existing research conducted in fields relevant to this thesis, namely car ownership research worldwide, car ownership and car-related research in Germany, and lastly, the application of ML in transportation research. Hereafter, the data preparation process and applied methodologies are elaborated in Chapter 5. Subsequently, Chapter 6 presents and discusses overall results, beginning with the interpretation of general influence factors relevant to car ownership levels in Germany. Additionally, the analysis of regional differences of attribute effects and the comparison of predictive performance are also reported in this part. Lastly, we will conclude our findings and give an

[10] Cf. Bund-Länder Demografie Portal (2020).

outlook on further research towards the topic.

2 Car Ownership Research

The following section reports on existing research studies regarding car ownership in terms of the decision why to own a car and the number of cars owned, respectively. First, the development of applied methods in car ownership research is presented and subsequently findings in developed and emerging countries are presented. Findings from research in Germany are excluded form this overview, as they will be reflected separately in Chapter 3.

2.1 Overview of Car Ownership Studies Over Time

Models and studies with a focus on private car ownership and its prediction have been developed since the 1930s,[11] resulting in a wide field of research concerning the subject, featuring a high amount of different models, focuses and research methodologies.[12] Ortúzar and Willumsen find that models can be classified into two different approaches: aggregate and disaggregate models.[13] Aggregate models use data aggregated at the regional or national level, while disaggregate models use data found at the household level.[14] Earlier research up to the 1970s has been focusing on the use of aggregate models, with studies using cross-sectional data[15], time-series data[16] or both[17]. The popularity can be explained by the ease of application of these models,[18] however, flexibility, accuracy and policy sensitivity are criticized for being very limited as aggregate models tend to fail to "capture the causal relationships underlying household behavior."[19] Moreover, they often suffer from collinearity and existing biases due to aggregation.[20] While aggregate models may be useful for forecasting vehicle ownership on a regional or national level, they are ineffictive in evaluating features that influence car ownership patterns, which is essential for the assessment and design of transportation policies. Therefore, more recent research focused on the application of disaggregate models. These models can capture existing heterogeneity of observations and causal relationships between the outcome and its determinants and deliver more precise results.[21] Thus, they overcome problems encountered with

[11] Cf. De Wolff (1938), p. 113; cf. Tanner (1958), (as qtd. in Whelan (2007), p. 206).
[12] Cf. Anowar et al. (2014), p. 442; cf. De Jong et al. (2004), p. 380.
[13] Cf. Ortúzar / Willumsen (2011), pp. 18-19.
[14] Cf. Yagi / Managi (2016), p. 9.
[15] Cf. Kain / Beesley (1965), pp. 171-172.
[16] Cf. O'Herlihy (1965), p. 171.
[17] Cf. Tanner (1978), p. 29.
[18] Cf. Ortúzar / Willumsen (2011), pp. 18-19.
[19] Kitamura / Bunch (1990), p. 477.
[20] Cf. Potoglou / Kanaroglou (2008a), p. 235.
[21] Cf. Eluru / Bhat (2007), p. 1038.

aggregate models, such as low accuracy and unobserved multicollinearity.[22]

The development of car ownership models in literature has been analyzed in recent literature reviews. De Jong et al. present an overview of aggregate and disaggregate models until 2002 and compare nine defined model types based upon 16 different criteria.[23] Potoglou and Kanaroglou compare models that focus on car ownership and the choice of alternative fueled vehicles.[24] A more recent review was published by Anowar et al., focusing on disaggregate models and their classification into four categories (exogenous static models, endogenous static models, exogenous dynamic models and endogenous dynamic models).[25] The following sections will report on findings from exisiting disaggregate models of car ownership.

2.2 Findings from Disaggregate Models of Household Car Ownership

Possible determinants of car ownership have been explored in a multitude of studies, in developed as well as in emerging countries. In order to create a comprehensive review of worldwide car ownership research to this date, the following two sections will, first, introduce the included studies and, secondly, report on findings.

2.2.1 Car Ownership in Developed Countries

To achieve a robust overview, studies from different parts of the developed world are analyzed. First, an overview of studies is given and afterwards, findings are presented grouped by categories. Results regarding more general influence factors are introduced first, followed by studies analyzing more current issues. Early studies were conducted by Dargay and Whelan in the United Kingdom (U.K.). Dargay's study analyzed car ownership in the U.K. based on mobility data from 1982 to 1995,[26] while Whelan designed a model to forecast future car ownership in Great Britain using travel and household survey data from 1971 to 1996.[27] Further studies in the European area comprise the results by Caulfield, as well as Matas and Raymond. Caulfield investigated multiple car ownership in the Dublin area using census data from 2006[28] and Matas and Raymond later examined features related to car ownership in Spain in the timespan from 1980 to 2000.[29] Moving on to studies with a focus on North America, a study by

[22] Cf. Potoglou / Kanaroglou (2008a), p. 235.
[23] Cf. De Jong et al. (2004), p. 380.
[24] Cf. Potoglou / Kanaroglou (2008a), p. 235.
[25] Cf. Anowar et al. (2014), pp. 442-443.
[26] Cf. Dargay (2002), p. 352.
[27] Cf. Whelan (2007), p. 207.
[28] Cf. Caulfield (2012), p. 133.
[29] Cf. Matas / Raymond (2008), p. 187.

Potoglou and Kanaroglou is included, who investigated the relationship between car ownership and urban form in Hamilton, Canada, using household data from 2005.[30] Moreover, Ryan and Han focused on socio-economic features and accessibility in their study in Hawaii in 1995[31] and Chu developed a model to predict car ownership in New York City based upon household survey data from 1997 and 1998.[32] Bento et al. focused especially on the influences of urban form and transit supply when analyzing transportation surveys of the year 1990.[33] The last two studies included analyzing data of the United States (U.S.) were conducted by Bhat and Guo using data from the year 2000.[34] Bhat and Guo examined car ownership and its relation to built-environment, public transport attributes and demographics in the San Francisco Bay area[35] Lastly, other than European and North American studies, we report on findings from Sun et al., who assessed demographic features of car ownership in the Osaka area in Japan from 1970 to 2010[36] and from Soltani, who examined car ownership in the city of Adelaide in South Australia in 2001.[37]

Investigated influence factors regarding car ownership can be broadly divided into two subcategories: built-environment features and socio-economic features. In terms of built-environment variables and their influence on car ownership, commonly included are different population density measures such as population centrality, household density, employment density or mixed density. Bento et al. found that living in a more densely populated city reduces the likelihood of owning one, two, and three vehicles compared to more sprawled cities. A 10 % increase in population centrality led, for example, to a reduction in the probability of owning two cars by about 1.5 %.[38] Results of a study by Soltani who investigated the Adelaide area also indicated that higher dwelling density had a negative influence on vehicle ownership while, in contrast, employment density was insignificant.[39] This is in line with other studies like Caulfield's, who found residential density to be one of the variables with the greatest impact on the number of cars a household owned.[40] Bhat and Guo only found a low influence of household and employment density on car ownership rates, which they attributed to the inclusion of (correlated) transport network variables. When transport network variables are removed from the model,

[30] Cf. Potoglou / Kanaroglou (2008b), p. 46.
[31] Cf. Ryan / Han (1999), p. 1.
[32] Cf. Chu (2002), p. 62.
[33] Cf. Bento et al. (2005), p. 466.
[34] Cf. Bhat / Guo (2007), p. 515; cf. Weinberger / Goetzke (2010), p. 2116.
[35] Cf. Bhat / Guo (2007), p. 523.
[36] Cf. Sun et al. (2014), p. 518.
[37] Cf. Soltani (2005), p. 2154.
[38] Cf. Bento et al. (2005), p. 474.
[39] Cf. Soltani (2005), p. 2159.
[40] Cf. Caulfield (2012), p. 132.

density measures became more negative and significant. The authors conclude that previous studies have used density measures as proxies for transportation measures such as street block density and transit accessibility.[41] Potoglou and Kanaroglou investigated the influence of mixed density, which considers both the employment density per acre and the number of households per acre. They found that as the number of jobs and households increased in a zone, the probability of owning two or more vehicles declined. However, the probability of owning one vehicle was not affected.[42]

Beyond density measures, popular influence factors are transportation network measures and public transport availability. Bhat and Guo found a highly significant negative influence of street block density on car ownership in general. However, they observed high amounts of heterogeneity in responses to street block density as 23 % of observed households increased car ownership as street block density increased.[43] Bento et al. included other transportation network measures like road density and cityshape, but found that only vehicles miles driven is affected and not the number of cars owned.[44] With regard to public transport availability and quality, most studies found a small negative influence on car ownership rates. Findings by Bhat and Guo showed that households living in areas where public transport is available, have a decreased likelihood of owning cars than households in areas with no availability. Additionally, the time that is needed to access public transport also influenced car ownership rates. A longer access time was associated with a positive influence on car ownership. They also analyzed the influence of commuting time and commuting cost, finding that longer commuting time is associated with higher car ownership levels, while higher commuting cost was associated with lower car ownership levels, which is in line with intuitive expectations.[45] Potoglou and Kanaroglou included the number of bus stops within walking distance into their model and found that a higher number of stops only decreased the probability to own three or more vehicles and did not affect first and second vehicle ownership. Results suggest that higher public transport availability may reduce car ownership but cannot erase it completely.[46] Caulfield used the same approach and found that household living in areas with a low amount of bus stops showed a greater likelihood of owning more than one car. He also investigated the influence of rail availability, where results indicated that missing access to rail transit increases both, the likelihood

41 Cf. Bhat / Guo (2007), p. 524.
42 Cf. Potoglou / Kanaroglou (2008b), p. 51.
43 Cf. Bhat / Guo (2007), p. 520.
44 Cf. Bento et al. (2005), p. 477.
45 Cf. Bhat / Guo (2007), p. 520.
46 Cf. Potoglou / Kanaroglou (2008b), p. 51.

to own one car and the likelihood to own several cars. However, even though in the analyzed area, in Dublin, 54 % of households had access to rail transit, car ownership rates were still at a high level.[47] Bento et al. also investigated bus and rail transit supply and found that higher bus supply resulted in minor negative effects on the probability of owning two vehicles, while rail transit had no influence on car ownership, only on average miles driven.[48] Soltani, in his study of Adelaide, found that higher public bus coverage of an area resulted in a decreased likelihood of owning and using private vehicles.[49]

Some variables are used in a wide range of studies, others, however, are used more scarcely. A built-environment variable that is included in some models is land-use diversity, referring to the heterogeneity of land uses in a given area, measuring for example the mixture of residential premises, schools, businesses and parks. Chu found that mixed-use development has a more significant negative influence on car ownership in New York City than employment or mixed density.[50] These results were confirmed by Potoglou and Kanaroglou in their study of Hamilton[51] and by Soltani in his study of Adelaide. Households in more homogenous areas were associated with higher car ownership rates than households in areas with diverse usage.[52] A different variable analyzed was the type of housing. Potoglou and Kanaroglou found that living in a single-family house increased the likelihood to own vehicles compared to living in other housing types.[53] Chu also finds a moderate positive influence of single-family housing on the probabilities of owning one or two vehicles.[54]

Concerning socio-economic features, there is a variety of variables that are included in different studies. A variable found to be highly significant for car ownership in almost every study is household income. Potoglou and Kanaroglou came to the result that in Hamilton, Canada, household income was the dominant feature in determining car ownership rates. Especially the medium income class increased the likelihood of a household to own one vehicle. In turn, the high-income class increased the probability of owning two vehicles but showed a lower significance, probably because a higher income rather leads to the purchase of more expensive cars than to several ones.[55] Matas and Raymond in their study in Spain found that income was

47 Cf. Caulfield (2012), p. 138.
48 Cf. Bento et al. (2005), p. 474.
49 Cf. Soltani (2005), p. 2158.
50 Cf. Chu (2002), p. 65.
51 Cf. Potoglou / Kanaroglou (2008b), p. 51.
52 Cf. Soltani (2005), pp. 2159-2160.
53 Cf. Potoglou / Kanaroglou (2008b), p. 50.
54 Cf. Chu (2002), p. 66.
55 Cf. Potoglou / Kanaroglou (2008b), pp. 50-51.

a significant variable and main contributing factor regarding car ownership in all three years, however, income elasticity was lower in rural areas than in urban areas regarding car ownership and generally declined when car ownership levels increased.[56] Dargay came to similar results when investigating car ownership in the U.K., as in his study car ownership increased with income but at the same income level, elasticity for rural households was lower than for urban ones. The reason for this phenomenon may be that rural areas tend to have a higher rate of car ownership and are, therefore, closer to a saturation point. Moreover, cars in non-urban regions are even more an essential rather than a luxury good compared to urban environments.[57] However, Chu found that, even in highly urbanized environments like New York City, income is one of the most important features in determining car ownership.[58]

The second category of socio-economic features that is commonly used is directed towards occupation and household size. Whelan developed a model to predict car ownership in Great Britain at the disaggregate level, focusing on demographic variables.[59] Results suggest that the number of employed adults strongly influences car ownership and the effects are greater than just the effect through additionally generated income. At the same income level, one additional worker increased the probability of car ownership for all categories (one or more cars, two or more cars, three or more cars) by up to 7 %.[60] Potoglou and Kanaroglou found that the number of working adults on car ownership positively influenced the probability of owning two or more cars as the variable characterizes households with higher mobility needs. On the other hand, part-time workers were associated with a lower likelihood to own both one or two cars, which might be explained by lower disposable income.[61] Caulfield showed that the individual's occupation also had an impact on car ownership. In his study of Dublin, results indicated that less skilled and unskilled employees and workers in the agricultural sector were unlikely to belong to a household that owned multiple cars. On the other hand, managers and professionals showed a higher likelihood of belonging to a household with several cars, even when controlling for different income levels.[62] Matas and Raymond found that household size had a positive association with car ownership and that the effect was more extreme for working adults. The coefficients proved to be larger in rural environments than in urban ones and increased over

[56] Cf. Matas / Raymond (2008), p. 200.
[57] Cf. Dargay (2002), p. 361.
[58] Cf. Chu (2002), p. 67.
[59] Cf. Whelan (2007), p. 205.
[60] Cf. Whelan (2007), p. 212.
[61] Cf. Potoglou / Kanaroglou (2008b), p. 50.
[62] Cf. Caulfield (2012), pp. 134-135.

time, which can be attributed to increased mobility needs in recent times.[63] Instead of looking at working adults, some studies also investigated the influence of the number of licensed drivers. Chu found that in New York City, the number of household members with a driving license was a more dominant determinant of car ownership than household size or the number of workers.[64] Potoglou and Kanaroglou also found different influences for both of the variables. While the number of workers only increased the probability of owning two or more cars, the number of licensed drivers increased the probability of owning one or more cars.[65]

Multiple studies investigated not only household size but also household composition and how the structure of a household affected the number of cars owned. Caulfield found that single households or parents living without their children had a decreased likelihood of owning more than one car, while childless couples and couples with children were likely to have several cars. Couples with children over the age of 19 were even most likely to own three cars, which can be attributed to the fact that these children have bought own cars.[66] These findings are in line with results by Potoglou and Kanaroglou who found that dummy variables for childless couples and couples with children showed a strong preference for owning two cars, while households compromising extended family or unattached individuals were likely to own three or more cars.[67] In their study of the Honolulu area in Hawaii, Ryan and Han developed a variable specification representing the order of household member classes. Their model showed positive coefficients, which decreased with decreasing importance of family members. The authors also found that households with children are unlikely to decide not to have a car.[68]

More recent studies on car ownership have focused on the influence of current issues on car ownership, such as the influence of the millennial generation in the U.S. analyzed by Lavieri et al.[69] and Knittel and Murphy[70], or the impact of carsharing options, investigated by Kim et al. in 2019 using data from Korea.[71] Lavieri et al. found that age, parenting status, and features of the place of residence had the greatest influence on the millennial generation's automobile choices. Higher age and income as well as the existence of children had a positive effect on car ownership and use. Transportation mode choice had a strong relationship with the place of

63 Cf. Matas / Raymond (2008), p. 196.
64 Cf. Chu (2002), p. 67.
65 Cf. Potoglou / Kanaroglou (2008b), p. 50.
66 Cf. Caulfield (2012), p. 136.
67 Cf. Potoglou / Kanaroglou (2008b), p. 50.
68 Cf. Ryan / Han (1999), p. 6.
69 Cf. Lavieri et al. (2017), p. 92.
70 Cf. Knittel / Murphy (2019), p. 1.
71 Cf. Kim et al. (2019), p. 126.

residence.[72] This goes in line with general car ownership studies, previously presented in this chapter. Knittel and Murphy confirmed these findings as in their study only small differences were found between the millenial generation and other generations in terms of car ownership and car use. Even though millenials lived in urban environments more often, married later and had larger families, the influence on car ownership is negligible.[73] Kim et al. analyzed the influence of the introduction of a carsharing program on vehicle ownership in Korea and found that 31 % of car sharing users reduced car ownership or deferred car purchases because of the possiblity of carsharing in both, the early and the mature phase of the program. Additionally, the proportion of people who disposed a private car due to carsharing rose from 2 % in 2014 to 4 % in 2018. Beside the accessibility of the car sharing services, customer service satisfaction and payable fees were found to be essential to have an effect on vehicle ownership. Households with a lower income and older members were more likely to reduce car ownership than others, while higher income households are unlikely to dispose of a car but will refrain from buying a second one. Furthermore, results indicated that members who use carsharing for commuting are more likely to decrease their car ownership than members who use carsharing for leisure or shopping trips.[74]

In general, studies conclude that both built-environment and socio-economic characteristics are important and have significant influence. Still, most studies find that household attributes are even more dominant.[75] However, when looking at a longer timespan, the large increase in the number of cars might be largely influenced by built-environment attributes as shown in the study by Sun et al. They consider (exogenous) demographic features like age, household life-cycle, residential area and (endogenous) features such as car ownership, number of car trips and travel duration in their study of the Osaka area in Japan.[76] During the considered timespan, average car ownership increased from 0.47 vehicles in 1970 to 1.32 vehicles in 2010. Using a simultaneous equation model, they found that the increase could be mainly attributed to changes in land use, transportation systems and geographical expansion. Demographic changes, on the contrary, even fostered a reduction in car ownership of 15 % in the 30 years in questions, that was offset by the increase induced by changes in the built-environment.[77]

---

[72] Cf. Lavieri et al. (2017), p. 91.
[73] Cf. Knittel / Murphy (2019), pp. 12-13.
[74] Cf. Kim et al. (2019), p. 123.
[75] Cf. Bhat / Guo (2007), p. 524.
[76] Cf. Sun et al. (2014), p. 517.
[77] Cf. Sun et al. (2014), pp. 525-526.

2.2.2 Car Ownership in Emerging Countries

Most car ownership studies and reviews have been conducted in developed countries, while studies on developing countries are more sparse and have only started to pick up in more recent years.[78] There is especially a lower amount of disaggregate studies and existing studies are often based on smaller surveys. Greater scale household databases or centrally conducted household surveys rarely exist in developing countries and private collection costs are often high.[79] Both Yamamoto and Sanko et al. examined and compared interactions among the ownership of different vehicle types, including cars, motorcycles and bicycles, in Asian countries with different develoment stages based on data before 2000.[80] Li et al., Jiang et al., Ma et al. and Wong all analyzed vehicle ownership in Chinese cities. Li et al. examined the influence of urban form, socio-economic and demographic factors on car ownership in 36 megacities in China based on data collected in 2006.[81] Ma et al. explored the correlation between the ownership of different vehicle types in the city of Hangzhou in China in 2015,[82] and Jiang et al. investigated the impact of land use and street characteristics on car ownership and usage based on travel surveys conducted in 2014 in Jinan.[83] Wong analyzed ownership of cars and motorbikes in the town of Macao in the south of China in 2009.[84] Zegras conducted a study in Santiago de Chile, exploring the influence of built-environment in terms of design, density and diversity on general motor vehicle use and ownership using 2001 travel data.[85] Choudhary and Vasudevan analyzed vehicle ownership in India, focusing primarily on differences between urban and rural households in 2011 and 2012.[86] A recent study conducted by Ha et al. investigated the factors that guide household vehicle ownership in Cambodia and how to predict them by comparing traditional statistical models with more novel ML methods.[87]

It should be noted, that the scope of research questions in developing countries is often broader. Most studies focus on general vehicle ownership and less on car ownership only, since other vehicle types, such as motorbikes and electric bikes, as well as non-motorized vehicles, such as standard bikes, are still much more predominant in developing countries than they are in de-

[78] Cf. Ma / Ye (2019), p. 648.
[79] Cf. Ma / Ye (2019), p. 663.
[80] Cf. Sanko et al. (2012), p. 200; cf. Yamamoto (2009), p. 352.
[81] Cf. Li et al. (2010), pp. 77-78.
[82] Cf. Ma et al. (2018), p. 7.
[83] Cf. Jiang et al. (2017), p. 521.
[84] Cf. Wong (2013), p. 213.
[85] Cf. Zegras (2010), p. 1801.
[86] Cf. Choudhary / Vasudevan (2017), p. 54.
[87] Cf. Ha et al. (2019), p. 71.

veloped ones.[88] Li et al. found that households in Chinese megacities that owned electric bikes or standard bikes were less likely to own a car.[89] Ma et al. analyzed the correlation between the ownership of different vehicle types in the city of Hangzhou in China and found a substitutive effect between cars, motorbikes and electric bikes, as well as a substitutive effect between motorcycles and electric bikes.[90] Studies that compare multiple vehicle ownership between developing and developed countries came to similar results. Yamamoto analyzed the difference between Kuala Lumpur and Osaka and confirmed that the ownership of cars and motorbikes had a substitutive relationship in Malaysia, whereas in Japan, the ownership relation was complementary.[91] This outcome was confirmed by Sanko et al., whose study compared car and motorcycle ownership in Bangkok, Kuala Lumpur and Nagoya. Results indicated a substitutive relation between cars and motorbikes in developing countries like Thailand and Malaysia, while in a developed country like Japan, the relationship was complementary.[92]

The most frequently analyzed features that influence car ownership in developing countries are the distance to the city center, population density and household income. The effect of distance to the city center in studies investigating car or vehicle ownership in large cities seems to be contrary to findings in developed countries. Li et al., in their study of different Chinese megacities, found that the variable had a negative influence on car ownership, meaning that households that live far away from the city center were more likely not to have a car, while households living close to it were more likely to own one.[93] Zegras came to a similar conclusion in his study of Santiago de Chile. Even though within a distance of 10 km, households were more likely to own a car, households living further away from the city center were unlikely to possess a car.[94] Secondly, population density is a variable often considered by studies. Li et al., Yamamoto and Wong found a negative relation between population density and vehicle ownership,[95] which is similar to findings in developed countries. Jiang et al., in their study of China, found no influence of population density on car ownership, but only on car use.[96] A third variable that was shown to have a strong influence on car ownership in developing countries is income. Car ownership was positively correlated with higher income in all presented studies in developing countries. Ha et al., Jiang et al., Li et al., Yamamoto and Zegras all find that higher income was

88 Cf. Ma / Ye (2019), p. 659.
89 Cf. Li et al. (2010), p. 83.
90 Cf. Ma et al. (2018), p. 13.
91 Cf. Yamamoto (2009), p. 360.
92 Cf. Sanko et al. (2012), p. 211.
93 Cf. Li et al. (2010), p. 82.
94 Cf. Zegras (2010), p. 1806.
95 Cf. Li et al. (2010), p. 76; cf. Wong (2013), p. 217; cf. Yamamoto (2009), p. 351.
96 Cf. Jiang et al. (2017), p. 529.

a dominant variable when analyzing car ownership.[97] Zhang et al. find that household income was one of the most important factors when investigating car ownership in eastern regions in China, where a 10 % increase in income would lead to an increase of 9 % in vehicle ownership probability.[98] Other variables examined include the number of workers and general household size, for which different studies generated mixed results. Yamamoto found that the number of workers in a household determined motorized vehicle ownership,[99] while Zegras found that households with one adult worker were more likely to own a car than households with three adult workers. This may be explained by the fact that a higher number of workers in emerging countries is often linked to a greater household and family size, where income is needed for basic needs and is not available for vehicle expenditures.[100] This goes in line with findings by Wong when analyzing car ownership and its relation to household size in Macao, China. A higher number of adult household members, workers, and non-workers reduced the likelihood of owning a car, which is contrary to developed countries.[101] However, there are different findings in studies by Yamamoto in China and Ha et al. in Cambodia, where household size in terms of adult members had a positive influence on vehicle ownership of any type.[102]

Further variables that have also been analyzed in the context of emerging countries include the existence of children and land-use diversity at the location of residence. The number of children had a significant positive effect on car and vehicle ownership in a multitude of studies. Studies in China provide evidence that households with children show a stronger preference for owning a car or a vehicle.[103] The same was found in India, where the presence of children had a positive influence on vehicle ownership for both rural and urban households, although the impact on urban households was greater.[104] Zegras found a similar effect as households with up to four children were more likely to own one or two cars in his study. However, the influence on the ownership of three vehicles was negative, reflecting the reduced purchasing power and limited number of licenses in households with children.[105] Zegras also analyzed the influence of land-use diversity and other built-environment variables in Santiago de Chile and found that high land-use diversity and better bus services reduced general vehicle ownership choices. At

[97] Cf. Ha et al. (2019), pp. 79-80; cf. Jiang et al. (2017), p. 528; cf. Li et al. (2010), p. 82; cf. Yamamoto (2009), p. 365; cf. Zegras (2010), p. 1803.
[98] Cf. Zhang et al. (2017), p. 18.
[99] Cf. Yamamoto (2009), p. 359.
[100] Cf. Zegras (2010), p. 1803.
[101] Cf. Wong (2013), p. 216.
[102] Cf. Yamamoto (2009), p. 359; cf. Ha et al. (2019), pp. 79-80.
[103] Cf. Yamamoto (2009), p. 359; cf. Wong (2013), p. 217.
[104] Cf. Choudhary / Vasudevan (2017), p. 57.
[105] Cf. Zegras (2010), p. 1803.

the same time, proximity to metro stations only influenced second and third vehicle choices.[106] Ao et al. investigated the influence of built-environment in rural China and found a significant effect on car ownership as the accessibility to public services and public transport decreased the probability of owning two or more cars in rural regions.[107]

In general, the main influence factors chosen by studies in both emerging and developed countries, are widely similar, even though in developed countries a wider variance of influence factors was examined. Regarding the results, most variables show equivalent effects and, thus, do not seem to vary with the state of development of a country. However, some influence factors display opposing effects compared to results in developed countries in various studies. The factor which contributes the most to these differences seems to be the existing budget constraint for a higher number of households in developing countries.

3 Transportation Research in Germany

This section seeks to present some of the existing findings in the field of transportation research in Germany, commencing with car ownership research and then continuing with studies on topics such as car use and mobility behavior. To this date, research on car ownership explicitly has received very little attention in Germany, with only three major studies over the last decades. Ritter and Vance analyzed determining factors of car ownership and predicted future car ownership based on mobility data from 1999 to 2009, focusing mainly on how a change in average family size would translate into nationwide car numbers.[108] Moreover, Vance and Hedel investigated the effect of urban form features on car ownership and use in Germany between 1996 and 2003.[109] Finally, Keller and Vance combined household survey data with satellite images to explore the influence of landscape patterns on car ownership and use between 1996 and 2009.[110] The first study differenciated between different car ownership levels, while the latter only distinguished households with and without cars.

Ritter and Vance primarily focused on household size and certain policy-relevant variables, such as fuel and insurance prices, public transportation system and land use.[111] They found expected positive associations of higher age, a higher number of licensed drivers, household size, number

[106] Cf. Zegras (2010), p. 1812.
[107] Cf. Ao et al. (2019), pp. 34-35.
[108] Cf. Ritter / Vance (2013), p. 74.
[109] Cf. Vance / Hedel (2008), p. 63.
[110] Cf. Keller / Vance (2013), p. 6.
[111] Cf. Ritter / Vance (2013), p. 74.

of walking minutes to public transport and income with car ownership. Higher fuel costs, low public transport pricing and urban neighborhoods, on the other hand, had a negative influence on car ownership.[112] When simulating future car ownership rates in Germany assuming a yearly increase of 0.8 % in income, an increase in the number of cars was projected until 2030, even though population decreased. However, when no increase in income was assumed, the total number of cars even reduced slightly.[113] In the study conducted by Vance and Hedel, based on data from 1996 to 2003, urban form variables were found to be significant for car ownership and kilometers driven even when controlling for selectivity and endogeneity biases.[114] The two density measures, namely road density and commercial outlet density, had negative coefficients, implying that in more dense environments the need for car use is smaller. Walking distance to public transport had a positive relationship with car ownership, indicating that a further distance to the nearest transit stop increased the likelihood of owning a car. In terms of socio-economic variables, the strongest influence was exerted by the number of male workers in the household. A unit increase led to an increase in the probability of car ownership of 0.13 %.[115] Keller and Vance confirmed most of Vance and Hedel's results in their later study using geospatial data in addition to household surveys. They also found that a higher number of walking minutes to public transport increased the likelihood of car ownership. The availability of rail services was not significant for car ownership in their study. Moreover, higher business density decreased the probability of owning cars, as well as the distance driven and, unexpectedly, children reduced the likelihood of owning cars.[116] Regarding the variables derived from satellite images, only the share of open space exerted a significant positive influence on car ownership,[117] once more demonstrating that less dense environments stimulate car ownership.

The dimension of car use is strongly linked to car ownership and has been investigated in multiple studies with different focuses. Buehler and Kuhnert compared car use in the U.S. and Germany in the years 2001 and 2002.[118] Later on, Eisenmann and Buehler created detailed car use profiles from 2010 - 2014 household surveys in Germany and California and divided them by hierarchical cluster analysis into clusters featuring similar car-usage characteristics.[119] Best and Lanzendorf examined the influence of gender[120] and Frondel and Vance analyzed the

[112] Cf. Ritter / Vance (2013), p. 80.
[113] Cf. Ritter / Vance (2013), p. 83.
[114] Cf. Vance / Hedel (2008), p. 63.
[115] Cf. Vance / Hedel (2008), p. 57.
[116] Cf. Keller / Vance (2013), p. 15.
[117] Cf. Keller / Vance (2013), p. 16.
[118] Cf. Buehler / Kunert (2010), p. 16.
[119] Cf. Eisenmann / Buehler (2018), p. 174.
[120] Cf. Best / Lanzendorf (2005), p. 110.

impact of fuel taxes and fuel efficiency[121] on car use in Germany. Buehler and Kuhnert found that Americans undertook 20 % more trips and had significantly higher mobility needs than the German population.[122] Moreover, they were more car-dependent and drove longer distances, independent of socio-economic and geospatial factors. Looking at their results on distance traveled by car for Germany alone, nearby public transit stops, as well as higher population density and land-use diversity reduced distances. A higher income, number of licenses, number of workers and male household members increased car use. They concluded that dense neighborhoods with a diverse mix of usages led to lower car use.[123] Looking at car usage in detail, Eisenmann and Buehler found that the same eight types appeared in Germany and California, namely "standing cars, moderate-range cars, day-to-day cars, workday cars, weekend cruisers, long-distance cars, short-haul cars and all-rounders"[124], only varying in size. Interesting findings include that there is no dominant car usage cluster in Germany, however, leisure day cars and short-haul cars are relatively rare. Additionally, in California, all groups featured a higher share of battery electric vehicles and hybrid electric vehicles.[125]

Best and Lanzendorf assessed how the change in gender roles, with more women pursuing careers, affected car use in Cologne in 1997. Results suggested that car use by men and women did not differ a lot. Multivariate analysis showed that part-time employment and pursuing a job generally increased car use. At the same time, the existence of children lowered car use by women while, in turn, increasing male car use. Therefore, higher participation of women in the labor market increased car use, while the traditional gender roles in terms of child care led to less car use.[126] Frondel and Vance analyzed the effect of fuel price and fuel efficiency on vehicle kilometers traveled during the period from 1997 to 2015 in Germany. Estimated elasticities showed that even though increased fuel prices would have reduced driving substantially, higher fuel efficiency increased driving to an even greater extent. This indicates that decreased car use from fuel taxes may be offset by increased efficiency standards.[127]

Another line of research focuses on mobility behavior in Germany over time. Wittwer and Hubrich analyzed changes and developments in mobility behavior drawing from national mobility surveys conducted from 2008 to 2013 and investigated topics such as car availability, ur-

[121] Cf. Frondel / Vance (2018), p. 990.
[122] Cf. Buehler / Kunert (2010), p. 17.
[123] Cf. Buehler / Kunert (2010), p. 19.
[124] Eisenmann / Buehler (2018), p. 171.
[125] Cf. Eisenmann / Buehler (2018), p. 176.
[126] Cf. Best / Lanzendorf (2005), p. 120.
[127] Cf. Frondel / Vance (2018), p. 989.

ban mobility and differences between younger and older inhabitants. Kuhnimhof et al. studied changes in young adults' travel and mobility patterns in Germany from 1976 to 2007. Wittwer and Hubrich found that car availability showed an increase of 7.6 % from 2008 to 2013, even though car availability among 18 to 35 year-olds had declined. However, car availability of people over 60 had increased by 10 % in the same time span and had overtaken young adults' car availability in 2010.[128] In terms of the young generation, they found that they used car-sharing options on average every one to two months, had higher bike usage and were more regular public transport users than older generations.[129] Kuhnimhof et al. came to similar results. While car use increased for all ages until 2000 in Germany, it has decreased among young adults since then. Identified reasons included the higher use of public transport and multimodal travel behavior as well as the fading of gender differences in automobile use, which was characterized earlier by higher car ownership and use of young men.[130]

A topic that has received more attention in recent years is the introduction of battery electric vehicles and their acceptance. Letmathe and Suares developed total cost of ownership models for different vehicle types in order to test the competitiveness of battery vehicles[131] and Jakobsson et al. investigated the suitability of multi-car households for the use of electric vehicles based on data from 2010.[132] Letmathe and Suares computed the total cost of ownership for ten different electric and hybrid cars, as well as for internal combustion engine vehicles in the same price range. Results were confirmed by multiple scenario simulations and indicated that only a minimal number of electric vehicles were economically competitive without, but also with governmental subsidies. Especially the small vehicle segments proved to be uneconomical.[133] Jakobsson et al. studied different driving patterns in multi-car households in Germany and Sweden to evaluate suitability for battery electric vehicles and found that second cars offer a better fit for battery vehicle adoption than first cars or single-household cars as driving patterns were more stable and long-distance driving happened less frequently. For 70 % of second cars, the required range to cover complete driving need was only 220 km, for 70% of first cars it was 390 km. In further steps, they, like Letmathe and Suares, took the total cost of ownership of battery electric vehicles into account, which, in contrast, decreased the difference in fit for battery vehicle adoption between first and second cars, especially in Germany. Nevertheless,

[128] Cf. Wittwer / Hubrich (2016), p. 4308.
[129] Cf. Wittwer / Hubrich (2016), p. 4313.
[130] Cf. Kuhnimhof et al. (2012), p. 443.
[131] Cf. Letmathe / Suares (2017), p. 317.
[132] Cf. Jakobsson et al. (2016), p. 3.
[133] Cf. Letmathe / Suares (2017), p. 314.

also in this scenario, second cars showed a better fit for battery electric vehicles than first cars.[134]

On the whole, existing results regarding car ownership in Germany confirm the main influence factors that were found in other countries, presented in section 2.2. Beside household income, public transport availability, the number of licensed drivers and household size, the density and diversity of the environment seem to influence car ownership decisions. However, most studies have been conducted based on data up to the year 2009 and not on recent surveys. Rising traffic and environmental concerns, but also relatively new mobility options such as car-sharing, might have changed car ownership behavior to some extent over the last decade, reinforcing the research need covered by this thesis. Research in other fields shows the interrelation of transportation research topics, as well as current trends that may shift mobility behavior in the future, such as the use of battery electric vehicles.

4 Artifical Intelligence in Transportation Research

Regarding the employed methodology, most of the above-mentioned vehicle ownership literature and general literature in transportation research follows a statistical approach using either ordered or unordered response mechanisms. In recent times, however, the application of ML has been expanding from the use in information technology to other fields, such as transportation research.

Until the early 2000s, traditional statistical methods based on random utility theory were considered as the main tool in academic research. They had proven their accuracy over a long time horizon and provided the advantage of easy interpretation of model coefficients. ML models, on the other hand, tend to represent a so-called “black box” where model parameters are difficult to interpret directly.[135] Since then, several studies have examined and compared the predictive power of ML algorithms and traditional models. Early applications in the field focused on mode choice simulation and mainly used aggregated data. Nijkamp et al. compared neural networks (NNs) and aggregate logit models regarding their ability to forecast the modal split between train and road transportation flows. They found that multi-layer feed-forward NNs did not offer significantly higher effectiveness in predicting modal split.[136] Further studies confirmed that the forecasting abilities of NN and random utility models were similar with regard to aggregated

[134] Cf. Jakobsson et al. (2016), pp. 13-14.
[135] Cf. Cantarella / De Luca (2005), p. 124.
[136] Cf. Nijkamp et al. (1996), p. 337.

data.[137]

Later work, however, extended the use of ML algorithms to disaggregate mode choice data and contrasted their performance with random choice models. Hensher and Ton compared the predictive power of artificial NNs and nested logit models regarding commuter mode choice. Findings did not clearly indicate a superior approach.[138] Nevertheless, equally analyzing the prediction of mode share, Cantarella et al. found that the multi-layer feed-forward NN outperformed random utility models and served as an indicator of best obtainable performance.[139] A further study regarding mode share prediction that used another ML algorithm, support vector machines (SVM), found that the SVM outperformed an MNL with regard to both fitting and testing results, and an NN with regard to testing results in several scenarios.[140] Lastly, Hagenauer and Helbich, as well as Cheng et al., implemented a RF approach. Hagenauer and Helbich found the RF to be the superior method to predict mode shares among six others, including MNL, SVM and NNs.[141] Cheng et al. compared the RF approach to two other ML models, gradient boosting and SVM, and an MNL. They found that the RF outperformed all other models in terms of prediction ability and emphasized that the RF, compared to the other ML algorithms, offered the possibility to compute relative importance measures for variables.[142]

Beside the field of mode choice prediction, a lower amount of studies exist in other transportation research areas such as traffic forecasting and accident analysis. In traffic forecasting, Lingras and Adano, analyzing the prediction of average and peak traffic volumes, found that multiple regression models and NNs produced similar errors but that NNs were better in generating predictions when the data is insufficient for regression models.[143] Similar results are found by McFadden when predicting operating speed on highways. Artificial NNs and regression models exhibited comparable predictive power, but artificial NNs were able to overcome the limitations of regression models.[144] Nevertheless, study results in fields like incident detection and accident type classification problems reinforced the better performance of NNs in contrast to logit models.[145]

[137] Cf. Reggiani / Tritapepe (2000), p. 111; cf. Schintler / Olurotimi (2019), p. 131.
[138] Cf. Hensher / Ton (2000), p. 171.
[139] Cf. Cantarella / De Luca (2005), pp. 153-154.
[140] Cf. Zhang / Xie (2008), p. 141.
[141] Cf. Hagenauer / Helbich (2017), p. 273.
[142] Cf. Cheng et al. (2019), pp. 7-8.
[143] Cf. Lingras / Adamo (1996), p. 306.
[144] Cf. Mcfadden et al. (2001), p. 17.
[145] Cf. Hashemi et al. (1995), p. 254; cf. Ivan / Sethi (1998), p. 336; cf. Khan / Ritchie (1998), p. 311.

Karlaftis and Vlahogianni conducted a comprehensive review of studies that compared statistical and ML models in transportation research, including fields beyond those mentioned here, such as infrastructure maintenance and environment pollution.[146] Unexpectedly, in their review and others, we have only found two studies that compared the performance of ML algorithms and random utility models for predicting household vehicle ownership decisions. First, Mohammadian and Miller examined the predictive capability of a NN and a nested logit model with regard to predicting household vehicle choice in terms of vehicle classes (e.g., subcompact, compact, mid-size, large, special purpose vehicle, or van) in Canada. They found that the NN displayed superior predictive power of 70.38 % of correct cases compared to 49.20 % correct cases for the nested logit model.[147] The second study was conducted by Ha et al., who analyzed predictive capability regarding the number of household vehicles in Cambodia. Results demonstrate the superiority of the ML algorithms used, a NN and a RF algorithm, compared to an MNL regarding both accuracy and the control of unbalanced alternatives. Which algorithm prevailed regarding prediction performance was dependent on the number of outcome categories used.[148] It should be noted that both of these studies conducted their analysis based on data collected in one city, not via a nationwide survey. Moreover, even though the latter study by Ha et al. focused on ownership levels, the authors predicted vehicle ownership, including cars, motor- and regular bikes, not car ownership specifically.

In a variety of transportation research topics, ML algorithms proved their superiority with regard to predictive performance. This paper tries to provide further evidence regarding the comparison of predictive power between the MNL model and a RF algorithm in the field of household vehicle choice. We try to close the research gap regarding ML performance in car ownership prediction drawing on a nationwide survey in a country with high car ownership rates.

5 Data and Methodology

The following chapter will report on the data source used in this thesis. Furthermore, it seeks to present the methodology applied to answer each of the proposed research questions.

5.1 Data Assembly

The data set used is taken from the study ”Mobilität in Deutschland” (MID) conducted in 2017, a mobility survey financed by the German Federal Ministry of Transport and Digital Infras-

[146] Cf. Karlaftis / Vlahogianni (2011), pp. 391-394.
[147] Cf. Mohammadian / Miller (2002), p. 99.
[148] Cf. Ha et al. (2019), pp. 83-84.

tructure. After MID 2002 and MID 2008, MID 2017 constitutes the third edition of the study. The aim of the survey is the collection of representative data covering socio-economic information and moblity-related behavior for households in Germany. It is, in addition to other transportation surveys, used for transportation planning and supplies quantitative information to support political decision making. The overall sample is constructed by randomly selecting households from registration office data, as well as mobile and fixed line data. Throughout one year, selected households are surveyed regarding their mobility behavior in two steps. First, a household interview is conducted and general household mobility data is surveyed. In a second step, all household members are questionened regarding the trips on a given day, including their purpose, length and transportation modes used. The data collection of the most recent study took place during 12 months between May 2016 and September 2017.[149] Participants were able to complete the survey in a written form, online, or via phone. A total number of 2,613,880 households were contacted, with a total return rate of 6 %.[150]

The data is available in various forms, distinguished by a different number and degree of detail of included sociodemographic and spatial variables, due to data protection. We use the standard data set that is characterized by the highest degree of detail in general, however, geospatial data is given only in terms of federal state and aggregated measures, such as region types. The MID survey comprises multiple data sources featuring different observational units such as households, persons, cars, trips, stages of trips and journeys, with the household data set being the most aggregated set. In this thesis the datasets on households and persons are used, as they aggregate all necessary data for the analysis. The original data set includes 156,420 households, of which 33,389 belong to the nationwide basis survey and 123,031 to integrated regional collections.[151] After removing households with missing or ambiguous data for relevant variables, the final data set contains 77,779 household observations. The dependent variable is constructed by grouping the variable indicating the number of cars owned per household into four groups: households owning zero, one, two or three and more cars. The corresponding shares of households in the aforementioned groups are 11%, 50%, 31% and 8%.

5.2 Evaluating Feature Effects on Car Ownership

As a first research objective, this study seeks to analyze which attributes influence car ownership in Germany and to which extent. This chapter aims to explain the choice of model made to

[149] Cf. Bundesministerium für Verkehr und Infrastruktur (2019b), pp. 15-16.
[150] Cf. Bundesministerium für Verkehr und Infrastruktur (2019b), p. 30.
[151] Cf. Bundesministerium für Verkehr und Infrastruktur (2019a), p. 5.

examine car ownership patterns and present its underlying theory. Focus is placed on a critical review of the assumptions implied by the model, the model's derivation and the interpretation of its results. Furthermore, an overview of the set of variables chosen to be included is presented.

5.2.1 Model Selection and Further Considerations

As presented in Chapter 2.1, the application of disaggregated models is prevalent in recent literature on car ownership. There are several disaggregate limited dependent variable models of both ordered and unordered form that have been used in previous literature in the field.[152] Exemplary, the ordered logit is a form of ordered-response mechanism, while the MNL represents an unordered-response model. Ordered-response mechanisms found on the hypothesis that observed car ownership levels are related to an underlying, continuous variable representing car ownership propensity. Each level of car ownership is associated with an "ordered partition of the real line"[153], separated by threshold values. For the unordered response mechanism, on the contrary, the discrete outcomes are not assumed to relate to a latent variable but rather to follow random utility maximization. Random utility theory suggests that each car ownership level is associated with a specific utility value and the level with the highest utility is chosen. Transferred to the econometric specifications of both mechanisms, ordered response, thus, assumes a continuous probability density function for the random component of car ownership propensity. In contrast, unordered response assumes a continuous multivariate probability density function for random components of utilities across all discrete levels of car ownership.[154] Bhat and Pulugurta found that the unordered model outperformed an ordered model across all four datasets analyzed and for all different measures of model fit. They concluded that the unordered response mechanism had a better ability to represent households car ownership decision.[155] Since then, a large amount of studies on car ownership presented in Chapters 2.2 used the MNL. Thus, to determine the influence of chosen attributes on car ownership in Germany, this study utilizes an MNL model.

The MNL is the simplest discrete choice model, as it does not assume normality, linearity or homoscedasticity, which makes it a popular model for different analyses. However, the MNL has properties that give rise to certain concerns, which should be addressed in this section. The model assumes independence of irrelevant alternatives, which implies that, as explained

[152] Cf. Chu (2002), p. 66; cf. Zegras (2010), p. 1812.
[153] Bhat / Pulugurta (1998), p. 62.
[154] Cf. Bhat / Pulugurta (1998), p. 62.
[155] Cf. Bhat / Pulugurta (1998), pp. 73-74.

by Cheng and Long, the ”choice between two alternative outcomes is unaffected by what other choices are available”[156]. While this assumption is restrictive and makes the MNL unsuitable in different contexts, it can be neglected in the context of car ownership. This is confirmed by Long and Freese, and McFadden, who state that as long as outcome categories are distinct and not substitutive of each other, the MNL is suitable.[157] Moreover, the MNL presumes non-perfect separation, meaning that outcomes are not perfectly separable by predictors,[158] however, this is not a concern in the used data set as no perfect separation by one of the independent variables is present. A further concern is directed towards endogeneity. The explanatory variables included in the model cover a broad range of potential influences on car ownership, but the possibility of a correlation with unobserved factors influencing general travel decisions cannot be completely eliminated. An example of such bias is the joint determination of car ownership choices (or, more general, travel behavior choices) with those regarding residential choice, also known as self-selection. Households that do not have or cannot afford a car choose to live in neighborhoods where a car is not necessary and vice versa.[159] The second source of endogeneity bias may be omitted variable bias, where unobserved factors, e.g., attitudes, that influence car ownership also correlate with explanatory variables, thereby distorting coefficients. If present, the resulting endogeneity biases would limit the causal interpretation of the model's coefficients. As Ritter and Vance, and Zegras, this paper does not correct potential endogeneity.[160] It, therefore, refrains from a causal interpretation of effects between residential attributes and car ownership and practices descriptive interpretation and the inference of mere association instead.

5.2.2 Theoretical Background of the Multinomial Logit Model

The MNL, compared to ordered response models, ignores the information about the ordinal nature of car ownership data. Instead, it uses the framework of random utility theory. Each household associates utility values with distinct car ownership levels and decides on the alternative with the highest utility. Hence, households consider a set of J elements representing different levels of car ownership. Assuming perfectly informed households and rational decision making, random utility maximization implies that the utility of an alternative consists of a

[156] Cheng / Long (2007), p. 584.
[157] Cf. Long / Freese (2006), p. 191; cf. McFadden (1974), p. 113.
[158] Cf. Starkweather / Moske (2011), p. 1.
[159] Cf. Mokhtarian et al. (2009), p. 389.
[160] Cf. Ritter / Vance (2013), p. 76; cf. Zegras (2010), p. 1797.

deterministic and a random component:[161]

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n}$$
$$with \quad V_{j,n} = \beta_{j,n} * X_n,$$

where Utility $U_{j,n}$ is the true utility of household n for car ownership level j out of J, $\beta_{j,n}$ is a parameter vector mapping household attributes to total utility and $\varepsilon_{j,n}$ is a random term capturing unobserved utility and random components. The vector X_n represents household characteristics. A household n will choose outcome $m \in J$ only if:[162]

$$U_{m,n} > U_{j,n}, \quad \forall j \neq m, j \in J.$$

According to Ben-Akiva and Lerman, utility maximization therefore implies that the probability P that household n chooses car ownership level $m \in J$ is defined by:[163]

$$\begin{aligned} P_n(m) &= Pr(U_{m,n} > U_{j,n}), && \forall j \neq m, j \in J \\ &= Pr(V_{m,n} + \varepsilon_{m,n} \geq V_{j,n} + \varepsilon_{j,n}), && \forall j \neq m, j \in J \\ &= Pr(\varepsilon_{m,n} \leq V_{m,n} - Vj, n + \varepsilon_{m,n}), && \forall j \neq m, j \in J \end{aligned}$$

The MNL assumes error terms identically and independently distributed as a Gumbel distribution with a scale parameter $\mu > 0$ and returns the following predicted choice probabilities for owning a number of cars m equal to:[164]

$$P_n(m) = \frac{exp(V_{m,n})}{\sum_{j=1}^{J} exp(V_{j,n})}$$

The interpretation of the MNL is not as straightforward as linear regression or binary logistic regression. For a dependent variable with J levels, the MNL returns $J - 1$ models, resulting in a coefficient for every explanatory variable for each value of the dependent variable relative to its base category. For numerical variables, the coefficients represent the change in log odds for car ownership outcome i relative to the reference group for a unit increase in the corresponding variable, assuming the other variables remain constant. In our case, the dependent variable has four levels and the reference group is the ownership of zero cars, therefore returning three coefficients for each explanatory numerical variable. For categorical variables with N levels, on the other hand, $(N - 1) * (J - 1)$ coefficients are returned, one for every level except the variables base category. To give an example, the variable "region" consists of three levels: "urban",

[161] Cf. Ben-Akiva / Lerman (1985), p. 57.
[162] Cf. Ben-Akiva / Lerman (1985), p. 47.
[163] Cf. Ben-Akiva / Lerman (1985), p. 101.
[164] Cf. Ben-Akiva / Lerman (1985), p. 104.

”suburban” and ”rural”, with “urban” being the base category. A total number of six coefficients is returned. The coefficients represent the change in log odds for car ownership outcome i (owning one car, two cars or three or more cars) relative to owning zero cars for a change in the region variable from urban to suburban or from urban to rural, respectively.

By exponentiating the logit coefficients, the relative probabilities or odds ratios can be derived, representing the change in odds for outcome i relative to owning zero cars for a unit increase of the corresponding variable, or for a change in the level of categorical variables, respectively. Odds ratios greater than one correspond to a rise in the odds of belonging to the outcome category relative to owning zero cars. In contrast, ratios less than one correspond to a decrease in odds. However, these ratios can only provide insights on the relative probability of being in an outcome category relative to the base category. The interpretation are of the form that, e.g., for a unit increase in the variable x, the relative probability of owning one car is 10 % higher than the likelihood of owning zero cars. The actual likelihood of being in an outcome category, e.g., the change in probability of owning one car for a unit increase in the variable x, is not reported. To analyze the effect of a unit increase (or the change in a categorical variable respectively) on actual probability, average marginal effects (AMEs) are computed. Marginal effects report ”the rate at which the outcome y changes at a given point in covariate space, with respect to one covariate dimension and holding all covariate values constant”[165]. AMEs are determined by computing marginal effects for every observation using partial derivates and subsequently averaging estimates. AMEs, therefore, provide insights about the average influence of variables on each outcome category.[166] Furthermore, effect plots are used to display the change in probability for all outcomes over the complete value span of a single explanatory variable, that may be masked by the AME. Effect plots, introduced by Fox, can help in interpreting complex statistical models where an interpretation based on estimated coefficients only proves to be difficult.[167] This is the case for our model, as it features a logistic model with several outcomes and multiple higher-order terms. To construct effect plots, R’s effect package is used. To construct plots, covariates other than the analyzed attribute are assumed at typical values. Continuous variables are held at their average, and categorical variables are weighted according to the level’s proportion in the sample. The results depicted in plots therefore represent a certain combination of values and how a change in the value of the analyzed variable affects the probabilities for this value combination. Effects in plots therefore may differ from results generated from AMEs

[165] Leeper (2018), p. 7.
[166] Cf. Leeper (2018), pp. 7-8.
[167] Cf. Fox (1987), p. 348, cf. Fox / Hong (2009), p. 1.

that average effects for all existing households in the data and do not take into account different values of the covariate.

5.2.3 Definition of Explanatory Variables

The explanatory variables entered in the model are selected based on the review of related literature presented in Chapters 2.2 and 2.3 and the existing data limitations. The variables can be divided into three broad classes: socio-economic attributes, residential attributes and mobility-related attributes. An overview of variables and descriptive statistics can be found in Table 1.

As presented in Chapter 2.2, socio-economic characteristics of households may determine car ownership to a great extent. We, therefore, include a suite of socio-economic variables to estimate their influence. Household income strongly influenced car ownership levels in all previous studies. It is included as a numerical variable with values up to 9,000 EUR per month in total household income. The number of working adults and the number of license holders are, in addition to income, further features that were found to highly contribute to the explanation of car ownership in e.g. studies by Potoglou and Kanaroglou, as well as Whelan.[168] Based on the studies of Chu and Potoglou and Kanaroglou, who find different influences for the number of full-time workers, part-time workers and the number of license holders,[169] all three variables are included in the model. Part-time workers in our setting include all household members that work up to 35 hours a week, also including persons in marginal employment or internships. A binary attribute regarding the household's generation is included to determine possible effects of age or generation. In our model, a household is defined as belonging to a younger generation when more than half of its members belong to the generations Y (born between 1981 and 1995) or Z (born between 1996 and 2010). On the contrary, it is defined as belonging to an older generation when at least half of its members belong to generation X (1966-1980), the babyboomer generation (1956 to 1965) or an even older generation.[170] The influence of age on car ownership has been investigated by Lavieri et al., Knittel and Murphy as well as Ritter and Vance with mixed results as presented in Chapter 2.2.1. For the existence of children, the previous literature mainly reported an increase in the probability to own one or two cars.[171] In order to verify the effect in this study, the last socio-economic variable indicates if the household includes children under 18. We did not include specific variables capturing the cost of

[168] Cf. Potoglou / Kanaroglou (2008b), p. 50; cf. Whelan (2007), p. 212.
[169] Cf. Chu (2002), p. 65; cf. Potoglou / Kanaroglou (2008b), p. 50.
[170] Cf. Klaffke (2018), p. 1.
[171] Cf. Potoglou / Kanaroglou (2008b), p. 50; cf. Ryan / Han (1999), p. 6.

car ownership, such as purchase cost or maintenance cost in the model. However, according to Ryan, the intercept "for each alternative captures the average costs of ownership of the number of vehicles represented by that alternative"[172]. Ownership cost include purchase cost or leasing fees, as well as maintenance, insurance and fuel cost.

The most popular built-environment attributes included in previous studies are population density measures, public transportation availability and variables indicating land-use diversity. The dataset used does not include exact spatial data like street or population density for household neighborhoods, however, several variables indicating built-environment attributes are available. First, we include the quality of public services as well as public transport quality as computed by the MID survey. Both variables are rated on a scale of one to four, with one representing low quality and four representing excellent quality. The rating for public services has been developed by analyzing the distance of 21 million addresses to 13 different public services, namely, doctors, pharmacies, supermarkets, drug stores, shopping malls, grocery stores, banks, ATMs, hairdressers, corner shops, post offices, restaurants and fuel stations. The computation also included the average distance to public services per municipality type to allow comparisons among the same municipality type. The rating for public transport, on the other hand, has been computed by taking into account the total number and average quality of public transport stops of different types in a distance of one kilometer to the household's residence. The evaluation of a stop is made based on the location of the stop, using parameters such as street type, reachability and location inside or outside of a municipality, as well as on a density measure that considers the number of stops in the corresponding geographical grid cell.[173] Additionally to the quality of public transport, the distance in kilometers to the next train, bus and metro stations with 28 departures on workdays is inserted. The variable is categorized into four levels, indicating distances less than 0.5 km (1), between 0.5 and 2.5 km (2), between 2.5 and 5 km (3) and distances greater than 5 km (4). The inclusion is believed to offer information beyond that of the general public transport rating and allows the analysis of effects among different means of public transport.

As a proxy for population density and the amount of open space, we include an attribute stating if the household lives in an urban, suburban or rural environment. The variable is constructed by aggregating the region statistical type "RegioStaR 7", defined by the Federal Office for Building and Regional Planning as follows. The urban category includes metropolises, regiopolises and

[172] Ryan / Han (1999), p. 2.
[173] Cf. Bundesministerium für Verkehr und Infrastruktur (2019c), pp. 7-8.

large cities. The sub-urban category is formed by middle-sized towns and suburban areas in urbanized and rural regions as well as central cities in rural regions. Lastly, the rural category is made up of provincial and rural areas.[174] The last two residential attributes are housing type and a garage dummy. As seen in a study by Bhat and Pulugurta, a housing type variable can help to add information about the close neighborhood of a household as the other residential attributes reflect a more aggregated scale.[175] Housing types encompassed are single-family houses, multi-family houses and apartmentbuildings. The influence of the availability of a garage on car ownership has not been addressed by past literature, however, mostly because of data limitations and is, thus, included in the model.

The last category includes mobility-related attributes of the household. The first two variables are concerned with the average number of trips per household on a given day and the average trip length. To guarantee comparability of these variables, households with reporting dates of trips on weekends or holidays were excluded, as those days usually show different trip patterns than regular workdays. Average trip length is constructed by dividing the sum of the total daily trip length of all household members by the total number of trips per household. Moreover, the amount of motorbikes and regular bikes per household are included, which has been shown to reduce car ownership in emerging countries, however, the influence in developed countries is yet to be determined. The last variable in this category is concerned with the use of carsharing and has not been considered by most studies. As shown in a recent study by Kim et al., carsharing may reduce car ownership or prevent the purchase of further cars and is therefore considered in the model.[176] The variable features three levels: households with one carsharing subscription, households with several carsharing subscriptions and households that do not use carsharing.

Before analyzing, the data was prepared. All analyses and calculations are conducted using the statistical software R. As the MNL is based on maximum likelihood estimation, a large number of observations is needed to guarantee a stable model. With a higher number of covariates, the number of necessary observations increases. Long and Freese propose to have at least ten observations for every coefficient estimated.[177] We therefore applied this rule on all variables with a limited number of categories and removed or merged levels with too few observations. Apart from that, outliers in a dataset can distort subsequent regression results and lead to errors and

[174] For further information refer to Bundesministerium für Verkehr und digitale Infrastruktur (2020).
[175] Cf. Bhat / Pulugurta (1998), pp. 68-69.
[176] Cf. Kim et al. (2019), p. 136.
[177] Cf. Long / Freese (2006), p. 65.

overestimated coefficients. For real continuous variables, outlier removal according to the out-

Variable	Description	Mean[1]	Min	Max
cars	Number of cars	1.371	0	5
licenses	Number of Licenses	1.808	0.0	4.0
income	Monthly household income in thousand Euro	3.471	0.0	9.0
full-time	Number of full-time workers	0.664	0.0	3.0
motorbikes	Number of motorbikes / mopeds	0.190	0.0	4.0
av.triplength	Average trip length in kilometers	7.500	0.0	33.1
part-time	Number of part-time workers	0.294	0.0	3.0
bikes	Number of bikes / pedelecs	2.212	0.0	10.0
av.trips	Average number of daily trips	3.015	0.0	8.3
generation	1 if categorized as older generation	0.914	0.0	1.0
children	1 if household has children	0.193	0.0	1.0
garage	1 if a garage is available	0.075	0.0	1.0
urban	1 if households lives in an urban environment (basecategory)	0.374	0.0	1.0
suburban	1 if households lives in an suburban environment	0.439	0.0	1.0
rural	1 if households lives in an rural environment	0.186	0.0	1.0
ps.1	1 if quality of public services: 1 (basecategory)	0.078	0.0	1.0
ps.2	1 if quality of public services: 2	0.390	0.0	1.0
ps.3	1 if quality of public services: 3	0.389	0.0	1.0
ps.4	1 if quality of public services: 4	0.143	0.0	1.0
no.carsharing	1 if carsharing is not available	0.952	0.0	1.0
carsharing	1 if carsharing available	0.035	0.0	1.0
carsharing.m	1 if multiple carsharing options available	0.012	0.0	1.0
singlefamily	1 if residence is a singlefamily-Home (basecategory)	0.573	0.0	1.0
multifamily	1 if residence is a multifamily-Home	0.293	0.0	1.0
apartment	1 if residence is an apartmentbuilding	0.036	0.0	1.0
other	1 if residence is another housing type	0.099	0.0	1.0
pt.1	1 if quality of public transport: 1 (basecategory)	0.065	0.0	1.0
pt.2	1 if quality of public transport: 2	0.507	0.0	1.0
pt.3	1 if quality of public transport: 3	0.342	0.0	1.0
pt.4	1 if quality of public transport: 4	0.086	0.0	1.0
metro.1	1 if nearest metro/tram station: air-line distance below 500m (basecategory)	0.158	0.0	1.0
metro.2	1 if nearest metro/tram station: air-line distance between 500m to 2500m	0.141	0.0	1.0
metro.3	1 if nearest metro/tram station: air-line distance between 2500 to 5000m	0.063	0.0	1.0
metro.4	1 if nearest metro/tram station: air-line distance above 5000m	0.637	0.0	1.0
train.1	1 if nearest train station: air-line distance below 500m (basecategory)	0.110	0.0	1.0
train.2	1 if nearest train station: air-line distance between 500m to 2500m	0.548	0.0	1.0
train.3	1 if nearest train station: air-line distance between 2500 to 5000m	0.195	0.0	1.0
train.4	1 if nearest train station: air-line distance above 5000m	0.146	0.0	1.0
bus.1	1 if nearest bus stop: air-line distance below 500m (basecategory)	0.765	0.0	1.0
bus.2	1 if nearest bus stop: air-line distance between 500m to 2500m	0.154	0.0	1.0
bus.3	1 if nearest bus stop: air-line distance between 2500 to 5000m	0.050	0.0	1.0
bus.4	1 if nearest bus stop: air-line distance above 5000m	0.031	0.0	1.0

[1] Mean for categorical variables represents the categories percentual share

Table 1: Variables description and statistics[178]

lier labeling rule was applied to avoid highly influential points. In the outlier labeling rule, the interquartile range of a variable is multiplied by a parameter and then added to the third quartile and substracted from the first quartile. Observations outside of the generated bounds are labeled as outliers. Earlier research papers used a parameter of 1.5 for the outlier labeling rule, but sub-

[178] Own analysis.

sequent studies suggest that 2.2 be a more accurate representation,[179] which is, thus, used in this study. Subsequently, as the MNL cannot handle ordinal variables by definition, transformations to either continuous or categorical variables were applied. Treating ordinal variables as continuous bears two advantages. First, the interpretation becomes easier and second, the ordinal information inherent in the variable is not lost. According to Long, ordinary variables can be treated as continuous as long as "successive categories of the ordinal independent variable are equally spaced"[180]. This applies to all variables referring to countable objects, such as the number of licenses, the number of full- and part-time workers and the number of bikes and motorbikes.

The MNL is a form of logistic regression and must therefore comply with the basic assumptions of logistic regression. These are the linearity in the logit for continuous independent variables, as well as the absence of multicollinearity and strongly influential points.[181] To limit potential multicollinearity among explanatory variables in the MNL, several analyses of correlation were performed. First, the correlation between numerical and ordinal variables was examined using Spearman's correlation coefficient, results are reported in Appendix 1. Spearman's correlation coefficient was chosen as it is not only appropriate for numerical variables but also extends to an ordinal scale and does not assume a normal distribution of variables like Pearson's correlation coefficient.[182] Correlations up to a value of 0.6 (or -0.6, respectively) are considered as moderate,[183] therefore, variables that exhibit higher correlations are excluded. Secondly, to further control for correlations among categorical variables, among categorical and ordinal variables and among categorical and numerical variables the generalized variance inflation factor (GVIF) was used. The variance inflation factor (VIF) is an accepted way to measure multicollinearity and signals high multicollinearity for a value exceeding five.[184] For models featuring factor variables requiring more than one coefficient and therefore more than one degree of freedom, the GVIF, introduced by Fox and Monette, is reported, which takes into account the resulting different number of model coefficients and higher number of degrees of freedom:[185]

$$GVIF_i = \frac{det(R_{11}) * det(R_{22})}{det(R)},$$

where R_{11} is the correlation matrix for all levels of variable i, R_{22} is the correlation matrix

[179] Cf. Hoaglin / Iglewicz (1987), p. 1149.
[180] Long / Freese (2006), p. 269.
[181] Cf. Stoltzfus (2011), p. 1101.
[182] Cf. Hauke / Kossowski (2011), pp. 88-89.
[183] Cf. Akoglu (2018), p. 92.
[184] Cf. Hair et al. (2010), p. 200.
[185] Cf. Fox / Monette (1992), p. 180.

for all remaining variables excluding the intercept and R is the correlation matrix for all variables excluding the intercept. For continuous variables, the GVIF is the same as the VIF. To compare the GVIF across categorical variables with a distinct amount of levels, the following modification is recommended:[186]

$$GVIF_{mod_i} = GVIF_i^{1/(2*DF)},$$

with DF being the degrees of freedom of variable i, therefore the number of produced coefficients for this variable. The modified GVIF represents a linear measure that can be evaluated taking the square root of the regular VIF evaluation, thus, $GVIF_{mod}$ should not exceed $\sqrt{5}$. VIF and GVIF values for predictor variables are shown in Appendix 2. While household size was used in previous studies, it displayed high correlation coefficients with several variables, as well as the highest GVIF and was, thus, removed from the model.

After testing for multicollinearity between predictors, we use a forward selection algorithm to formulate a preliminary model. The forward selection process for the preliminary model is summarized in Appendix 3. The starting model includes zero variables and further variables are incrementally selected and added based on the highest contribution to model fit according to the Akaike Information Criterion (AIC). The AIC explains how close the fitted values are to the true expected values by returning the expected distance between the two.[187] All variables are included into the model by the algorithm, indicating that they improve model fit.

Next, we test the linearity assumption in the logit for continuous variables. This means, in logistic regression, there should be a linear relation between continous predictors and the logit-transformed outcomes. Begg and Gray propose the approximation of multinomial fit by fitting separate binary models and show that the returned coefficients are consistent with those returned by the MNL and can thus be used to assess general model fit.[188] This approach can be used to detect necessary transformations for continuous independent variables and influential observatios when fractional polynomial analysis is not supported. In our case, where the dependent variable consists of four groups, we fit a model for zero cars versus one car, zero cars versus two cars and zero cars versus three or more cars, using the standard package glm for logistic regression in R. Coefficient estimates are shown in Appendix 4.

For continuous variables, possible nonlinear patterns were detected using scatter plots between

[186] Cf. Fox / Monette (1992), p. 180.
[187] Cf. El-Habil (2012), p. 284.
[188] Cf. Begg / Gray (1984), p. 17.

predictors and logit values as well as binned residual plots as proposed by Gelman.[189] We add a quadratic term to the number of trips, the number of licenses and the number of part-time workers and transform the number of motorbikes to its logit to account for detected nonlinear patterns. As the latter variable contains zero values, we add a small constant of 0.0001 to zero-observations only, as suggested by Hu,[190] to enable logit-transformations. Observed plots before and after transformations for the three logit models are included in Appendices 5 to 10. Moreover, binned plots showing the structure of residuals for the separate logit models before and after the inclusion of nonlinear terms are shown in Appendices 11 to 13. All nonlinear terms inserted are chosen when using the forward selection algorithm again (see Appendix 14). The resulting model capturing nonlinear effects performed better than the preliminary model without nonlinear terms as confirmed by a significant likelihood-ratio test and a smaller AIC, as presented in Table 2. We inspect the existence of highly influential points as extreme data points may alter the quality of a logistic regression model. Influential points are found by looking at observations with high standardized residuals with an absolute value exceeding three.[191] We inspect our separate logistic models for influential points and remove 70 observations. The resulting MNL model is used for the analysis of influences on car ownership levels in Germany.

	χ^2	DF[1]	p-value
Likelihood-ratio test	3677	12	< 0.0001
AIC: Preliminary model[2]	105236		
AIC: Final model	101583		

[1] Degrees of freedom.

[2] AIC shows a different value than in the forward selection algorithm as we removed highly influential points.

Table 2: Model validation measures for preliminary and final fit[192]

5.3 Evaluating the Influence of Regional Environment on Feature Effects

The second part of this paper examines the influence of the regional environment in which a household lives (represented in our model by the categories ”urban”, ”suburban” and ”rural”) on the magnitude of other covariates in the model. To study the influence of one variable on the effect of other variables in the model, interaction terms are used. Interaction terms are widely used in studies when it is assumed that an independent variable is influenced by the

[189] Cf. Gelman / Hill (2006), pp. 97-98.
[190] Cf. Hu (1972), p. 90.
[191] Cf. Gray / Woodall (1994), p. 111.
[192] Own analysis.

value of a different independent variable.[193] Therefore, a second MNL model is estimated, featuring additional variables that capture interaction effects between the variable ”region” and the other variables. Interaction terms are included and estimated in a second model for several reasons. The main reason is that once interaction effects are added, the main effects included in the interaction are not interpretable on a stand-alone basis anymore and the interpretation becomes primarily about significant interactions.[194] Secondly, none of the other car ownership studies presented included interaction terms as all of them focused on main effects. To be able to compare results we have chosen to analyze main effects in a first step and then focus on the interaction effects with the variable region in a second step. Unmeasured nonlinear terms that are not significantly distorting to main effect coefficients may distort interaction term coefficient and their significance as both examine nonadditive effects. Thus, we include squared terms of all continuous variables as covariates in the model, as recommended by Cortina.[195] Furthermore, we have chosen not to center continuous variables before the inclusion of interactions as all continuous variables possess meaningful zero-points, as suggested by Dalal and Zickar.[196]

The cost of additional interaction terms is only one degree of freedom and additional interaction terms, even if not significant, do not harm results.[197] To be able to analyze the entirety of possible effects, we add all possible interaction terms other than those for categorical variables that do not possess observations for a certain level in a certain region. These include the variables capturing multiple carsharing subscriptions, the distance to metro and bus stations and the quality of public transport. For example, regarding households in rural areas, our dataset does not comprise observations that report the highest rating of public transport quality and additionally do not own cars. Consequently, the estimation of coefficients for these levels would cause the model to be biased and unstable. Thus, no interaction terms are added for these attributes. As stated by Norton et al., when looking at interaction terms in a logistic model, the term's significance cannot be correctly deduced from the z-statistic in the regression output.[198] Thus, we analyze differences between regional environments for different values of variables and test for significance via confidence intervals using effect plots. Secondly, we examine interactions via a second measure, as suggested by Berry et al.[199] We compute AMEs for all three regions for all variables containing interactions and analyze existing differences. For more information

[193] Cf. Long / Freese (2006), p. 271.
[194] Cf. Williams (2015), p. 4.
[195] Cf. Cortina (1993), pp. 915-917.
[196] Cf. Dalal / Zickar (2012), p. 339.
[197] Cf. Mize (2019), p. 97.
[198] Cf. Norton et al. (2004), p. 154.
[199] Cf. Berry et al. (2012), p. 17.

regarding effect plots and AMEs, compare Chapter 5.2.2.

5.4 Prediction of Car Ownership

This paper aims to contribute to the research regarding the prediction of car ownership patterns. It therefore compares the prediction performance between a ML algorithm and a traditional statistical model. The ML algorithm that is chosen for comparison against the MNL used in the first part of this paper is the RF, as it has shown superior prediction performance in other transportation-related studies (see Chapter 4). This section explains the generation of predictions for both models and how prediction performance and variable importance is measured.

5.4.1 Generating Predictions Using Different Classifiers

Statistical models are traditionally established to analyze the effects of certain predictor variables on the outcome category and to interpret these effects accordingly to derive insights about the relationship. However, they can also be used to generate predictions.[200] Predictions of outcomes are generated based on the calculated predicted probabilities. Based on the formula presented in Chapter 5.2.2 and the estimated model coefficients, predicted probabilities for the outcomes are generated: $P(1 : \text{One car})$, $P(2 : \text{Two cars})$ and $(P(3 : \text{Three or more cars}))$. The predicted probability for the base category $P(0 : \text{Zero cars})$, for which no model coefficients are present, is calculated by substracting the sum of the predicted probabilities of the other categories from one. The outcome category with the maximum value of predicted probability is selected as the predicition of car ownership for the corresponding household. If the predicted probability of "Three or more cars" was the greatest among the predicted probabilities, this household is predicted to own three or more cars.

ML algorithms, such as RF, are mainly estimated to generate predictions for new data and not to infer conclusion or insights about the relationship between input and output parameters. The model is trained on existing data and can then predict the outcome categories of new data. The basic model behind the RF method is a single decision tree. Decision trees recursively partition observations into outcome categories based on input variables. Therefore, a first splitting variable and its optimal splitting point are found by analyzing all possible splits and deciding on the variable and split that will create the most homogenous subgroups based on an impurity measure such as the gini index. Gini impurity measures the probability of incorrect classification for random elements in a set when labeled simply according to class distributions. For

[200] Cf. Cortese (2020), p. 1.

a splitting variable X_i and a possible number of outcome categories $O_1, ..., O_J$ gini impurity $G(X_i)$ is given as:[201]

$$G(X_i) = \sum_{j=1}^{J} P(X_i = O_j)(1 - P(X_i = O_j)),$$

where $P(X_i = O_j)$ is the probability of randomly selecting a datapoint of class O_j in the subset. Gini impurity is calculated for the original node and the nodes created by the different splits. The split is chosen that is defined by the highest gini gain, which is formed by the difference between original node impurity and the impurities of the branches created by the split, weighted by the share of data points they comprise. In further steps, other splitting variables are chosen and for every split, the number of observations in the generated regions becomes smaller. This process is repeated until a certain stopping condition, such as end-nodes that only contain one outcome category (”pure” nodes), is met. However, one disadvantage of regular decision trees is their tendency to overfit the data. The RF algorithm is essentially a collection of decision trees and less prone to overfitting due to the combination of two main features. The first is bootstrapping, where each tree in the forest is trained with a randomly selected subsample of the training data created by sampling with replacement. This means that some rows of data are never used for training, they form the so-called ”out-of-bag” sample. Secondly, to avoid the correlation of trees inside the forest, the second feature of random feature selection is implemented. In contrast to regular decision trees, where all explanatory variables are available as splitting variables at each node, only a random subset is available for each tree in the forest.[202] Therefore, in the first step of the algorithm, for each tree, observations are bootstrapped and a subset of explanatory variables is randomly selected. In a second step, each tree is computed and split until its maximum depth. After all trees are computed, the model can be used to predict new data. Outcome categories for new observations are computed by all trees, subsequently majority voting is applied to determine the predicted outcome for each observation. By the use of this approach, prediction errors resulting from biased or noisy data are reduced.[203] An overview of the random forest method is given in Figure 1.

5.4.2 Tuning Process and Measurement of Prediction Performance

Our dataset was split into train and test set prior to model prediction for both MNL and RF, using a ratio of 80:20. The RF is tuned before the actual prediction, while the MNL, as a statistical model, does not have tuning parameters. The goal of the tuning process is to find parameters that

[201] Cf. Cheng et al. (2019), pp. 3-4.
[202] Cf. Breiman (2001), pp. 11-12.
[203] Cf. Breiman (2001), p. 29.

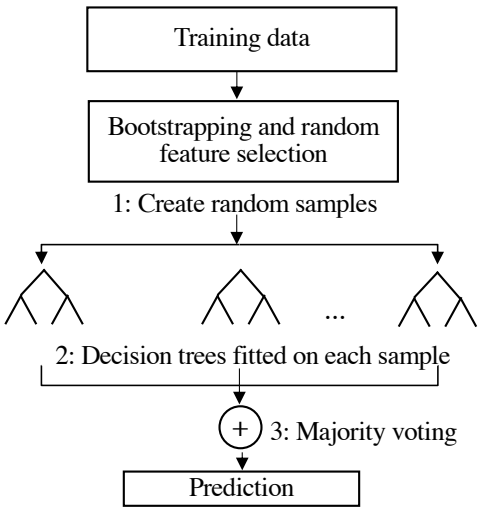


Figure 1: Operation mode of a random forest[204]

optimize the prediction performance of the model, while also avoiding overfitting the data. Traditional data science methods include k-fold cross-validation and bootstrap methods. However, Borra and Di Ciaccio and Kim find that repeated 10-fold cross-validation yields better results than bootstrapping.[205] We therefore apply repeated 10-fold cross-validation during the tuning process to find the optimal RF model. The resulting model with optimal parameters, as well as the MNL model, is then trained on the full train set before predictions are generated for the test set, that has not been used at all during the training process, also referred to as the validation set.

To evaluate and compare the predictions of the two approaches, we require a measure of correctly identified car ownership levels. Most studies will use the accuracy value that measures the fraction of true positives and true negatives among all predictions. Accuracy, though, may provide an overly optimistic measure when rating an imbalanced dataset.[206] As the number of observations belonging to each level of car ownership varies in our dataset, the prediction corresponds to an imbalanced classification task. We thus use macro-averaged balanced accuracy as well as Cohen's Kappa instead of accuracy, as these measures are more suitable to deal with imbalanced datasets.[207] In a multiclass setting like ours, balanced accuracy is calculated for every outcome category and is subsequently averaged. To compute, e.g., balanced accuracy of the outcome category of owning one car, two other measures, sensitivity and specificity are used. While sensitivity measures the fraction of correctly predicted households that own one car among all household observations owning one car, specificity measures the fraction of cases

[204] Own depiction based on Cheng et al. (2019), p. 4.
[205] Cf. Borra / Di Ciaccio (2010), p. 2988; cf. Kim (2009), p. 3744.
[206] Cf. Brodersen et al. (2010), p. 3121.
[207] Macro-averaged F1 was also tested, but did not lead to different results and is thus not included here.

correctly predicted as "not owning one car" among all observations of the categories zero, two and three and more cars. Balanced accuracy is calculated as the arithmetic mean of both:[208]

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity+Specificity}}{2},$$
$$\text{with Sensitivity} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}},\ \text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives+False Positives}}$$

This study uses the measure's macro-average, meaning that balanced accuracy will be estimated for each class and results will subsequently be averaged, with all classes having the same weight in the calculation of the average. Micro-averaging, on the other hand, takes into account the distribution of classes and computes a weighted average.[209] Therefore, the micro average evaluates the performance on large classes in the dataset, while the macro average evaluates the performance over all classes. According to Manning et al., the macro-average should be used to estimate the performance on smaller classes.[210] We, thus, decided to use the macro-average, as in the case of car ownership level prediction, each category has the same importance and should be determined correctly by the classifier, not only the class that is observed most often, e.g., owning one car in this study. Cohen's Kappa represents a measure of agreement between two raters, normalized by the agreement given by chance. For ML algorithms, one rater is based on the prediction, while the other rater represents actual observations. The agreement between the two raters is given by an index that is equal to one if there is perfect agreement. Using proportions, Fleiss et al. report the following equation based on the confusion matrix of the two raters:[211]

$$K = \frac{P_o - P_e}{1 - P_e} = \frac{\sum_{i=1}^{k} P_{ii} - \sum_{i=1}^{k} P_{i.} P_{.i}}{1 - \sum_{i=1}^{k} P_{i.} P_{.i}}$$

P_o are the proportions referring to the agreement between the two raters (the diagonal in the confusion matrix) for the categories 1 to k and P_e are the expected proportions of agreement given by chance. Moreover, $P_{i.}$ and $P_{.i}$ are the overall row and column proportions of the confusion matrix.[212] Kappa, thus, denotes how well the predictive model performs beyond guessing based on the distribution of the dataset.

5.4.3 Determining Variable Importance

To further determine the contribution of different variables on car ownership levels, we analyze variable importance for our MNL model and the used RF algorithm to analyze both results. For

[208] Cf. Brodersen et al. (2010), p. 3122.
[209] Cf. Van Asch (2013), pp. 4-6.
[210] Cf. Manning et al. (2008), p. 281.
[211] Cf. Fleiss et al. (2003), pp. 604-605.
[212] For an extensive example compare Fleiss et al., p. 605.

the MNL, the magnitude of statistical coefficients, while providing information on the general connection between outcome categories and independent variables, cannot be directly transferred to variable importance, as structures and measurements of variables differ. For example, the number of licensed drivers is measured on a numerical scale, while the quality of public transport is rated within the range of one to four. To overcome this issue, we follow other studies on car ownership research[213] and apply the standardized coefficients approach introduced by Menard.[214] A statistic similar to the standardized coefficient in ordinary least squares regression is obtained by computing the product of the variable's coefficients β_i and its standard deviation SD_i for all outcomes $1, .., O$ and summing the absolute values:

$$\beta_{standardized} = \sum_{o=1}^{O} |\beta_o * SD_o|$$

It should be noted that standardized coefficients are utilized only to compare predictor variables, as they offer a comparable measure of the relationship between predictors and the dependent variable when predictors are measured on different scales.[215] For coefficient interpretation, on the other hand, unstandardized coefficients are more appropriate, as a unit increase of "one standard deviation" is not intuitively interpretable for most of our variables, which are, thus, interpreted in their standard units and categories.

This paper uses mean decrease gini (MDG) to determine variable importance for the RF algorithm, following other research in the field.[216] Including all trees in the forest, a weighted mean of the improvement in gini impurity (see Chapter 5.4.1) for each splitting variable is computed. Accordingly, variables that create more homogenous subnodes have a greater effect on the outcome.[217] Another approach used by earlier research is mean decrease accuracy (MDA), where the importance of a variable is determined by the increase in error rate in the out-of-bag sample after its values have been randomly permuted. A higher increase in error signals higher variable importance. However, Breiman excluded this measure in his RF manual, as MDA becomes volatile when many predictor variables are present.[218] A study by Strobl indicates that MDA and MDG might be biased by differences in variable structure and results should, thus, be interpreted with caution.[219]

[213] Cf. Zegras (2010), p. 1803. cf. Ha et al. (2019), p. 76.
[214] Cf. Menard (2004), p. 219; cf. Menard (2011), p. 1416.
[215] Cf. Menard (2011), p. 1421.
[216] Cf. Cheng et al. (2019), p. 4; cf. Ha et al. (2019), p. 76.
[217] Cf. Charpentier (2015), p. 1.
[218] Cf. Breiman (2007), pp. 14-15.
[219] Cf. Strobl et al. (2007), p. 19.

6 Results and Interpretation

The following chapter will demonstrate the findings generated with regard to the presented research questions. First, Chapter 6.1 starts with an interpretation and comparison of the general impact of variables on car ownership levels in Germany. Next, in Chapter 6.2, these impacts are contrasted for urban, suburban and rural environments. Lastly, the results from testing the predictive performance of an MNL model and a RF algorithm when generating predictions for new data are presented and compared in Chapter 6.3.

6.1 Analysis of Car Ownership Levels in Germany

The following section will present results regarding the determinants of car ownership in Germany. Positive impacts are expected for variables that either reinforce the advantages of owning one or more cars or that increase the opportunity costs of using other modes of transport. We expect positive signs for the number of full- and part-time workers as a reflection of greater mobility needs, as well as for the number of driver's licenses since more licenses may increase conflicts regarding car use and lead to the acquisition of another car. Longer distances to bus, train or metro stations are also expected to have a positive effect on car ownership. Furthermore, households with children would be associated with a higher level of car ownership as children cause additional mobility needs. A higher income, the availability of a garage, a higher average trip length and a higher number of trips are hypothesized to have a positive influence on car ownership. Regarding the variable reflecting household composition in terms of age or generation, it is believed that an older age composition has a more positive impact on car ownership than a younger one, as younger people are thought to have greater environmental consciousness. Lastly, for the region attribute, it is supposed that rural and suburban regions have a more positive influence on car ownership due to a wider spread of recreational and necessary activities, while urban areas may have a negative effect due to higher population densities in cities. Adverse effects are accordingly expected for the number of motorbikes and bikes or pedelecs, as they offer potential to satisfy mobility needs. The same reasoning may be applied to carsharing membership as well as a higher quality of public transport, while a higher rating of public services at the place of residence is also hypothesized to reduce levels of car ownership. Moreover, car ownership may be less likely for households living in multi-family houses or apartment buildings as this is an indicator of limited parking space in the neighborhood. As the intercept represents the cost of car ownership, we consequently expect it to be negative and to increase with a higher number of cars.

6.1.1 Model Validation

The final model uses 19 explanatory variables. Since some of these variables are discrete or have added nonlinear terms, the actual number of estimated coefficients in the model is 37. Sample size guidelines for the MNL recommend at least 10 cases per estimated parameter,[220] the number of cases in the selected model is 1900. To validate the model, we apply overall model evaluations, statistical tests of each explanatory variable and goodness-of-fit measures which are reported in Table 3. Overall model evaluations include likelihood ratio tests, score tests and Wald tests, which are used to test the improvement of the model over the intercept-only model. Moreover, the AIC is compared between the two. The AIC explains how close the fitted values of a model are to the true expected values by returning the expected distance between the two.[221] As it does not contain predictors, the null model is an appropriate comparison model, because observations would be predicted to fall into the outcome category with the highest number of observations.[222] In all three tests the full model performed better than the constrained model with a p-value smaller than 0.0001, indicating that at least one independent variable is significant in predicting the level of household car ownership. The lower AIC value compared to the null model confirms the validity of the model. The individual model

Test	χ^2	DF[1]	p-value
Likelihood Ratio test	58479	111	<0.0001
Score test	23826	111	< 0.0001
Wald test	25668	111	<0.0001
Measure	Value		
AIC: intercept only	159.840		
AIC: Final	101.583		
Pseudo R-Squared	Value		
McFadden	0.3659		
Cox and Snell	0.5647		
Nagelkerke	0.6295		

[1] DF = degress of freedom

Table 3: Overall model evaluation[223]

coefficients were validated using z-tests, testing the hypothesis that the coefficient is zero and therefore not significant. All variables were significant, although some variables include levels of no significance, e.g. the variable indicating the distance to the next busstop. However, unsignificant levels of variables are kept in the model as categorical variables can only be included as a whole in order not to change the variable's reference level. Table 4 in the following section displays coefficients and their significance. Applied goodness-of-fit measures, which assess

[220] Cf. Hair et al. (2010), p. 318.
[221] Cf. El-Habil (2012), p. 284.
[222] Cf. Peng et al. (2002), p. 6.
[223] Own analysis.

model fit against actual observations, include R^2 measures and the Hosmer and Lemeshow (H-L) test. In ordinary least squares regression, R^2 represents the part of variance of the dependent variable explained by independent variables, however, in logistic regression this is not the case. Nevertheless, according to Peng, Lee and Ingersoll, the R^2 indices by Cox and Snell, Nagelkerke and McFadden provide a variation of the R^2 concept in linear regression but should only be used as additional indices to more useful ones, such as the HL-test.[224] Models with larger pseudo R^2 dominate models with smaller values.[225] McFadden states that, with regard to his index, values of pseudo R^2 from 0.2 to 0.4 indicate very good fit.[226] This in in line with the value reported for our model, as shown in Table 3. Regarding further methods for assessing model fit, the H-L test is widely used for multinomial and standard logistic models. The test is sensitive to larger sample sizes, as the "rejection of the null hypothesis may be caused by the large sample size rather than the true deviation between model and observation."[227] Thus, in this paper, we use the procedure by Lai and Liu, that propose the use of a bootstrap method to repeatedly estimate the H-L test under the standard sample size of 500 observations.[228] We do not reject the null hypothesis using the proposed method, which indicates good model fit.[229]

6.1.2 Model Results

Table 4 reports the estimated model coefficients representing log-odds. Corresponding odds ratios are presented in Appendix 15. The estimates show mostly the expected effects. The cost of car ownership represented by the intercept is negative and increases with the number of cars, which is in line with expectations. A higher income, a higher average trip length and average number of trips and a higher number of license holders is positively associated with car ownership of all levels, in comparison to owning no cars. Likewise, a higher number of full-time and part-time workers, a farther distance to bus, metro or trainstations, the availability of a garage and the existance of children and a more suburban or rural area increase the likelihood of owning one or more cars compared to owning no car, as expected. Conversely, a higher number of bikes, a higher share of younger people, better quality of public services and public transport, as well as housing types different from a single-family home reduce car ownership. A rather unexpected sign is found for a higher number of motorbikes. Additional motorbikes increase the likelihood of car ownership for all levels, which is contradictory to our hypothesized effects

[224] Cf. Peng et al. (2002), p. 6.
[225] Cf. El-Habil (2012), p. 282.
[226] Cf. McFadden (1977), p. 35.
[227] Lai / Liu (2018), p. 9.
[228] Cf. Lai / Liu (2018), pp. 3-4.
[229] The null hypothesis for the H-L test states good model fit, therefore an insignificant result is desirable.

and shows that motorcycles are not seen as an alternative to cars, but rather as an addition. Due to the nonlinearity of the MNL model, the interpretation of log-odds and odds ratios is limited to their sign and, also, to the comparison with the base category. To allow for a deeper understanding of the model, we will move on to the interpretation of standardized coefficents, AMEs and the effects of a variable over its value span for typical household values. AMEs and standardized coefficients are presented in Table 5. The depicted plots show the change in probabilities for the change in a specific attribute over its value range, as explained in Chapter 5.2.2. For most of the variables, the plots show similar effects as the AMEs.

	1 vs. 0 cars			2 vs. 0 cars			3+ vs. 0 cars		
Variable	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.
(Intercept)	-4.140	0.220	***	-13.329	0.267	***	-20.590	0.433	***
licenses	4.713	0.085	***	8.690	0.124	***	11.012	0.243	***
suburban	0.892	0.054	***	1.442	0.062	***	1.664	0.085	***
rural	1.078	0.093	***	1.713	0.102	***	2.039	0.123	***
income	0.786	0.039	***	1.824	0.048	***	1.989	0.070	***
licenses squared	-1.015	0.024	***	-1.632	0.029	***	-1.777	0.046	***
ps.2	-0.170	0.113		-0.356	0.119	***	-0.357	0.130	***
ps.3	-0.506	0.115	***	-0.899	0.122	***	-1.073	0.136	***
ps.4	-0.841	0.120	***	-1.415	0.130	***	-1.609	0.152	***
income squared	-0.059	0.004	***	-0.130	0.005	***	-0.130	0.007	***
carsharing	-1.769	0.065	***	-2.636	0.085	***	-2.886	0.143	***
full-time	-0.022	0.035		0.396	0.038	***	0.652	0.045	***
m.carsharing	-1.986	0.097	***	-2.775	0.129	***	-3.063	0.239	***
motorbikes (log)	0.075	0.008	***	0.115	0.009	***	0.174	0.009	***
av.triplength	0.106	0.008	***	0.184	0.009	***	0.195	0.012	***
multifamily	-0.516	0.046	***	-0.779	0.053	***	-0.834	0.073	***
apartmentbuilding	-0.601	0.073	***	-1.008	0.100	***	-1.009	0.194	***
other	-0.630	0.057	***	-0.750	0.068	***	-0.536	0.093	***
pt.2	0.009	0.136		0.041	0.142		-0.042	0.152	
pt.3	-0.139	0.141		-0.276	0.149	*	-0.451	0.165	***
pt.4	-0.271	0.150	*	-0.592	0.165	***	-0.570	0.206	***
av.triplength squared	-0.003	0.000	***	-0.005	0.000	***	-0.005	0.000	***
generation	0.756	0.048	***	0.861	0.061	***	0.980	0.095	***
part-time	0.124	0.110		0.454	0.119	***	0.581	0.138	***
children	0.190	0.064	***	0.240	0.068	***	-0.160	0.078	**
bikes	-0.057	0.015	***	-0.108	0.016	***	-0.109	0.019	***
av.trips	0.075	0.009	***	0.083	0.011	***	0.103	0.015	***
metro.2	0.165	0.050	***	0.436	0.064	***	0.351	0.108	***
metro.3	0.352	0.084	***	0.681	0.100	***	0.742	0.145	***
metro.4	0.367	0.061	***	0.679	0.076	***	0.724	0.118	***
train.2	0.101	0.049	**	0.150	0.061	**	0.166	0.090	*
train.3	0.259	0.069	***	0.460	0.081	***	0.547	0.111	***
train.4	0.201	0.098	**	0.441	0.109	***	0.669	0.134	***
garage	0.110	0.078		0.287	0.085	***	0.393	0.101	***
bus.2	0.135	0.061	**	0.234	0.067	***	0.303	0.080	***
bus.3	0.230	0.153		0.345	0.160	**	0.450	0.173	***
bus.4	0.042	0.185		0.244	0.194		0.365	0.208	*
part-time squared	-0.295	0.077	***	-0.354	0.082	***	-0.373	0.095	***

Coeff. = Coefficients, Std. Err. = Standard Error, Sig. = Significance

Table 4: Coefficients of the MNL Model[230]

[230] Own analysis.

Variable	AME: Zero cars	AME: One car	AME: Two cars	AME: Three cars	Stand. Coeff.	Rank	Rel. Imp.
licenses	-0.121	-0.098	0.132	0.086	18.2	1	0.37
income	-0.028	-0.039	0.055	0.013	7.8	2	0.16
av.triplength	-0.004	-0.002	0.005	0.001	3.3	3	0.07
suburban	-0.051	-0.030	0.064	0.017	2.0	4	0.04
rural	-0.059	-0.036	0.071	0.024	1.9	5	0.04
ps.4	0.047	0.036	-0.066	-0.017	1.4	6	0.03
carsharing	0.139	-0.035	-0.103	-0.023	1.3	7	0.03
motorbikes	-0.004	-0.002	0.003	0.003	1.2	8	0.02
ps.3	0.026	0.031	-0.044	-0.013	1.2	9	0.02
multifamily	0.028	0.011	-0.032	-0.006	1.0	10	0.02
m.carsharing	0.162	-0.042	-0.096	-0.024	0.9	11	0.02
metro.4	-0.022	-0.022	0.037	0.006	0.9	12	0.02
full-time	-0.001	-0.055	0.041	0.017	0.8	13	0.02
old	-0.047	0.039	0.012	0.007	0.7	14	0.01
other	0.033	-0.017	-0.025	0.009	0.6	15	0.01
part-time	-0.002	-0.031	0.031	0.009	0.6	16	0.01
train.3	-0.015	-0.014	0.022	0.007	0.5	17	0.01
apartmentbuilding	0.033	0.023	-0.051	-0.006	0.5	18	0.01
bikes	0.003	0.004	-0.006	-0.001	0.5	19	0.01
train.4	-0.012	-0.024	0.022	0.014	0.5	20	0.01
av.trips	-0.004	0.003	0.001	0.001	0.5	21	0.01
metro.3	-0.021	-0.025	0.039	0.007	0.4	22	0.01
ps.2	0.008	0.017	-0.023	-0.003	0.4	23	0.01
pt.3	0.008	0.013	-0.011	-0.010	0.4	24	0.01
pt.4	0.016	0.027	-0.040	-0.004	0.4	25	0.01
metro.2	-0.011	-0.025	0.036	-0.000	0.3	26	0.01
bus.2	-0.007	-0.007	0.010	0.005	0.2	27	0.00
bus.3	-0.012	-0.005	0.011	0.007	0.2	29	0.00
garage	-0.007	-0.016	0.018	0.008	0.2	30	0.00
train.2	-0.006	-0.001	0.006	0.001	0.2	31	0.00
bus.4	-0.003	-0.025	0.020	0.008	0.1	32	0.00
pt.2	-0.001	-0.003	0.007	-0.004	0.0	33	0.00

Stand. Coeff. = Standardized Coefficients, Rel. Imp. = Relative Importance

Table 5: Average marginal effects and standardized coefficients[231]

Standardized model coefficients report the importance of attributes in the model with regard to the outcome. The most influential attributes in this study are the number of licenses, household income and the average trip length. While all of these attributes are hardly influenceable by policy-makers, most of the variables with moderate importance offer the chance to impact household choices. Especially the quality of public services near the place of residence seems to be of high importance and has a higher influence on car ownership levels than the quality of public transport. Moreover, both carsharing attributes are ranked among the ten most influential variables in the model. Lastly, the housing attribute representing parking space also shows moderate importance, indicating that limited parking space influences car ownership decisions. Out of all public transport variables, the availability of busses appears to have the lowest impact on a household's car ownership decisions. The distances to train and metro stations exert higher impacts. Moreover, the existence of children and the availability of a garage are of low rele-

[231] Own analysis.

vance, according to standardized coefficients. We will continue with a more in-depth analysis of the attributes and their influence on different levels of car ownership.

In literature, household income is seen as an essential variable as only sufficient financial means allow to own and maintain vehicles.[232] When looking at the AMEs and the effect plot an increase in income leads to a decrease in the predicted probabilities of zero and one car (by 3% and 4 % for a small increase in income, respectively) and an increase in the probabilities of owning two or three cars, which is also visible in Figure 2. The increase in predicted probability of owning two cars is higher than the increase for three cars. This may be based upon the reason that high-income households, instead of buying a higher number of cars, invest in more expensive and luxury cars. In our study, the greatest AMEs are evoked by the variable capturing the number of licenses per household, where another license leads to changes in the predicted probabilities for all categories of 9 % and more. The corresponding plot visualizes the nonlinear pattern that is masked by the AMEs. The probability of owning zero cars steadily declines with a higher number of license holders while the likelihood of owning one car only declines for a number of license holders greater than one. For four driver's licenses the probability of owning three or more cars is the highest in our sample, indicating that a significant proportion of people holding a driver's license want to possess their own vehicle, irrespective of the number of existing cars in a household. These results confirm the findings of earlier studies by Potoglou and Kanaroglou, and Chu, who found the influence of the number of licenses on the number of owned vehicles to be greater than the influence of the number of workers and household size.[233]

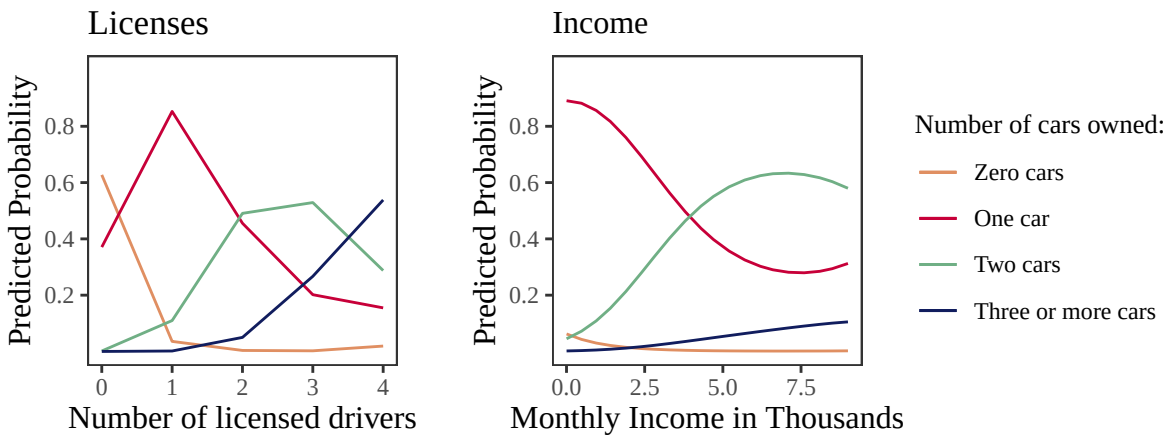


Figure 2: Influence of the number of lisence holders and monthly income on predicted probabilities[234]

[232] Cf. Roorda et al. (2000), p. 75.
[233] Cf. Chu (2002), p. 67; cf. Potoglou / Kanaroglou (2008b), p. 50.
[234] Own analysis.

The average number of trips and the average trip length is a representation of a household's mobility needs. The AMEs in Table 5 show that a higher number of trips reduces the probability of owning no cars and conversely, increases the probabilities of possessing one, two and three cars, however, by less than 1 % on average. The effect does not change over the value span of the variable, as shown in Figure 3. Average trip length, on the other hand, shows a nonlinear pattern, mostly affecting the probabilities for one and two cars. While the AME for owning two cars is positive, the plot only shows an increase in the predicted probability up until a medium trip length. For longer average trip lengths, however, the opposite is the case, which may indicate that for short to medium trip lengths people tend to favor a car, while for longer trip lengths other modes of transport may propose an alternative. The results are to be interpreted with caution, since the variable's computation as an average may mask substantial differences between household members.

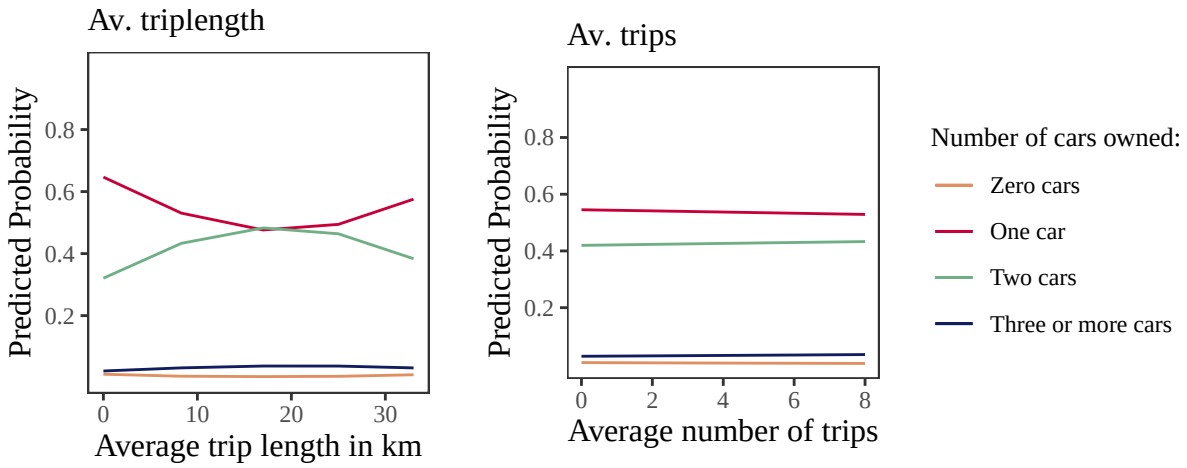


Figure 3: Influence of average trip length and the average number of trips on predicted probabilities[235]

We find that both, a higher number of full-time workers and a higher number of part-time workers, increase the probability of owning two or more cars, however the influence of full-time workers is more pronounced with AMEs of up to 6 % for an increase in the number of workers. This is contradictory to results by Potoglou and Kanaroglou, who found that part-time workers in Canada were less likely to own two cars.[236] A possible explanation for the different findings could be the higher occurence of part-time employment in Germany when children are present. In Germany, 73 % of women with children work in part-time models,[237] while in Canada only 21 % of women in dual earner households with children work part-time.[238] Therefore, part-time

[235] Own analysis.
[236] Cf. Potoglou / Kanaroglou (2008b), p. 50.
[237] Cf. Statistisches Bundesamt (2019).
[238] Cf. Statistics Canada (2015).

employment in Germany may have a higher correlation with having children, and the higher mobility needs generated by those children, than it does in Canada, and the findings may stem from this difference. According to Table 5, an increase in both full-time and part-time workers has, on average, a negative effect on owning zero cars and one car. Nevertheless, for a high amount of part-time workers the predicted probability of owning zero cars for an average household rises in Figure 4. This may be attributed to existing budget constraints in households with only part-time workers.

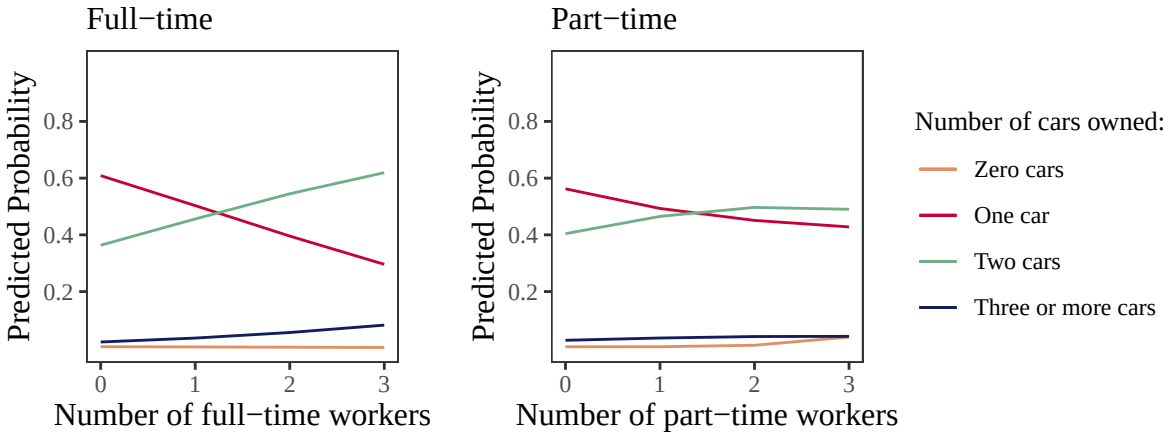


Figure 4: Influence of the number of full-time and part-time workers on predicted probabilities[239]

When looking at the age composition, the AMEs indicate that older household compositions on average reduce the likelihood of owning zero cars by 5 % and increase the probabilities of owning cars in general, which may indicate that younger people are less inclined to own more cars. In the effect plot regarding household generation in Figure 5, the effect of a household belonging to an older generation on owning one car is negative, in contrast to the AMEs, indicating an even stronger tendency to own more cars. The results complement those of Ritter and Vance, who found a positive effect for the influence of age on car ownership in Germany for the age group between 40 and 64 years.[240] Studies on generation effects in other countries, though, were not able to show significant effects, thus, the interpretation of results remains ambigous. The effect may also be due to economic constraints in the younger age gap. Young families and households may, even at the same income, spend a lower share on cars, as other expenses such as education, mortgage fees or similar dominate. As a last socio-economic variable we look at the effect of the existence of children. The existence of children on average only increases the probability of owning one and two cars while it decreases the probability of the other al-

[239] Own analysis.
[240] Cf. Ritter / Vance (2013), p. 79.

ternatives. The fact that the probability for three cars is reduced might be due to the fact that households with children rather buy one or two larger cars instead of three cars, especially as in a traditional family with young children, only two adults are present that possess driving licenses. Moreover, third cars are usually a luxury good, as can be seen in Figure 2, where the probability for three cars strongly increases for high-income households. Additional expenses associated with children may reduce purchasing power and thereby the probability to own three cars. Still, our hypothesis that children create the need for additional trips and mobility and thereby promote the purchase of additional cars is confirmed and in line with results of other studies on car ownership. However, the effect according to the standardized coefficent in Table 5 is rather small.

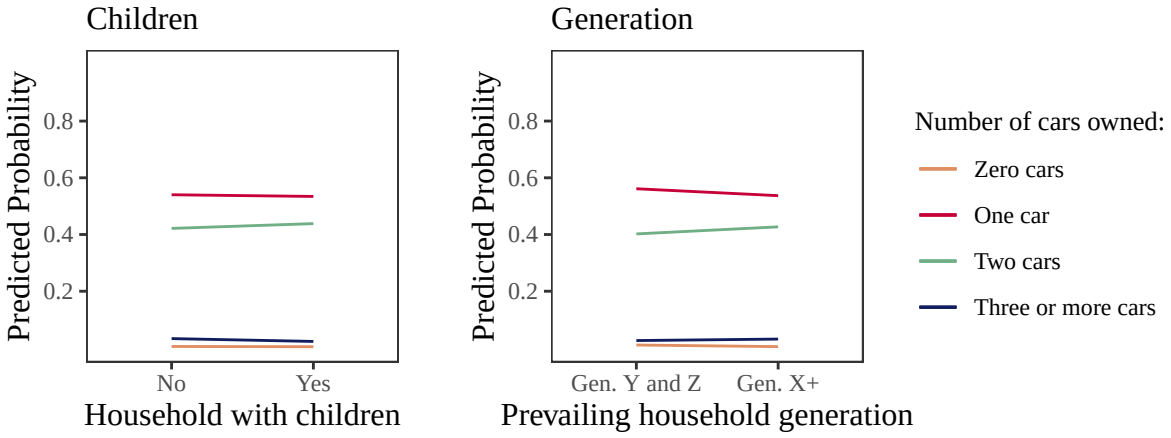


Figure 5: Influence of children and the prevailing generation on predicted probabilities[241]

Next, we analyze mobility-related attributes. Two variables regarding carsharing were introduced to analyze the effect of carsharing subscriptions on car ownership. The AMEs for one carsharing subscription show that in comparison with households that do not use carsharing, the probability of owning two cars is reduced by over 10 % and the probability of owning zero cars is increased by 14 %. The likelihoods of owning one car or more than three cars are only reduced slightly when looking at AMEs. In Figure 6, the probabilities for both owning zero cars and one car are increased for the average values assumed by the effect plot. The coefficient for the subscription to multiple carsharing services is significant as well but only shows marginally higher AMEs for several levels, implying that already the subscription to one carsharing service is sufficient to reduce car ownership. From the results it can be concluded, that the influence of carsharing is twofold. On the one hand, carsharing may promote households to stop owning cars at all. Additionally, carsharing can especially shift second-car ownership

[241] Own analysis.

towards the ownership of only one car. Which influence prevails may depend on the existence of regular mobility needs that cannot be satisfied by carsharing, such as day trips or trips with longer distances, that are not encompassed in the area of the carsharing subscription.

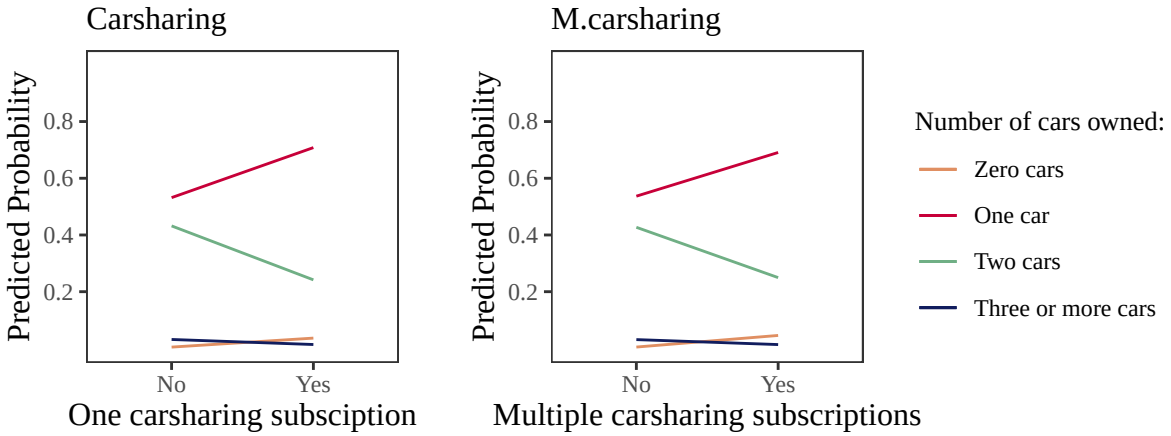


Figure 6: Influence of the quality of carsharing subscriptions on predicted probabilities[242]

The number of motorbikes and standard bikes has not been introduced yet into studies of car ownership in western countries but was often used as an explanatory variable in studies concerning emerging countries, where bikes and motorbikes are both used as a substitution to cars. Interestingly, in our study, bikes and motorcycles show contrary effects on car ownership in AMEs and the effect plot in Figure 7. On average, a higher number of bikes increases the likelihood of owning one car and zero cars while decreasing the likelihood of two and more cars, even tough the effects are small. The opposite is true for motorcycles. This indicates, that while bikes might reduce the need for more cars to some extent, motorcycles do not, or are at least not bought for this purpose. This might be due to the fact that motorcycles need an additional license in Germany, are more expensive than bikes and are bought by people who enjoy driving motorized vehicles in general.

Lastly, we examine the built-environment attributes introduced into the model. The housing type and the garage attributes are used as indicators of parking space at the place of residence. In comparison to living in a single-family home, both the residence in an apartment building and in a multi-family home reduce the likelihood of owning two and more cars and increase the likelihood of owning one or no cars. The AMEs are stronger for apartment buildings, which is in line with expectations, as more households reside in the same amount of space, compared to multi-family homes. The same impacts are observable for the garage-dummy.

[242] Own analysis.

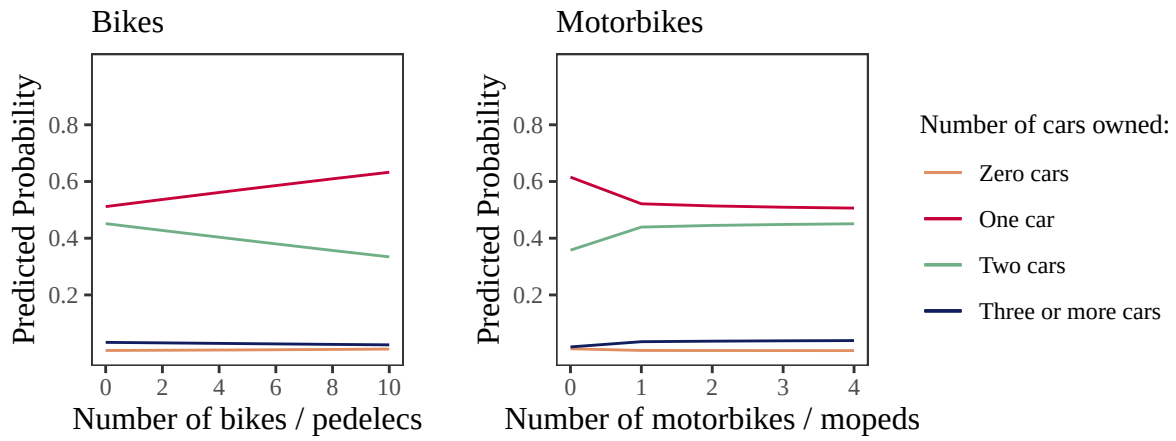


Figure 7: Influence of the number of bikes and motorbikes on predicted probabilities[243]

Households owning garages are more likely to own two and more cars, suggesting that larger parking space stimulates additional car ownership. We also investigate the effects of the regional attribute. It can be observed that, in comparison to living in an urban environment, suburban and rural regions increase the likelihood of owning two and more cars at an AME of 7 % and 2 %, while reducing the likelihood of owning no cars or only one car. The difference in predicted probabilities is not as high as expected between rural and suburban areas, indicating that suburban environments already promote the ownership of more than one car to a quite high extent. The observed effects are confirmed by the behavior of the effect plots in Figure 8.

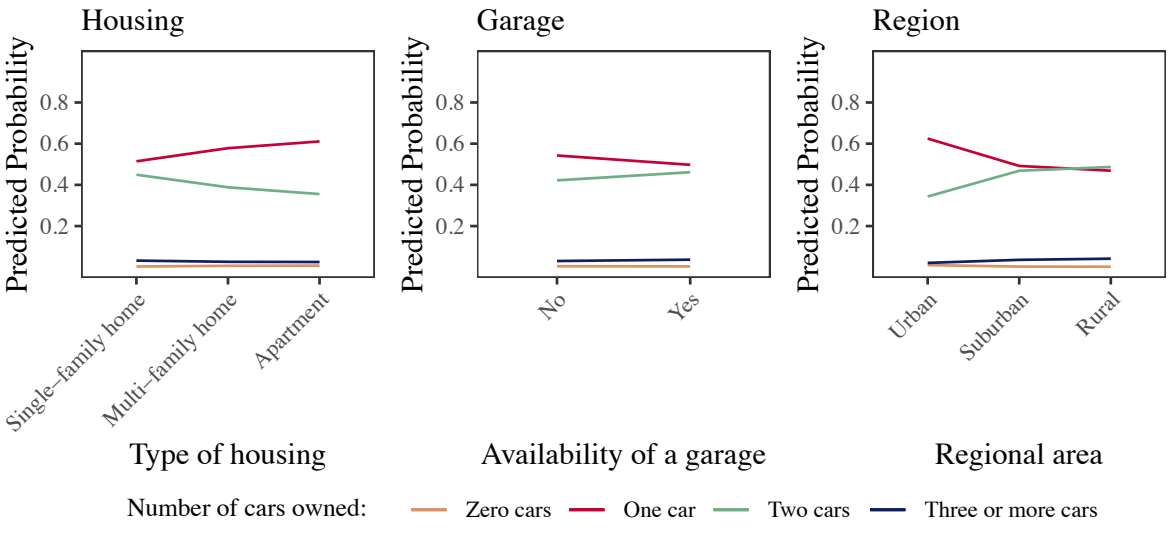


Figure 8: Influence of housing type and garage availability on predicted probabilities[244]

As expected, a higher quality of public transport has a negative effect on the probability to own

[243] Own analysis.
[244] Own analysis.

two cars and three or more cars. However, the coefficient corresponding to an improvement on the quality scale from one to two was not statistically significant as displayed in Table 4, which is also visible in Figure 9. This implies that households are only affected in their ownership choice when public transport reaches a notably higher quality. For public services in general, all improvements are statistically significant and reduce the likelihood of owning two and more cars, while promoting the possession of one car as well as no car. For both variables the AMEs increase with increasing quality, suggesting that a further improvement of relatively high quality may still impact car ownership choices. The AME for the same level of quality is stronger for public services, suggesting that an increase in quality and availability of public services is a more efficient mean of lowering car ownership levels than increasing the quality of public transport. However, both, higher public transport quality and higher quality of public services in the neighborhood may reduce high levels of car ownership, but do hardly decrease the ownership of one car.

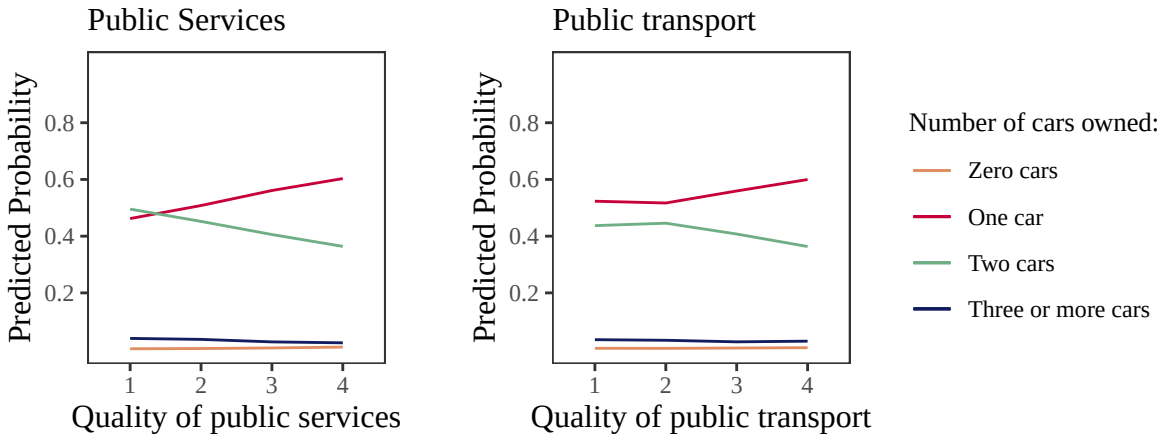


Figure 9: Influence of the quality of public transport and public services on predicted probabilities[245]

The variables capturing the distance to bus, train and metro stations show the expected signs and effects. A farther distance to public transport stations leads, on average, to a decrease in the probability to own zero or one car and an increase in the probability of owning two or more cars. As can be seen from Figure 10, however, the largest changes in probability take place at different distances for the three modes of transport. Regarding the distance to train stations the greatest difference is observed between a medium and far distance, for metro stations the greatest impact, at least for the one- and two-car cases, already take place at the short- to medium distance stage. As trains often cover trips of longer distance at a faster pace, the acceptance of a longer trip to the stop itself might be more accepted than for the other two modes. Metro

[245] Own analysis.

stations are most often found in cities, where longer distances to public transport are less likely to be accepted than outside of cities, as trips are shorter in general. Regarding the bus attribute, some levels are not significant and no substantial differences in impacts are given between different distances to the next stop. This may be caused by substantial heterogeneity of length and speed of bus routes.

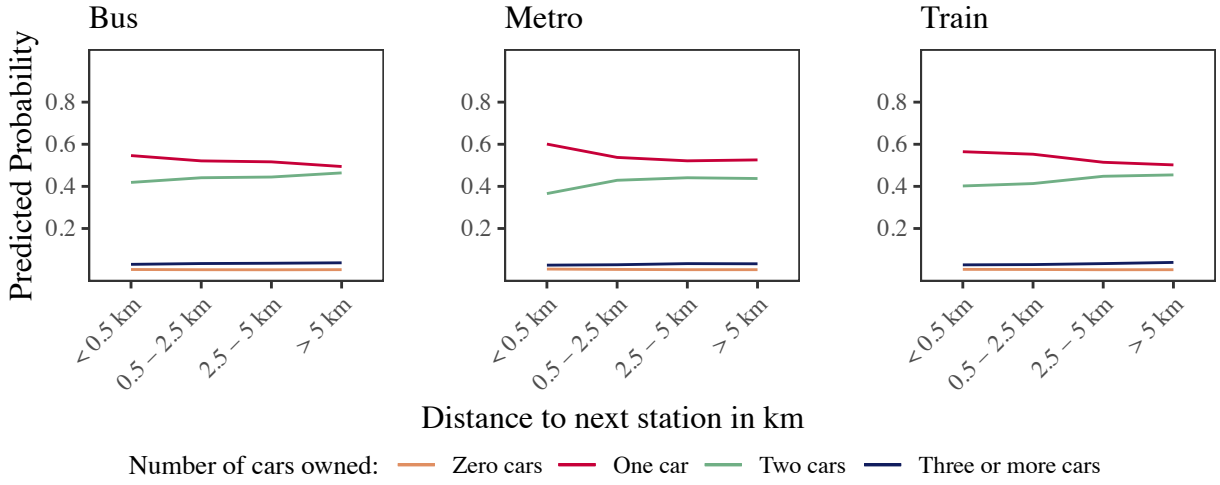


Figure 10: Influence of the distance to public transport stations on predicted probabilities[246]

It is evident when looking at the performance of AMEs and the intercepts in plots in general, that one car is considered more of a necessity than two or three or even more cars. Most of the variables reduce second- and third-car ownership while in turn fostering the ownership of zero or one cars or the other way around. This indicates that the reduction of two and more cars to a lower number may be influenced more easily by diverse factors, while incentivizing the change from one to zero cars is much more difficult.

6.2 Variance in Feature Impact in Different Regional Environments

This chapter will report on the findings from the analysis of the model's interaction effects of different independent variables with the variable "region", including the levels urban, suburban and rural. While variables indicating the regional environment of a household have been included in existing literature, the analysis of interaction effects with other variables has not been presented yet. Model validation for the model including interactions can be found in Appendix 16 and corresponding coefficients can be found in Appendix 17. The model displays a lower AIC value and significant overall model evaluations compared to the base model and the null model with a p-value smaller than 0.0001, which confirms that the included interaction terms

[246] Own analysis.

offer information beyond the information gained from the model only featuring main effects.

Regarding the magnitude and influence of predictors in different regions, we expect a higher influence of carsharing in urban regions, due to a higher availability of carsharing services and lower car need in general. Likewise, the effect of both, additional part- and full-time workers, as well as additional license holders is anticipated to be smaller for urban environments, as work and other activities are more reachable than in non-urban regions. Due to shorter distance, the effects for bikes are expected to be elevated in urban and suburban areas. The effect of more parking space indicated by housing types and garages should be higher in urban environments as parking space is scarce in general. The effect of children is believed to be stronger in less urban regions, as in dense urban environments with a close variety of activities, children may provoke less needs for additional car ownership. Similar influences across regions are anticipated for the quality of public services, an older age compositions, the influence of trainstations and the effect of higher income.

The effect plots allow for the comparison of the influence of different regional environments across the range of independent variables for varying levels of car ownership. Moreover, they aid in the assessment of significant differences between regional effects. Additionally, in Appendix 18, AMEs for the analyzed variables for all three regional categories are shown. In this chapter, effect plots are only reported and interpreted for car ownership levels that demonstrate differences between regions for both slopes and intercepts. In the case that not all car ownership levels are displayed, complete effects plots for all car ownership levels can be found in Appendices 19 to 24. Moreover, effect plots of interactions with the variables income, average trip length, average number of trips, number of motorbikes and distance to train stations are only displayed in Appendices 25 to 29, as they do not remarkable differences for any car onwership levels. In general, our expectation that the probability to own zero cars or one car is significantly higher in urban regions is confirmed by the behavior of effect plots. Across all variables, the intercepts for urban effects are above those of suburban and rural effects and are significantly different for at least certain values of the indepent variable in question. The intercepts for the probability of having two or three and more cars are, in turn, below those of other areas. Suburban and rural regions, on the other hand, show similar intercepts across most variables analyzed. An equal pattern is present for slopes and AMEs. Both show larger differences and magnitudes for urban environments, while slopes and AMEs are closer for other environments. All differences in AMEs are significant. We will proceed with a more detailed analysis and

interpretation of the attributes in question. The order of attributes follows the chapter above, thereby starting with socio-economic attributes, continuing with mobility-related characteristics and lastly, reporting on built-environment attributes.

The number of license holders shows only significant differences for urban regions in our plots. We can observe in Figure 11 that the predicted probabilities for urban households decrease (with respect to owning one car) and increase (with respect to owning three cars) at a slower pace than they do for suburban and rural households, which is in line with hypothesized results. The effect on the probability of owning two vehicles is similar between the three regions. The AMEs in Appendix 17 display equivalent effects, as they show differences of 4 % between urban and suburban or rural regions for the one- and three-car probabilities. Possibly due to limited space and shorter distances in urban surroundings, households seem less inclined to buy more cars when more household members acquire driver's licenses.

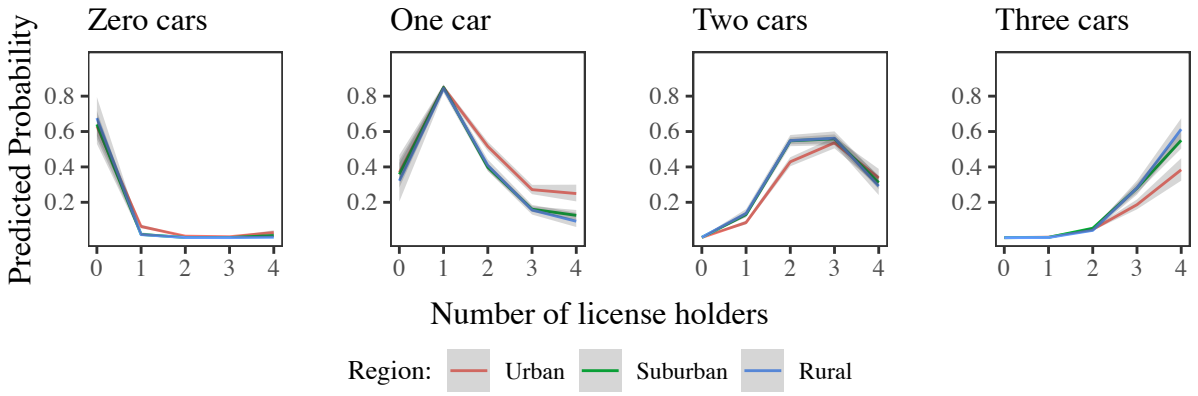


Figure 11: Influence of the interaction between the variables "licenses" and "region" on predicted probabilities[247]

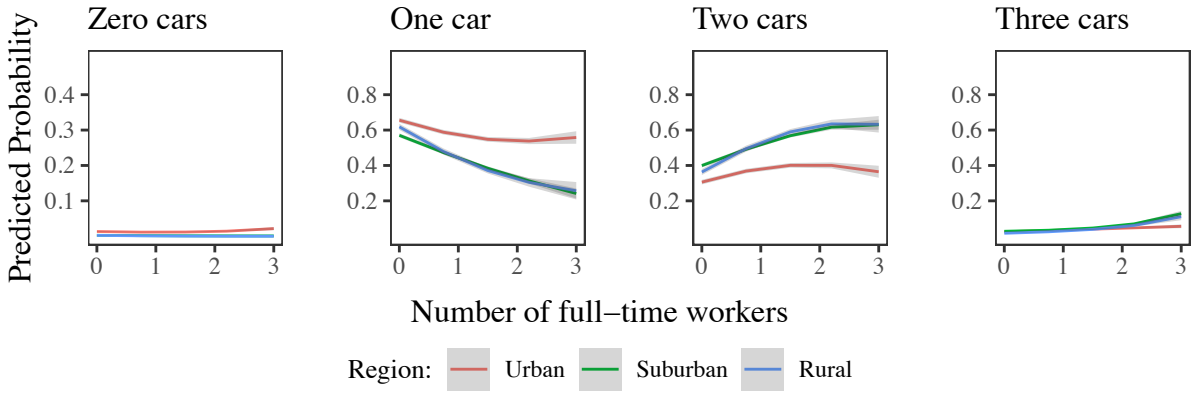


Figure 12: Influence of the interaction between the variables "full-time" and "region" on predicted probabilities[248]

[247] Own analysis.
[248] Own analysis.

The influence of the number of workers differs across regional environments and the differences are the same for effect plots in Figure 12 and Figure 13 and the corresponding AMEs. The probability of owning one car decreases with AMEs of up to 8 % for a higher number of workers in suburban and rural settings, while the average marginal decrease in urban environments stays below 4 %. Similarly, the likelihood to own two cars in urban households remains almost unchanged for a higher number of workers, while suburban and rural households show a substantial increase. The same pattern can be observed for part-time workers. Regarding the ownership of zero and three cars, we find differences between regions but also between part-time and full-time workers. In part-time households, a higher number of part-time workers increases the probability of owning zero cars for urban households, while for full-time workers, the effect is a lot smaller. This might be explainable by budget constraints, however, those constraints seem not to apply to suburban or rural households, as a car might be of more necessity there. For three cars, the opposite can be observed. The probability of owning three or more cars increases with more full-time workers for rural and suburban environments only. For part-time workers, however, the lines stay flat. The lower impact in urban regions is likely caused by the on average higher accessibility of the place of work by means of public transport.

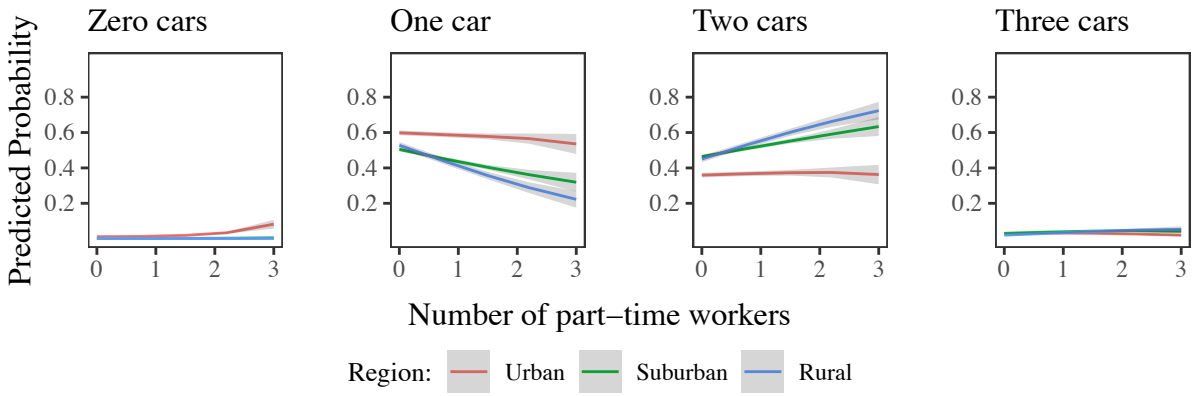


Figure 13: Influence of the interaction between the variables ”part-time” and ”region” on predicted probabilities[249]

AMEs of having children show negative influences of the same size in all environments for the probability to own zero and three cars. This indicats that our interpretations in Chapter 6.1.2 hold across all regions. Nevertheless, the existence of children promotes contrary effects regarding the decision to own one or two cars, as displayed by both, the effect plots in Figure 14 and the AMEs for these cases. In urban areas, households with children have a higher probability of owning one car and a lower likelihood to own two vehicles than households without children.

[249] Own analysis.

For suburban and rural areas, opposite effects can be observed. The expectations that the effect itself is smaller in urban environments due to shorter distances and higher density of activities for families is not confirmed, as the effects show equal magnitudes. This indicates that, while additional mobility needs generated by children foster car ownership in all environments alike, in rural and suburban areas households rather change from being a one-car household to being a two-car household, while urban households start owning cars when children are born.

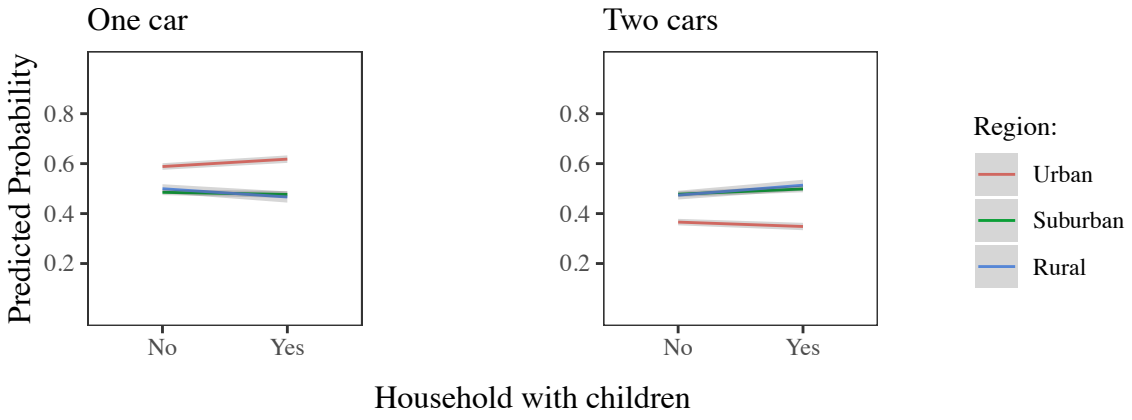


Figure 14: Influence of the interaction between the variables ”children” and ”region” on predicted probabilities[250]

The behavior of the variable representing older household compositions provides interesting insights into the effects of the variable in Chapter 6.1.1. In Figure 15, we can observe that a more senior age composition decreases the probability of owning one car and increases the likelihood of owning more cars in urban households. At the same time, the opposite is true for suburban and rural households. Likewise, in urban environments, an older age composition reduces the probability of owning zero cars, while there is no visible influence in suburban and rural regions. The effect analyzed in the previous chapter seems to only be valid in urban environments, where younger households own fewer cars. In more rural environments younger households tend to have an even higher likelihood to own several cars than older households. A possible explanation might be that younger households have higher mobility needs than older households in general. In urban areas, these higher mobility needs are catered to by public transport. In other regions, however, a car is needed.

Regarding the influence of carsharing, significant differences between the regions can be observed. While in urban and suburban environments, the probability of owning zero cars increases by an AME of 16 % and 17 %, respectively, for households with carsharing subscriptions, the likelihood for rural households hardly changes. Regarding the probability to own one

[250] Own analysis.

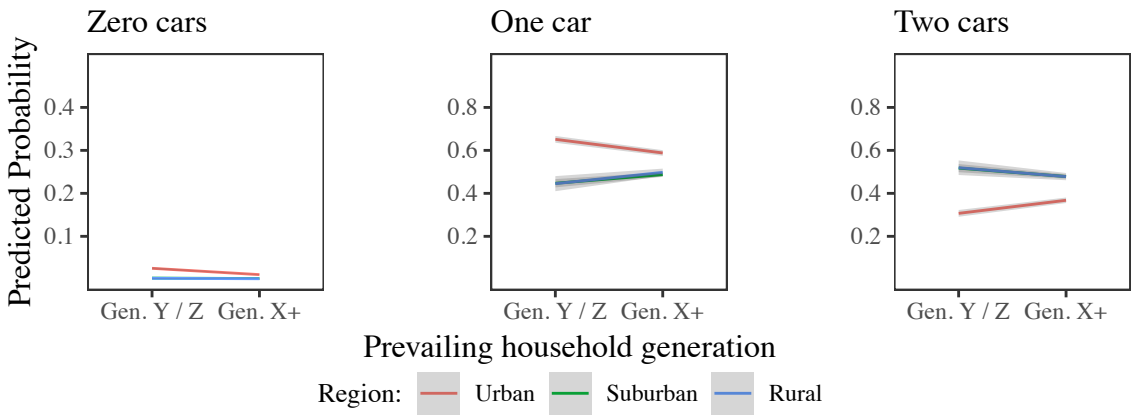


Figure 15: Influence of the interaction between the variables ”generation” and ”region” on predicted probabilities[251]

car, there is no significant difference between regions anymore when a household is subscribed to carsharing, meaning that when carsharing is available and used by households, the differences in ownership probabilities among different environments vanish to some extent. For the ownership of two cars, the slopes in Figure 16 show the highest increase for the suburban environment, which is confirmed by the highest AME of 12 % for this level. Even though there is no significant difference to rural areas as stated by the confidence intervals, the results indicate that carsharing programs are more likely to foster changes in car ownership in urban and suburban environments than in rural settings. One reason might be that in urban environments, stationary carsharing is more dominant, in comparison with free-floating carsharing in cities.[252] The lower influence of carsharing might be explainable by the sparseness or distance to carsharing stations in stationary carsharing and, consequently, the lower convenience for users.

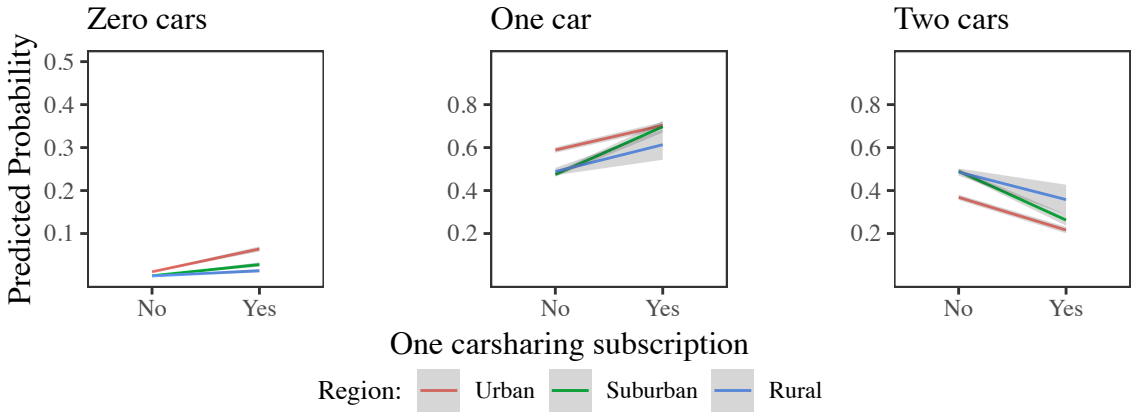


Figure 16: Influence of the interaction between the variables ”carsharing” and ”region” on predicted probabilities[253]

[251] Own analysis.
[252] Cf. Bundesverband Carsharing (2020), p. 1.
[253] Own analysis.

Regarding the number of bikes depicted in Figure 17, the slopes for urban and suburban households are positive for the probability to own one car and negative for the probability to own two cars. For rural environments, the opposite is true. This confirms our hypothesis that bikes and pedelecs can meet mobility needs and may reduce car ownership in regions where distances are short to moderate. For the probability of owning zero cars only the AME for urban environments is positive, however the effect only amounts to 0.3 %. Therefore, bikes and pedelecs, in contrast to our results for carsharing, cannot replace all types of trips for most households even when distances are short.

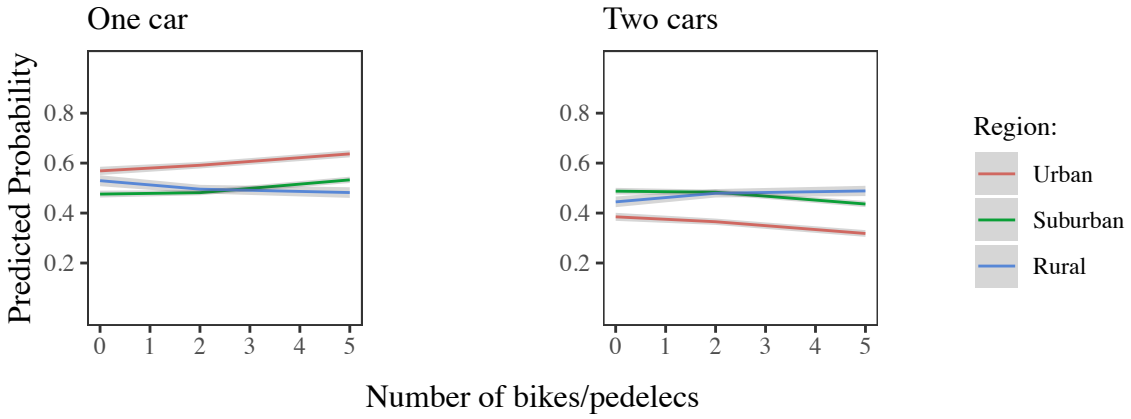


Figure 17: Influence of the interaction between the variables ”bikes” and ”region” on predicted probabilities[254]

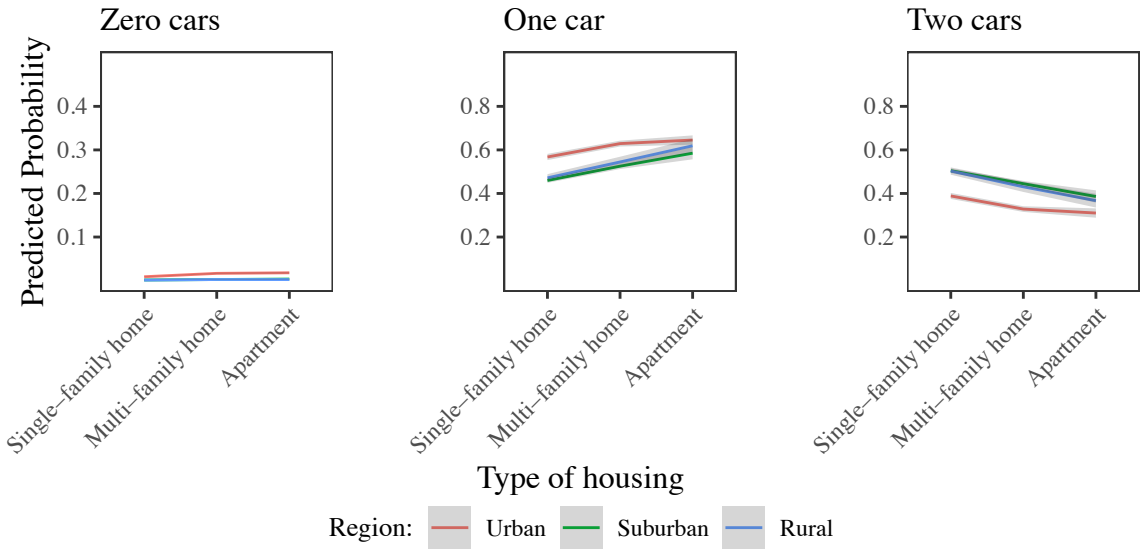


Figure 18: Influence of the interaction between the variables ”housing type” and ”region” on predicted probabilities[255]

For the influence of the variable housing type, a difference can be observed between urban and other households in Figure 18. For non-urban households, the change in probabilities between

[254] Own analysis.
[255] Own analysis.

a multi-family home and an apartment building is higher than for urban households. This can be explained by the fact that multi-family homes in more rural areas provide more parking space than in urban areas, where the availability of parking space is similar for both building types. Additionally, according to the confidence intervals shown in grey, we can observe significant differences in probabilities only for single- and multi-family homes. The residence in an apartmentbuilding does not exhibit significantly different influences on car ownership levels in distinct environments, as parking space there is low in general. Moreover, we can observe that for urban households, the building type does affect the probability to own zero cars to a small extent, while it does not affect households in rural or suburban regions, indicating that parking space is more rare. This phenomenon is confirmed by the behavior of the plots and AMEs for the variable ”garage”. Urban households are much more influenced in their ownership decision regarding one or two cars by the non-availability of a garage than suburban or rural households, possibly because finding a parking space in close distance is more difficult and garages can, thus, offer a higher convenience than in other regions. This can be seen by the differences in AMEs of up to 2 % and the significant differences in Figure 19 when no garage is present. For non-urban households the change is less pronounced, indicating that the impact of one additional private parking space is less important than the general parking situation in the close neighborhood, indicated by the housing variable.

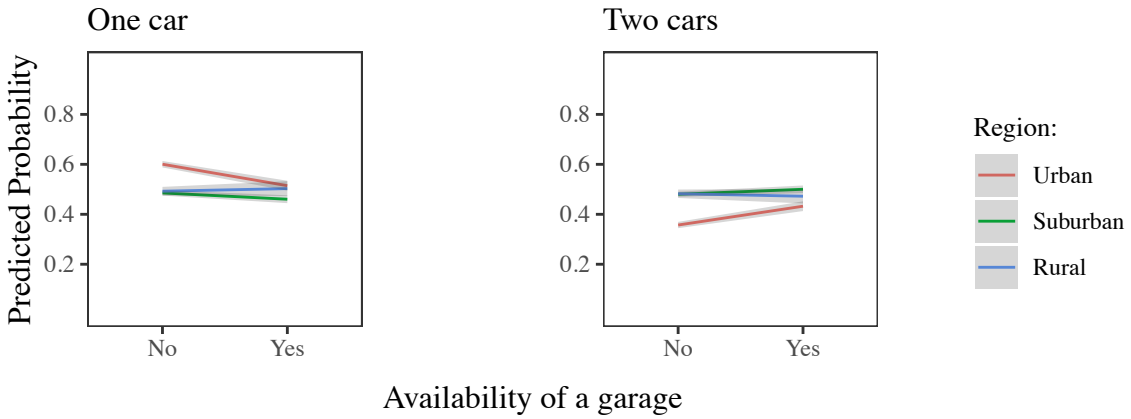


Figure 19: Influence of the interaction between the variables ”garage” and ”region” on predicted probabilities[256]

Regarding the likelihood of owning zero cars, a higher quality of public services increases the probability for urban households to a significantly higher extent than for the other regions as can be seen in Figure 20. This is in line with our expectation that the complete elimination of car ownership is not possible in all regions for most households. For the other cases, the slopes

[256] Own analysis.

and AMEs are stronger for urban areas to a small extent, but show similar patterns in general.

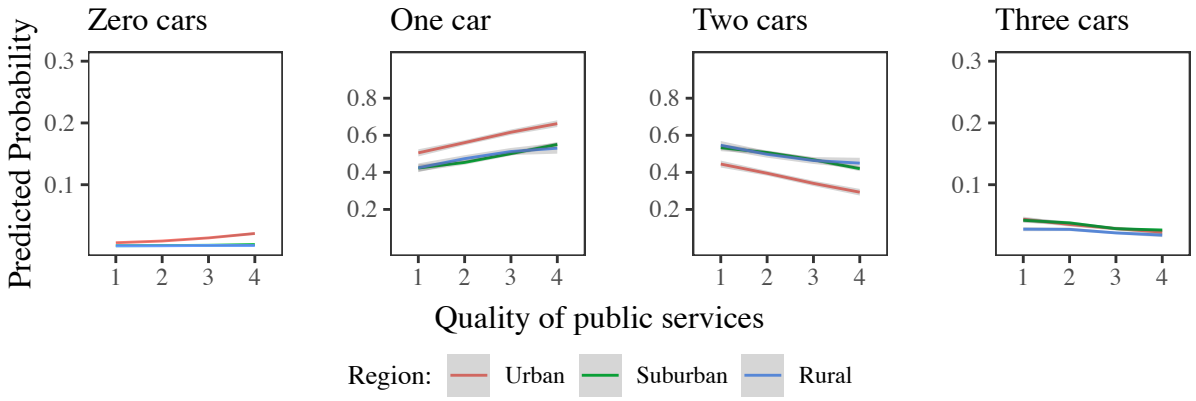


Figure 20: Influence of the interaction between the variables ”quality of public services” and ”region” on predicted probabilities[257]

In summary, most of the expected effects are validated by the model. The effect of carsharing appears more effective in reducing car ownership levels in urban and suburban environments. Additional workers and license holders exert less influence in urban environments. Similarly, the impact of the number of bikes is more pronounced in urban and suburban environments were distances are shorter and bike paths are often superior. Less parking space in neighborhoods reduces the ownership of several cars in all regions, however, to an even higher extent in non-urban regions, as in urban regions parking space is low in general. Furthermore, the influence of a better quality of public services are significant across all regions, but more pronounced in urban environments. Children and the prevailing household generation generated different results than expected for some levels of car ownership. Children increase car ownership in all regions to a similar extent, however, they promote the ownership of one car in urban households, while rather promoting the ownership of two cars in other regions. Similarly, a younger prevailing household age only indicates lower car ownership in urban regions, while showing even higher car ownership in other regions.

6.3 Comparing the Prediction Capability of Statistical and Machine Learning Classifiers

Our last research question is directed towards the difference in prediction performance between statistical regression models and ML algorithms. This section will report on the results generated from comparing the prediction capabilities of the MNL model used in Chapter 6.1 and a RF algorithm. It will begin with explaining the training process for the RF classifier, then compare predictions and lastly compare the estimated variable importances of both models.

[257] Own analysis.

6.3.1 Tuning Process

The prediction performance of ML classifiers is impacted by the choice of model parameters. This section reports on performance differences for different parameter combinations applied on training data and how the final RF specification is determined. To optimize the RF algorithm, different combinations of model parameters are tested to achieve a higher prediction accuracy. RF models are influenced by three main parameters, namely the number of splitting variables used per tree m, the total number of trees in the forest n and the minimum nodesize t. The model is tuned using the GridSearch algorithm in R's caret package. As decision trees are internally able to capture nonlinear terms,[258] we do not need to include the nonlinear terms that were specifically introduced into our MNL model. For simplicity reasons, the final model specification was achieved in two steps. First, the optimal number of trees was determined using the forest's out-of-bag error (defined as $1-$accuracy). In a second step, the minimum node size and

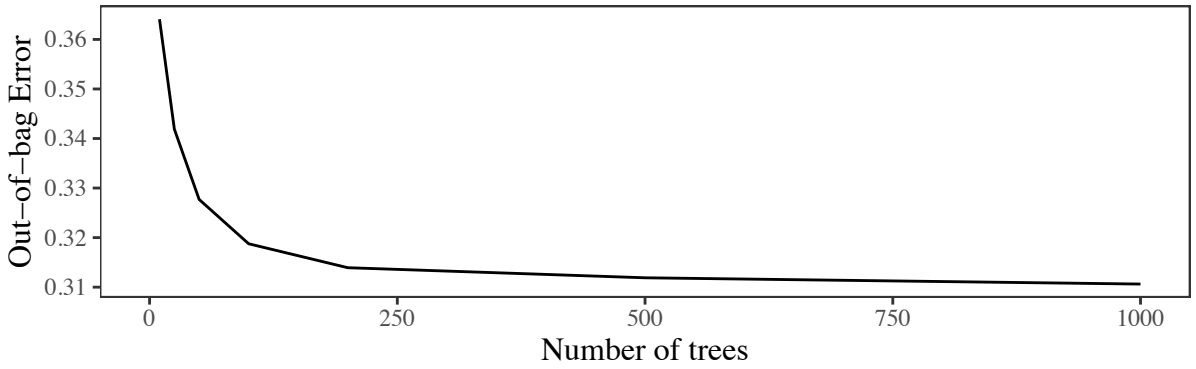


Figure 21: Out-of-bag error for different numbers of trees in the random forest algorithm[259]

the number of splitting variables were determined by comparing balanced accuracy values that, in turn, were calculated by a repeated 10-fold cross-validation process. Generally speaking, more trees will yield better accuracy, but at the cost of higher computational cost and diminishing improvement of accuracy.[260] Thus, following Cheng, we did not include the determination of optimal tree size directly into the tuning process, as this decision is rather about finding the right cut-off value than an optimal value.[261] We constructed RF algorithms with the number of trees ranging from ten to 1200. The comparison is shown in Figure 21. We can observe that forests beyond 500 trees only yield minor performance improvements, which is why we continue with a given size of 500 trees per forest. The default number of splitting variables is $\sqrt{p}$,

[258] Cf. Lafond et al. (2017), p. 130.
[259] Own analysis.
[260] Cf. Oshiro et al. (2012), p. 166.
[261] Cf. Cheng et al. (2019), p. 5.

with p being the number of explanatory variables in the model. However, Breiman and Cutler recommend trying values smaller and larger than the default.[262] Therefore, beside the proposed value of six for our case of 33 variables (including dummy variables generated for categorical variables), we test values from two to 33. Regarding nodesize, we test values ranging from one to 640. Results are shown in Figure 22. The best performance is found for 15 splitting variables and a nodesize of 40, which we settled on for our final model.

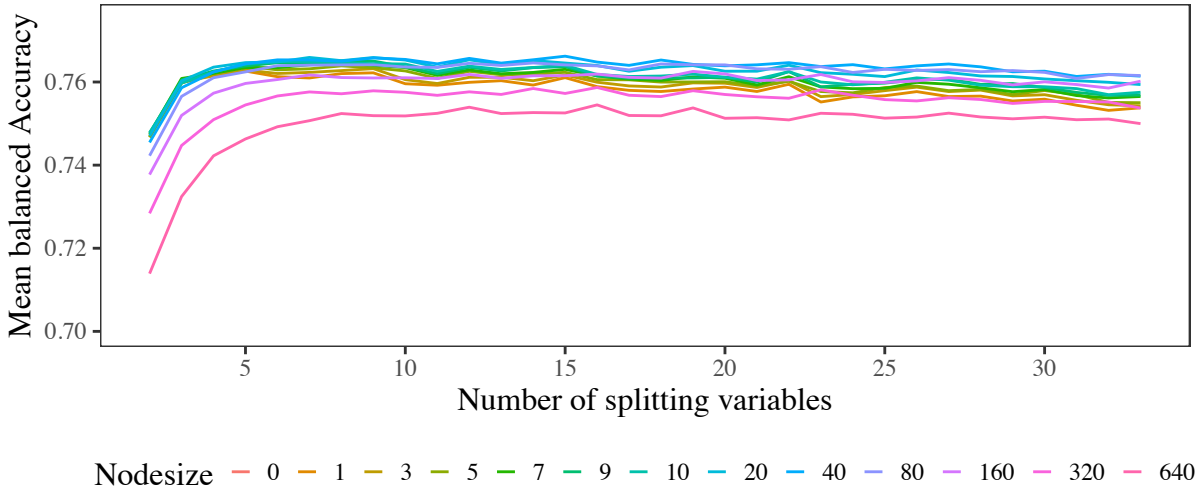


Figure 22: Mean balanced accuracy for different nodesizes and number of splitting variables[263]

6.3.2 Predicition Capability and Variable Importance

To verify the method that exhibits the best prediction performance regarding car ownership levels, a comparison is made between the developed MNL model and the RF method. With regard to train and test data, the same datasets are used to evaluate both models.

Measure		In-Sample Predictions		Out-of-sample Predictions	
		RF	MNL	RF	MNL
Balanced Accuracy	Overall	0.853	0.778	0.790	0.782
Balanced Accuracy	Zero cars	0.914	0.850	0.870	0.853
Balanced Accuracy	One car	0.820	0.741	0.757	0.745
Balanced Accuracy	Two cars	0.814	0.725	0.742	0.730
Balanced Accuracy	Three and more cars	0.863	0.795	0.793	0.802
Kappa		0.625	0.471	0.500	0.478

Table 6: Comparison of car ownership predictions[264]

Table 6 shows the overall balanced accuracy, balanced accuracy per category and Kappa of both models for in-sample and out-of-sample predictions. The RF decays in prediction performance

[262] Cf. Breiman / Cutler (2020), p. 1.
[263] Own analysis.
[264] Own analysis.

when using testing data compared to using training data, with a decrease of 6 % in overall balanced accuracy and 12 % in Kappa. The balanced accuracy of the MNL on the other hand even increases at a minor degree for the testing data, which is in line with observations made by Ha et al.[265] Both models show Kappa values of around 50 %, which indicate moderate model fit, as defined by Landis and Koch.[266] The study conducted by Ha et al. regarding vehicle ownership yields substantially lower Kappa values of below 31 % for all models, which might be due to a different study design and the lower amount of explanatory variables used.[267]

Our results demonstrate that, with regard to predicting car ownership levels, the RF algorithm outperforms the MNL model in balanced accuracy and Kappa. Moreover, by looking at the balanced accuracy measure for each category alone, it should be noted that the RF outperforms the MNL among all separate car ownership levels other than three and more cars. We, thus, conclude and confirm findings of previous papers, that the RF algorithm may improve prediction accuracy in comparison to traditional statistical methods when applied to transportation data. This result is in line with previous research in the field as listed in Chapter 4, however, the application to German data and to the prediction of the number of cars owned has not been presented before. The superiority may stem, to some extent, from the nonlinear nature of some variables, that is not automatically captured by traditional statistical models. Nevertheless, it should be noted that studies analyzing the prediction of other transportation-related issues, such as travel mode choice prediction, find higher discrepancies between RF and MNL predictions. Cheng et al. find discrepancies in overall model accuracy between an MNL and a RF of up to 22 % and Hagenauer and Helbich of over 30 %.[268] Though, the study conducted by Ha et al. regarding vehicle ownership prediction equally finds only performance differences of small magnitude. Thus, the magnitude of performance differences might be predetermined to some degree by the issue in question.

For a more in-depth analysis, we additionally report the variable importance measure of the RF in comparison to the standardized coefficients estimated for the MNL model. The comparison allows to reassess the importance of car ownership determinants in Germany analyzed in Chapter 6.1.2. The results are shown in Table 7. For both models, the variable importance measures exhibit similar patterns. The number of licensed drivers, income and the average trip length are the most important factors in both models. Other variables that have been ranked with high

[265] Cf. Ha et al. (2019), p. 84.
[266] Cf. Landis / Koch (1977), p. 165.
[267] Cf. Ha et al. (2019), p. 84.
[268] Cf. Cheng et al. (2019), p. 7; cf. Hagenauer / Helbich (2017), p. 277.

Variable	RF Results		MNL Results	
	Rank	Rel. Imp.	Rank	Rel. Imp.
Licenses	1	34.7%	1	36.9%
Income	2	23.6%	2	15.8%
Av.triplength	3	10.5%	3	6.7%
full-time	4	5.7%	13	1.6%
Av.trips	5	4.2%	21	0.9%
Bikes	6	3.9%	19	0.9%
metro.4	7	3.5%	12	1.7%
carsharing	8	1.4%	7	2.7%
motorbikes	9	1.4%	8	2.5%
generation	10	1.3%	14	1.5%
part-time	11	1.3%	16	1.2%
suburban	12	1.0%	4	4.0%
multifamily home	13	0.9%	10	2.0%
pt.2	14	0.9%	33	0.1%
rural	15	0.6%	5	3.8%
pt.4	16	0.6%	25	0.8%
pt.3	17	0.6%	24	0.8%
children	18	0.5%	28	0.5%
ps.2	19	0.5%	23	0.9%
ps.4	20	0.4%	6	2.7%
other	21	0.4%	15	1.2%
train.4	22	0.4%	20	0.9%
train28.3	23	0.3%	31	0.4%
m.carsharing	24	0.3%	11	1.8%
bus.2	25	0.3%	27	0.5%
ps.3	26	0.3%	9	2.5%
garage	27	0.3%	30	0.4%
train.3	28	0.1%	17	1.0%
apartmentbuilding	29	0.1%	18	1.0%
metro.3	30	0.1%	22	0.9%
metro.3	31	0.0%	26	0.7%
bus.3	32	0.0%	29	0.5%
bus.4	33	0.0%	32	0.2%

[1] Rel. Imp. = Relative Importance

Table 7: Comparison of Variable Importance[269]

importance by both models are the number of workers, the prevailing age of the household, the availability of carsharing and the region in which a household lives. Least important factors include the existence of children, the availability of a garage and all of the indicators for bus stops. From these results we see our results from Chapter 6.1.2 confirmed, as the RF algorithm demonstrates comparable results regarding the importance of the analyzed attributes.

7 Conclusion

Household car ownership and usage is one of the main contributors to environmental problems and its reduction in the future is necessary to limit further climate change. Therefore, sustainable transportation planning requires insights into the factors influencing the amount of

[269] Own analysis.

privately held vehicles. Even though a multitude of studies exist that examine car ownership in other countries, studies on car ownership in Germany are rather scarce and based on past data. Thus, this paper analyzes the effects of a variety of car ownership determinants in Germany based on a recent national mobility survey by using an MNL model. The model suggests that the number of licensed drivers and household income are the most important drivers of car ownership, as has been shown in previous research in other developed countries. However, other determinants on which policy-makers have more influence and that can reduce car ownership have been determined. Especially carsharing subscriptions show a large influence on the probability to own a lower number of cars and should be promoted more strongly to reduce car dependency. Moreover, results indicate that households place more focus on the quality of public services at the place of residence when making car ownership choices than they do on the quality of public transport. Nevertheless, higher public transport quality, as well as shorter distances to regularly served stops of different means of public transport show the ability to significantly decrease levels of car ownership. Furthermore, the positive coefficients of the rural and suburban dummies, as well as negative coefficients for attributes indicating less parking space, suggest that with regard to car ownership decisions, households take density measures into account. Lastly, the number of bikes also exerts a negative influence on car ownership levels, promoting the efforts carried out to create more bike-friendly infrastructure in many regions.

As especially cities and metropolitan areas in Germany experience rising car traffic and associated problems such as fine dust pollution, we contribute to previous research by placing particular focus on exploring the variance of variable impact in different regional environments. In terms of regional differences, we find that a number of variables display influence varying in strength and direction among the different regional environments. Especially urban and suburban environments seem to be able to reach a reduction in car ownership levels when stimulating carsharing services. Additionally, the effects of the attributes for bikes and the quality of public services are stronger in urban areas, suggesting that the advancement of all three may result in a significantly larger proportion of households owning one or even zero cars. Interestingly, in our study, a younger household composition does not generally translate into lower car ownership. This suggests that the current ”Fridays For Future”-movement has not fully reached present car ownership levels yet, but this may change in future years. Our results can aid policy makers concerning future transport-related decisions, as they offer evidence on relationships between socio-economic and mobility-related characteristics and auto ownership. The findings may additionally provide new input for cities and local authorities regarding future urban design and

transport planning.

Lastly, the study has implications for researchers in the field of mobility. We assess the performance of statistical and ML classifiers regarding the prediction of car ownership levels. Our analysis revealed that the predictive capability of a RF classifier was better than that of an MNL model. Thus, it reinforces the application of ML in future car ownership forecasts and other transportation research fields. Additionally, the MNL model used can serve as a basis for more comprehensive modeling approaches and scenario simulations in the future.

8 Limitations and Outlook

Although this study provides new interesting insights about car ownership in Germany, certain limitations have to be considered. First, one shortcoming stems from the problem of self-selection, referring to the interrelation of residential choice and travel behavior. Biases may be present in the model as household car ownership choices presumably impact the choice of living environment and not only the other way around. Further research could focus on overcoming this problem by estimating a separate model regarding residential choice and introducing generated predictions of residential choice into a car ownership model. Secondly, even though the dataset used is large, inherent biases may be present, that can harm the generalization of coefficients to Germany on the whole. Third, unobserved variables correlated to either car ownership or explanatory variables used in the model may distort and limit the validity of results.

Looking ahead, the results presented offer several potential directions for future research. A possible next step would be to validate the results presented using local household surveys or focus groups. Moreover, when examining smaller areas, attributes derived from geospatial data could be added, to derive more detailed results for built-environment attributes such as population, street and road density. Other variables that potentially impact car ownership could also be collected and analyzed. Those may include environmental attitudes, the impact of initiatives such as the “Job-ticket” or the “Job-Rad” in Germany or the presence and effect of so-called “car-pride”, referring to the symbolic significance of a car as a form of status symbol. Lastly, regarding the prediction of car ownership, the performance of other ML algorithms such as artificial NNs could be tested and compared to the findings of this paper in the future.

Appendix

Appendix 1: Overview of spearman correlation coefficients between explanatory variables

	Av.trips	Quality.PS	Quality.PT	Distance Train	Distance Bus	Distance Metro	Av.triplength	Fulltime	Bikes	Monthly Income	Licenses	Householdsize	Parttime	Motorbikes
Av.trips	1	0.04	0.06	–0.04	–0.03	–0.04	0.15	0.06	0.12	0.08	–0.02	–0.03	0.08	0.01
Quality.PS		1	0.36	–0.53	–0.28	–0.58	–0.12	0	–0.05	–0.03	–0.15	–0.11	–0.06	–0.1
Quality.PT			1	–0.34	–0.29	–0.13	–0.12	–0.03	–0.07	–0.07	–0.13	–0.12	–0.06	–0.08
Distance Train				1	0.23	0.24	0.11	0.02	0.04	0.03	0.12	0.09	0.05	0.08
Distance Bus					1	0.26	0.11	0.03	0.06	0.02	0.12	0.09	0.06	0.1
Distance Metro						1	0.1	–0.03	0.04	0	0.14	0.09	0.05	0.09
Av.triplength							1	0.29	0.18	0.24	0.24	0.2	0.14	0.12
Fulltime								1	0.33	0.53	0.36	0.42	0.12	0.2
Bikes									1	0.5	0.49	0.6	0.33	0.19
Monthly Income										1	0.58	0.6	0.3	0.16
Licenses											1	0.75	0.29	0.21
Householdsize												1	0.42	0.19
Parttime													1	0.14
Motorbikes														1

Appendix 1: Overview of spearman correlation coefficients between explanatory variables

Appendix 2: Overview of variance inflation factors for predictor variables

Variable	GVIF	DF[1]	Modified GVIF
Number of Licenses	2.795	1	1.672
Region	3.445	2	1.362
Monthly Income	1.753	1	1.324
Quality of public services	1.703	3	1.093
Number of full-time workers	1.580	1	1.257
Number of part-time workers	1.292	1	1.136
Household with children	2.988	1	1.729
Carsharing available	1.085	1	1.042
Multiple carsharing options available	1.044	1	1.022
Average trip length	1.125	1	1.061
Number of motorbikes	1.093	1	1.045
Number of bikes/pedelecs	1.739	1	1.319
Older generation	1.137	1	1.066
Type of housing	1.503	3	1.070
Quality of public transport	2.994	3	1.201
Distance to nearest trainstation	1.952	3	1.118
Distance to nearest bus stop	1.590	3	1.080
Distance to next metrostation	3.034	3	1.203
Availability of a garage	1.025	1	1.013
Average number of trips	1.055	1	1.027
Householdsize	5.411	1	2.326

[1] Degrees of freedom.

Appendix 2: Overview of variance inflation factors for predictor variables

Appendix 3: Forward selection process of the preliminary model

Variable	Degress of freedom	AIC-Value
Start	0	160148.9
Number of licenses	3	125067
Region	6	120597
Monthly Income	3	114830
Quality of public services	9	112552
Carsharing available	3	111294
Number of full-time workers	3	110021
Multiple Carsharing options available	3	109313
Number of motorbikes	3	108620
Type of housing	9	108133
Average trip length	3	107640
Quality of public transport	9	107287
Older generation	3	106952
Number of part-time workers	3	106699
Average number of trips	3	106506
Household with children	3	106400
Number of bikes/pedelecs	3	106305
Distance to nearest metro station	9	106246
Distance to nearest train station	9	106182
Availability of a garage	3	106160
Distance to nearest bus stop	9	106144
End	0	106144

Appendix 3: Forward selection process of the preliminary Model

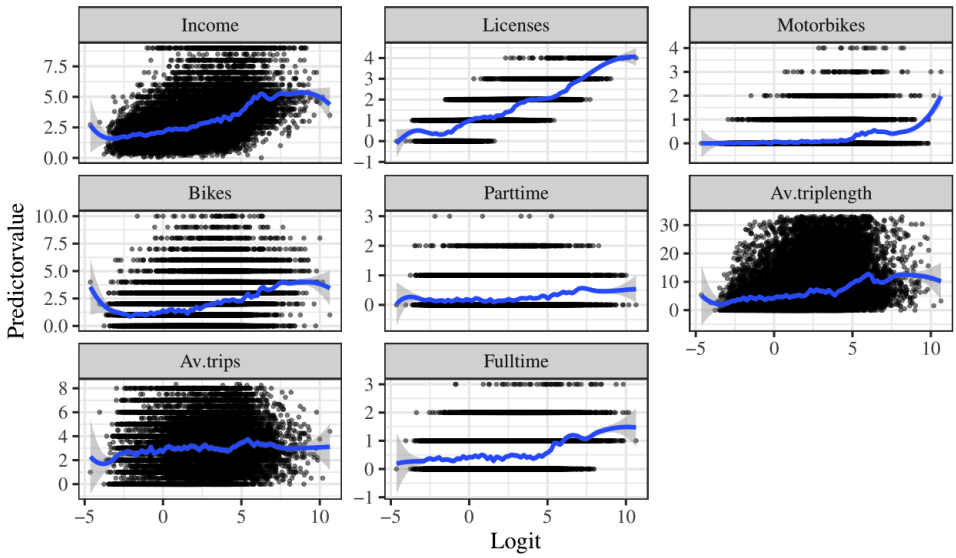
Appendix 4: Overview of coefficients for separate logits and the preliminary model

Variable	Coeff. Logit1	Coeff. MNL:1	Coef. Logit2	Coeff. MNL:2	Coeff. Logit3	Coeff. MNL:3
(Intercept)	-2.618	-2.674	-8.794	-8.384	-10.448	-14.837
licenses	1.999	2.017	3.268	3.476	2.550	4.841
suburban	0.812	0.793	1.319	1.340	1.437	1.582
rural	0.905	0.888	1.575	1.500	1.772	1.846
income	0.291	0.292	0.683	0.655	0.726	0.810
ps.2	-0.114	-0.123	-0.363	-0.293	-0.470	-0.297
ps.3	-0.447	-0.445	-0.856	-0.823	-1.192	-0.998
ps.4	-0.790	-0.778	-1.299	-1.360	-1.730	-1.567
carsharing	-1.893	-1.855	-2.898	-2.850	-3.275	-3.225
full-time	0.120	0.101	0.468	0.582	0.544	0.869
m.carsharing	-2.088	-2.052	-3.102	-2.964	-3.907	-3.329
motorbikes	0.436	0.497	1.016	0.799	1.330	1.173
multifamily	-0.479	-0.473	-0.890	-0.786	-1.046	-0.852
apartment	-0.574	-0.572	-1.174	-1.036	-1.486	-1.084
other	-0.575	-0.578	-0.949	-0.732	-0.743	-0.510
av.triplength	0.034	0.034	0.053	0.058	0.046	0.065
pt.2	-0.017	-0.002	0.171	0.047	0.134	-0.031
pt.3	-0.153	-0.129	-0.044	-0.252	-0.328	-0.429
pt.4	-0.325	-0.301	-0.446	-0.629	-0.507	-0.646
old	0.845	0.857	1.205	0.960	1.695	1.125
part-time	-0.257	-0.272	0.023	0.030	0.027	0.152
av.trips	0.122	0.119	0.136	0.145	0.224	0.167
children	0.322	0.291	0.428	0.424	0.256	0.033
bikes	-0.110	-0.102	-0.086	-0.146	-0.009	-0.157
metro.2	0.201	0.200	0.480	0.505	0.456	0.450
metro.3	0.337	0.344	0.791	0.722	0.652	0.839
metro.4	0.328	0.331	0.691	0.654	0.745	0.745
train.2	0.105	0.111	0.318	0.167	0.122	0.205
train.3	0.248	0.244	0.623	0.445	0.300	0.559
train.4	0.180	0.167	0.434	0.393	0.105	0.643
garage	0.105	0.115	0.425	0.293	0.267	0.396
bus.3	0.157	0.154	0.132	0.256	0.144	0.333
bus.3	0.163	0.166	0.364	0.275	0.244	0.395
bus.4	0.080	0.078	0.173	0.244	0.370	0.373

Coeff. = Coefficients

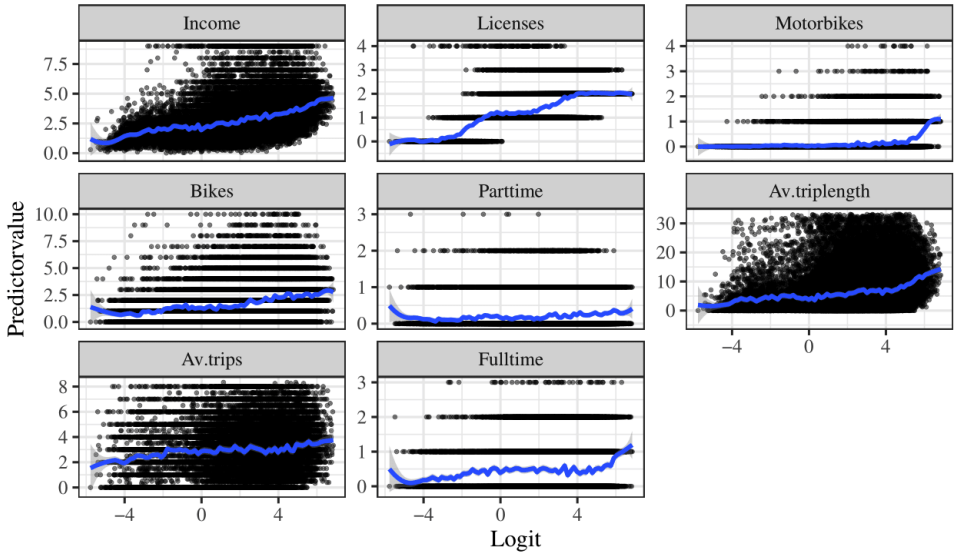
Appendix 4: Overview of coefficients of separate logits and the preliminary model

Appendix 5: Logit(One Car) - Relation between explanatory variables and the logit before transformations



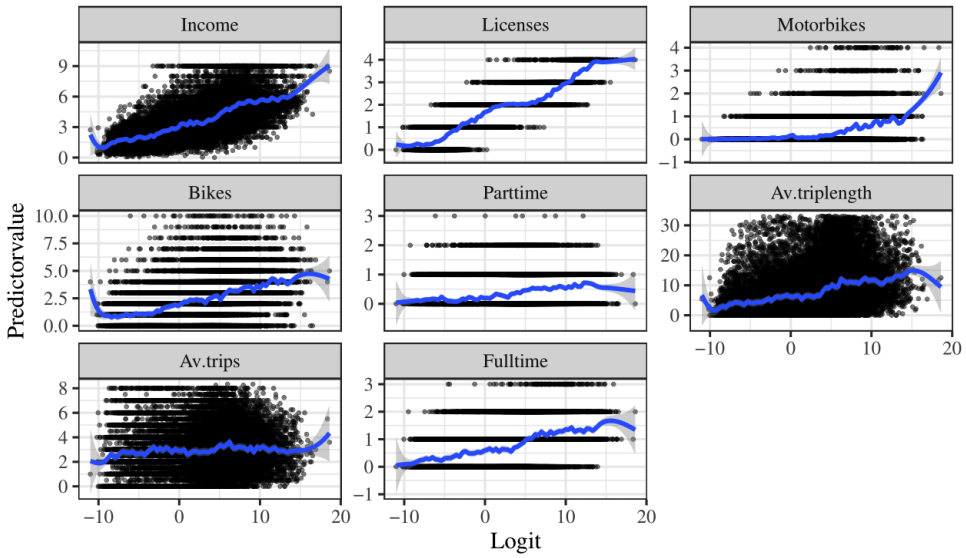
Appendix 5: Logit(One Car) - Relation between explanatory variables and the logit before transformations

Appendix 6: Logit(One Car) - Relation between explanatory variables and the logit after transformations



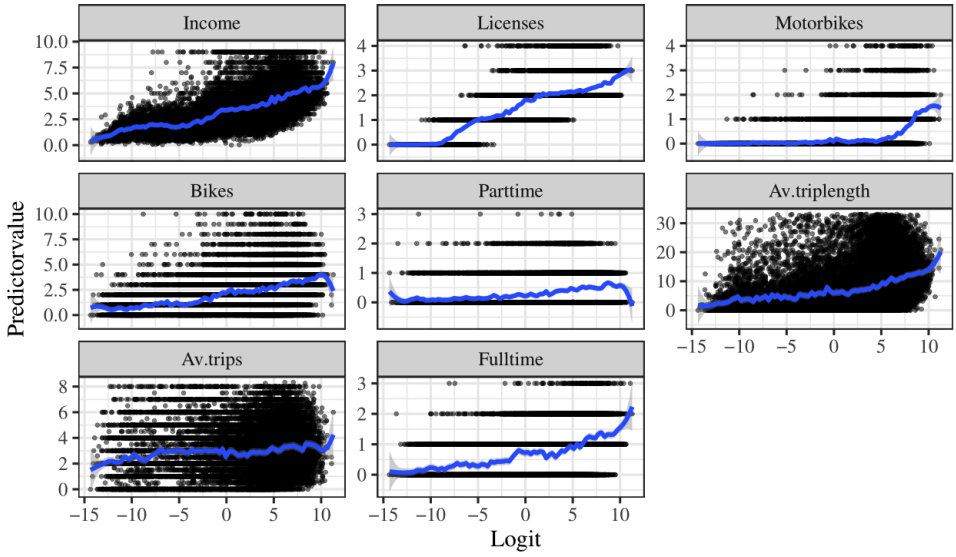
Appendix 6: Logit(One Car) - Relation between explanatory variables and the logit after transformations

Appendix 7: Logit(Two Cars) - Relation between explanatory variables and the logit before transformations



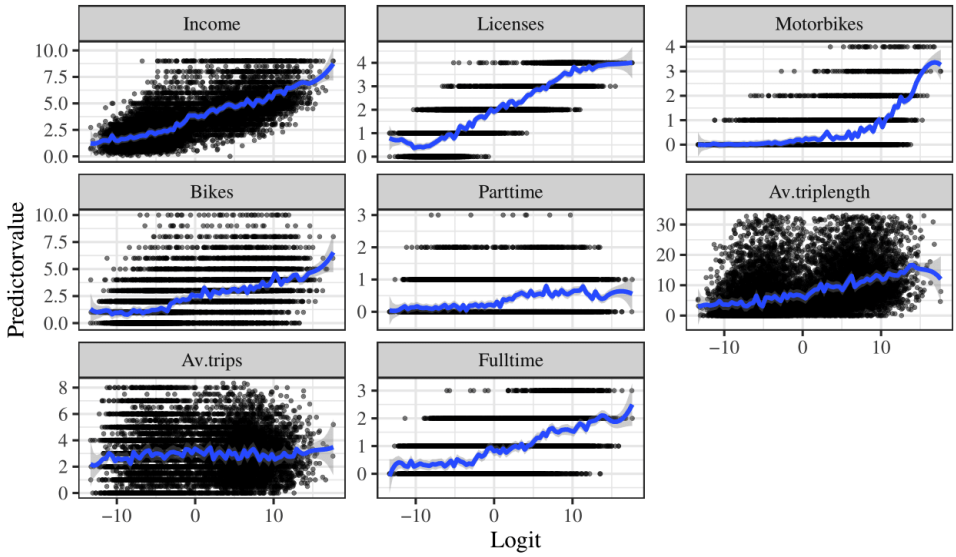
Appendix 7: Logit(Two Cars) - Relation between explanatory variables and the logit before transformations

Appendix 8: Logit(Two Cars) - Relation between explanatory variables and the logit after transformations



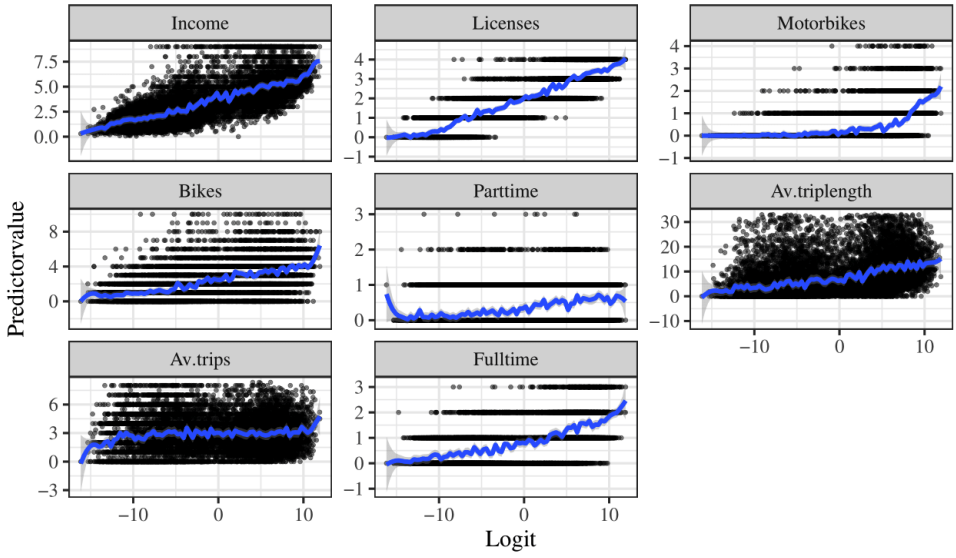
Appendix 8: Logit(Two Cars) - Relation between explanatory variables and the logit after transformations

Appendix 9: Logit(Three Cars) - Relation between explanatory variables and the logit before transformations



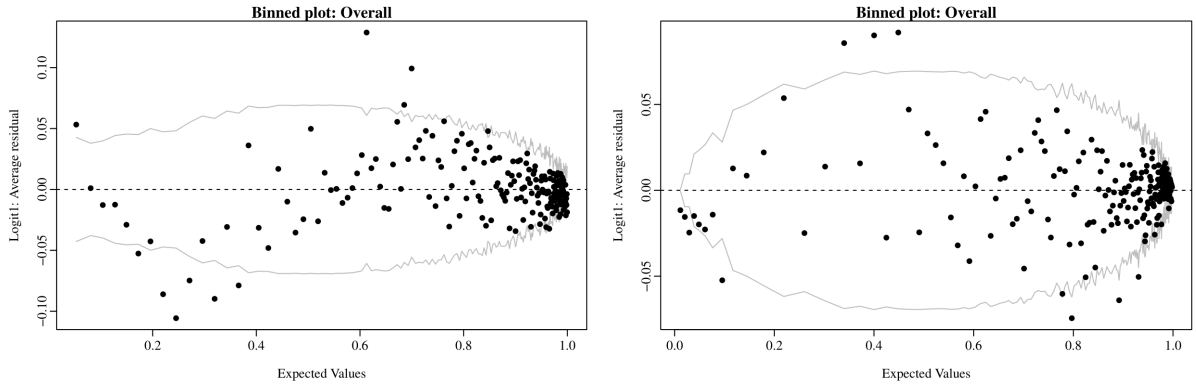
Appendix 9: Overview of variance inflation factors for predictor variables

Appendix 10: Logit(Three Cars) - Relation between explanatory variables and the logit after transformations



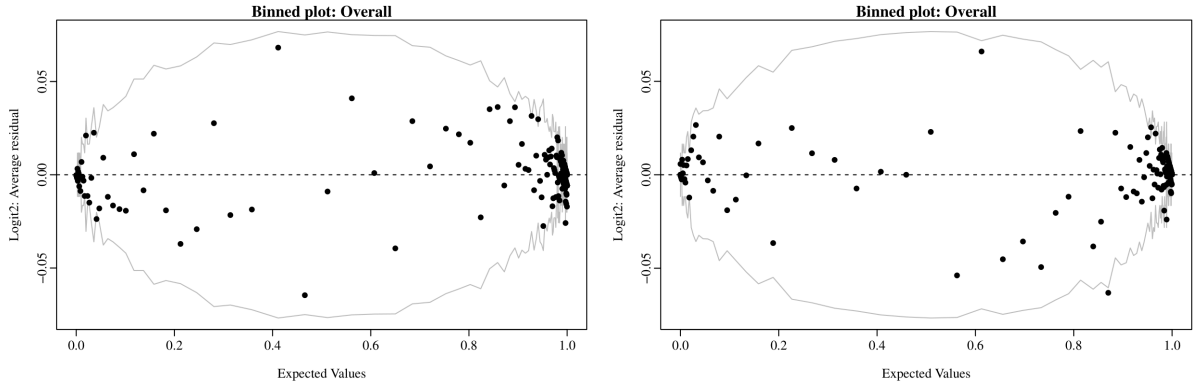
Appendix 10: Logit(Three Cars) - Relation between explanatory variables and the logit after transformations

Appendix 11: Logit(One car) - Binned Plots



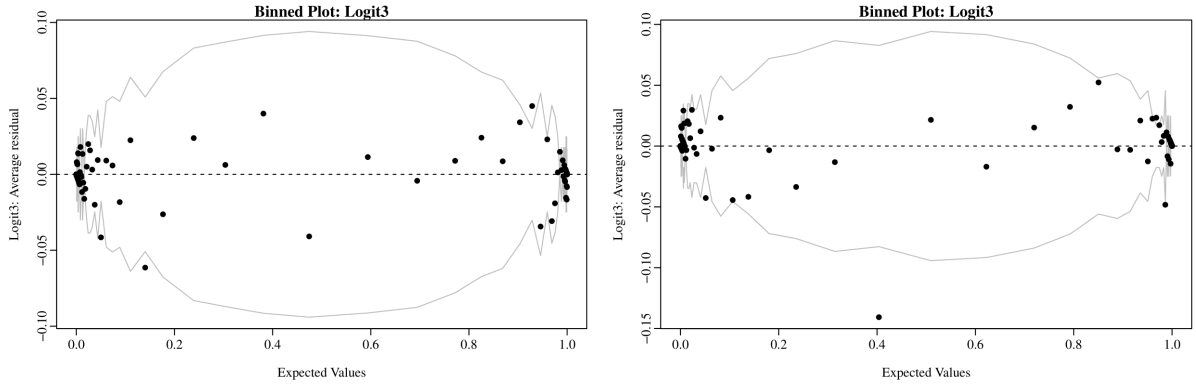
Appendix 11: Logit(One car) - Binned Plots

Appendix 12: Logit(Two cars) - Binned Plots



Appendix 12: Logit(Two cars) - Binned Plots

Appendix 13: Logit(Three cars) - Binned Plots



Appendix 13: Logit(Three cars) - Binned Plots

Appendix 14: Forward selection process of the final model

Variable	Degress of freedom	AIC-Value
Start	0	159839.9
Number of licenses	3	124434
Region	6	119883
Monthly Income	3	114067
Number of licenses squared	3	111111
Quality of public services	9	108781
Monthly Income squared	3	107321
Carsharing available	3	106149
Number of full-time workers	3	105194
Multiple Carsharing options available	3	104522
Number of motorbikes	3	103911
Average trip length	3	103461
Type of housing	9	103029
Quality of public transport	9	102698
Average trip length squared	3	102384
Older generation	3	102120
Number of part-time workers	3	101974
Household with children	3	101876
Average number of trips	3	101816
Number of bikes/pedelecs	3	101750
Distance to nearest metro station	9	101699
Distance to nearest train station	9	101629
Availability of a garage	3	101609
Distance to nearest bus stop	9	101595
Number of part-time workers squared	3	101583
End	0	101583

Appendix 14: Forward selection process of the final model

Appendix 15: Log-odds of the final MNL model

Variable	Log-odds: 1 vs. 0 cars	Log-odds: 2 vs. 0 cars	Log-odds: 3 vs. 0 cars
(Intercept)	0.016	0.000	0.000
licenses	111.358	5942.412	60619.880
suburban	2.440	4.227	5.281
rural	2.938	5.548	7.684
income	2.195	6.196	7.306
licenses squared	0.362	0.195	0.169
ps.2	0.844	0.700	0.700
ps.3	0.603	0.407	0.342
ps.4	0.431	0.243	0.200
income squared	0.943	0.878	0.878
carsharing	0.170	0.072	0.056
full-time	0.978	1.486	1.920
m.carsharing	0.137	0.062	0.047
motorbikes (log)	1.078	1.122	1.190
av.triplength	1.112	1.201	1.215
multifamily	0.597	0.459	0.434
apartmentbuilding	0.548	0.365	0.365
other	0.532	0.472	0.585
pt.2	1.009	1.042	0.959
pt.3	0.870	0.759	0.637
pt.4	0.762	0.553	0.566
av.triplength squared	0.997	0.995	0.995
old	2.130	2.364	2.665
part-time	1.132	1.575	1.789
children	1.210	1.272	0.852
bikes	0.945	0.898	0.896
av.trips	1.078	1.087	1.109
metro.2	1.180	1.546	1.420
metro.3	1.422	1.976	2.100
metro.4	1.443	1.973	2.062
train.2	1.106	1.162	1.181
train.3	1.296	1.585	1.729
train.4	1.223	1.555	1.952
garage	1.116	1.332	1.482
bus.2	1.144	1.263	1.353
bus.3	1.259	1.412	1.568
bus.4	1.043	1.277	1.441
part-time squared	0.744	0.702	0.688

Appendix 15: Log-odds of the final MNL model

Appendix 16: Model validation measures of the model including interactions

Test	χ^2	Degrees of freedom	p-value
Null model: Likelihood Ratio test	59541	276	<0.0001
Null model: Score test	27842	273	< 0.0001
Null model: Wald test	25469	276	<0.0001
Base model: Likelihood Ratio test	1061.7	165	<0.0001
Base model: Score test	1031	165	< 0.0001
Base model: Wald test	1000	165	<0.0001
Measure	Value		
AIC: intercept only	159.840		
AIC: base model	101.583		
AIC: interaction model	100.851		
McFadden	0.3725		
Cox and Snell	0.5712		
Nagelkerke	0.6368		

Appendix 16: Model validation measures of the MNL model including interactions

Appendix 17: Coefficients of the MNL model including interactions

	1 vs. 0 cars			2 vs. 0 cars			3+ vs. 0 cars		
Variable	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.
(Intercept)	-2.976	0.272	***	-11.859	0.373	***	-16.400	0.759	***
suburban	-1.466	0.430	***	-0.959	0.537	*	-3.625	0.967	***
rural	-2.379	0.730	***	-2.397	0.838	***	-5.799	1.246	***
licenses	3.958	0.108	***	7.939	0.196	***	9.050	0.459	***
licenses squared	-0.822	0.030	***	-1.445	0.043	***	-1.434	0.083	***
income	0.689	0.049	***	1.708	0.071	***	1.749	0.135	***
income squared	-0.048	0.005	***	-0.115	0.007	***	-0.104	0.012	***
ps.2	-0.270	0.153	*	-0.492	0.167	***	-0.586	0.200	***
ps.3	-0.629	0.157	***	-1.095	0.173	***	-1.247	0.220	***
ps.4	-0.970	0.163	***	-1.659	0.188	***	-1.881	0.273	***
carsharing	-1.593	0.067	***	-2.304	0.094	***	-2.454	0.172	***
full-time	0.127	0.085		0.616	0.113	***	0.654	0.179	***
workers squared	-0.114	0.045	**	-0.239	0.055	***	-0.183	0.077	**
av.triplength	0.090	0.010	***	0.175	0.013	***	0.206	0.021	***
av.triplength squared	-0.003	0.000	***	-0.005	0.000	***	-0.006	0.001	***
motorbikes (log)	0.077	0.010	***	0.119	0.011	***	0.182	0.013	***
multifamily	-0.534	0.065	***	-0.805	0.077	***	-0.935	0.116	***
apartmentbuilding	-0.583	0.092	***	-0.938	0.128	***	-0.990	0.257	***
other	-0.614	0.074	***	-0.848	0.093	***	-0.838	0.153	***
garage	0.189	0.103	*	0.535	0.119	***	0.720	0.159	***
pt.2	-0.056	0.149		-0.033	0.155		-0.125	0.166	
pt.3	-0.209	0.154		-0.349	0.162	**	-0.508	0.178	***
pt.4	-0.330	0.161	**	-0.616	0.176	***	-0.653	0.216	***
old	0.758	0.052	***	1.042	0.074	***	1.587	0.164	***
m.carsharing	-1.814	0.096	***	-2.459	0.131	***	-2.722	0.240	***
children	0.234	0.070	***	0.135	0.082	*	-0.164	0.117	
metro.2	0.136	0.050	***	0.346	0.065	***	0.267	0.109	**
metro.3	0.292	0.085	***	0.539	0.100	***	0.607	0.145	***
metro.4	0.331	0.062	***	0.574	0.076	***	0.620	0.119	***
av.trips	-0.048	0.031		-0.157	0.038	***	-0.324	0.054	***
tripsavg squared	0.016	0.004	***	0.031	0.005	***	0.059	0.007	***
part-time	-0.020	0.118		0.036	0.133		0.120	0.165	
part-time squared	-0.225	0.078	***	-0.230	0.085	***	-0.306	0.099	***
train.2	0.105	0.058	*	0.201	0.082	**	-0.095	0.137	
train.3	0.242	0.083	***	0.357	0.107	***	0.098	0.169	
train.4	0.572	0.382		1.025	0.405	**	0.741	0.466	
bikes	-0.006	0.032		-0.041	0.041		-0.093	0.065	
numped squared	-0.012	0.005	**	-0.017	0.006	***	-0.008	0.008	
bus.2	0.143	0.063	**	0.244	0.069	***	0.296	0.082	***
bus.3	0.350	0.173	**	0.477	0.180	***	0.573	0.193	***
bus.4	0.053	0.201		0.254	0.211		0.355	0.226	
suburban:licenses	1.683	0.204	***	1.786	0.294	***	3.240	0.592	***
rural:licenses	1.847	0.413	***	1.662	0.500	***	3.583	0.781	***
suburban:licenses_squared	-0.406	0.057	***	-0.415	0.070	***	-0.611	0.112	***
rural:licenses_squared	-0.357	0.167	**	-0.296	0.177	*	-0.534	0.205	***
suburban:income	0.278	0.096	***	0.265	0.117	**	0.472	0.180	***

Appendix 17: Coefficients of the MNL model including interactions

	1 vs. 0 cars			2 vs. 0 cars			3+ vs. 0 cars		
Variable	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.	Coeff.	Std. Err.	Sig.
rural:income	0.309	0.193		0.404	0.210	*	0.547	0.259	**
suburban:income_squared)	-0.030	0.013	**	-0.029	0.014	**	-0.050	0.018	***
rural:income_squared	-0.040	0.032		-0.055	0.033	*	-0.066	0.035	*
suburban:ps.2	0.171	0.290		0.273	0.304		0.314	0.335	
rural:ps.2	0.358	0.296		0.376	0.311		0.552	0.340	
suburban:ps.3	0.335	0.289		0.503	0.305	*	0.400	0.345	
rural:ps.3	0.446	0.307		0.561	0.325	*	0.633	0.366	*
suburban:ps.4	0.304	0.295		0.492	0.318		0.486	0.389	
rural:ps.4	0.839	0.370	**	1.112	0.400	***	1.105	0.481	**
suburban:carsharing	-0.874	0.230	***	-1.171	0.263	***	-1.383	0.373	***
rural:carsharing	-0.549	0.713		-0.372	0.774		-0.359	0.861	
suburban:workers	0.203	0.211		0.272	0.230		0.082	0.283	
rural:workers	0.141	0.810		0.501	0.819		0.495	0.841	
suburban:workers_squared	0.031	0.143		0.117	0.149		0.229	0.161	
rural:workers_squared	0.450	0.706		0.453	0.709		0.534	0.712	
suburban:av.triplength	0.061	0.018	***	0.057	0.021	***	0.038	0.029	
rural:av.triplength	0.040	0.035		0.027	0.037		0.044	0.043	
suburban:av.triplength_squared	-0.002	0.001	***	-0.002	0.001	***	-0.001	0.001	
rural:av.triplength_squared	-0.000	0.001		-0.000	0.001		-0.001	0.002	
suburban:motorbikes	-0.016	0.021		-0.012	0.021		-0.012	0.023	
rural:motorbikes	-0.055	0.036		-0.084	0.036	**	-0.094	0.038	**
suburban:multifamily	0.074	0.101		0.086	0.115		0.119	0.162	
rural:multifamily	-0.146	0.176		-0.174	0.199		0.033	0.252	
suburban:apartment	-0.124	0.163		-0.278	0.218		-0.259	0.400	
rural:apartment	0.101	0.418		-0.132	0.573		-0.489	1.260	
suburban:other	-0.152	0.131		0.050	0.155		0.329	0.219	
rural:other	0.777	0.339	**	1.216	0.360	***	1.408	0.402	***
suburban:garage	-0.223	0.166		-0.474	0.183	***	-0.548	0.223	**
rural:garage	-0.390	0.340		-0.777	0.359	**	-0.978	0.397	**
suburban:generation	-0.200	0.148		-0.648	0.171	***	-1.113	0.251	***
rural:generation	-0.089	0.406		-0.567	0.429		-1.392	0.475	***
suburban:children	-0.006	0.191		0.154	0.199		-0.008	0.223	
rural:children	0.084	0.450		0.332	0.458		0.275	0.471	
suburban:tripsavg	0.025	0.021		0.027	0.025		0.031	0.036	
rural:tripsavg	0.028	0.044		0.075	0.047		0.059	0.056	
suburban:part-time	0.222	0.116	*	0.439	0.125	***	0.599	0.151	***
rural:part-time	0.467	0.263	*	0.873	0.270	***	1.181	0.286	***
suburban:train.2	-0.044	0.109		-0.137	0.132		0.384	0.193	**
rural:train.2	0.063	0.246		0.109	0.273		0.268	0.334	
suburban:train.3	0.048	0.149		0.161	0.173		0.694	0.236	***
rural:train.3	0.191	0.286		0.427	0.312		0.685	0.374	*
suburban:train.4	-0.429	0.408		-0.637	0.434		0.063	0.501	
rural:train.4	-0.097	0.451		-0.213	0.482		0.222	0.554	
suburban:bikes	0.144	0.061	**	0.192	0.069	***	0.229	0.093	**
rural:bikes	0.129	0.135		0.257	0.142	*	0.238	0.159	
suburban:bikes_squared	-0.021	0.010	**	-0.027	0.011	**	-0.037	0.013	***
rural:bikes_squared	0.001	0.029		-0.005	0.030		0.003	0.031	

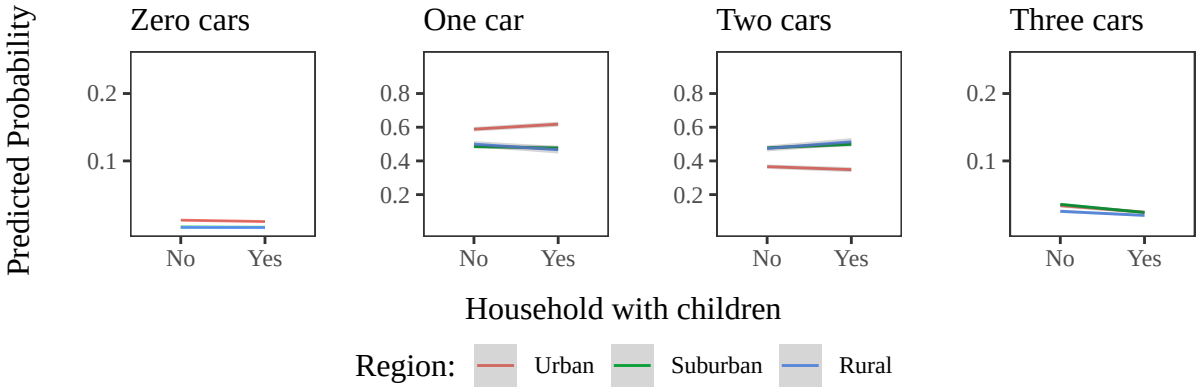
Appendix 17: Coefficients of the MNL Model including interactions (Continued)

Appendix 18: AMEs for interactions between independent variables and the variable "region"

	AME: 0 cars			AME: 1 car			AME: 2 cars			AME: 3+ cars		
Variable	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural	Urban	Suburban	Rural
carsharing	0.158	0.166	0.106	-0.048	-0.016	-0.032	-0.093	-0.120	-0.062	-0.018	-0.030	-0.012
garage	-0.015	0.001	0.006	-0.037	-0.015	-0.001	0.038	0.007	-0.005	0.014	0.007	-0.001
children	-0.015	-0.008	-0.009	0.031	0.005	-0.007	-0.003	0.024	0.032	-0.013	-0.021	-0.015
apartment	0.043	0.028	0.015	0.011	0.043	0.066	-0.046	-0.063	-0.057	-0.008	-0.009	-0.023
multifamily	0.039	0.017	0.021	0.005	0.021	0.017	-0.034	-0.030	-0.039	-0.010	-0.008	0.000
other	0.045	0.029	-0.004	-0.008	-0.027	-0.024	-0.033	-0.017	0.016	-0.004	0.015	0.012
income	-0.034	-0.025	-0.020	-0.036	-0.046	-0.046	0.057	0.057	0.053	0.013	0.014	0.014
licenses	-0.134	-0.111	-0.104	-0.077	-0.122	-0.122	0.148	0.140	0.123	0.063	0.093	0.102
bikes	0.003	-0.002	-0.003	0.005	0.008	-0.002	-0.008	-0.005	0.003	-0.001	-0.002	0.002
generation	-0.064	-0.023	-0.023	0.012	0.042	0.049	0.028	-0.022	-0.010	0.024	0.002	-0.016
part-time	0.007	-0.006	-0.013	-0.013	-0.033	-0.050	0.005	0.027	0.042	0.001	0.012	0.020
ps.2	0.017	0.004	-0.002	0.018	0.014	0.027	-0.027	-0.013	-0.027	-0.008	-0.004	0.002
ps.3	0.042	0.011	0.006	0.030	0.034	0.040	-0.058	-0.028	-0.039	-0.015	-0.017	-0.007
ps.4	0.071	0.027	0.004	0.035	0.046	0.051	-0.085	-0.055	-0.041	-0.021	-0.018	-0.015
train.2	-0.018	-0.011	-0.013	0.005	-0.024	-0.032	0.022	0.019	0.042	-0.009	0.016	0.004
train.3	-0.008	-0.002	-0.006	-0.001	-0.001	-0.011	0.020	-0.007	0.022	-0.011	0.010	-0.005
train.4	-0.040	-0.006	-0.015	-0.022	-0.033	-0.031	0.067	0.016	0.034	-0.005	0.023	0.012
av.trip length	-0.004	-0.004	-0.003	-0.003	-0.002	-0.003	0.006	0.005	0.004	0.001	0.001	0.001
workers	-0.003	-0.012	-0.014	-0.038	-0.058	-0.076	0.031	0.052	0.067	0.011	0.018	0.024

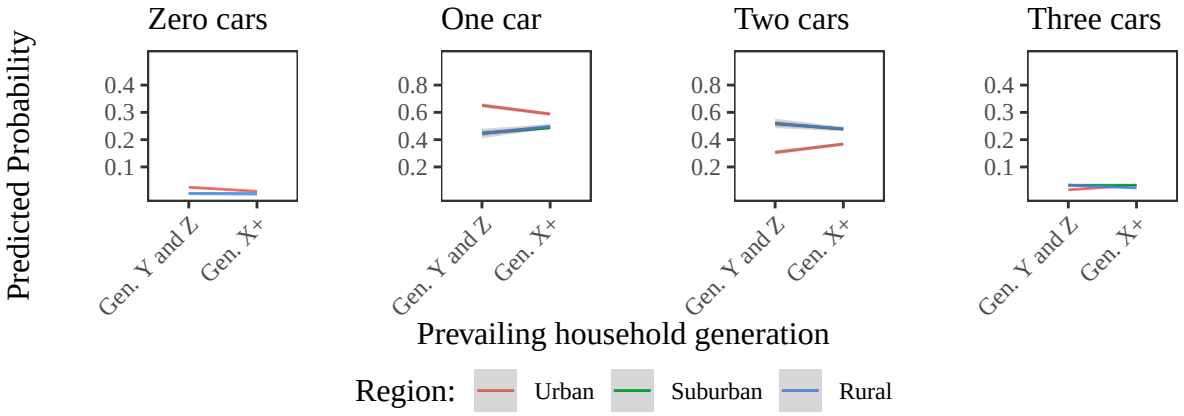
Appendix 18: AMEs for interactions between independent variables and the variable "region"

Appendix 19: Complete effect plots for the interaction between the variables ”children” and ”region”



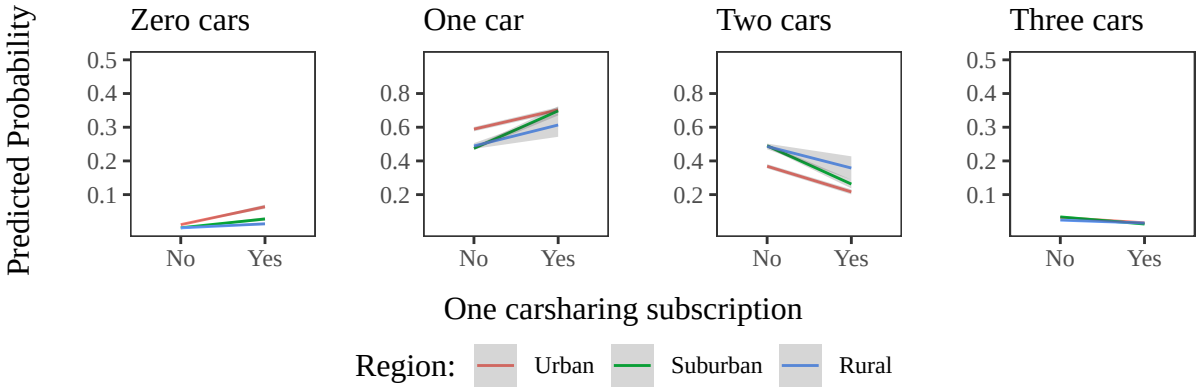
Appendix 19: Complete effect plots for the interaction between the variables ”children” and ”region”

Appendix 20: Complete effect plots for the interaction between the variables ”generation” and ”region”



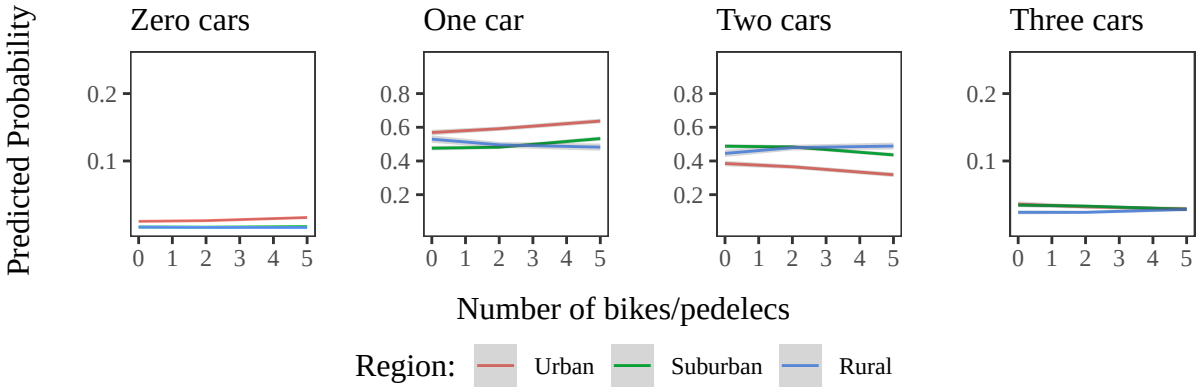
Appendix 20: Complete effect plots for the interaction between the variables ”generation” and ”region”

Appendix 21: Complete effect plots for the interaction between the variables ”carsharing” and ”region”



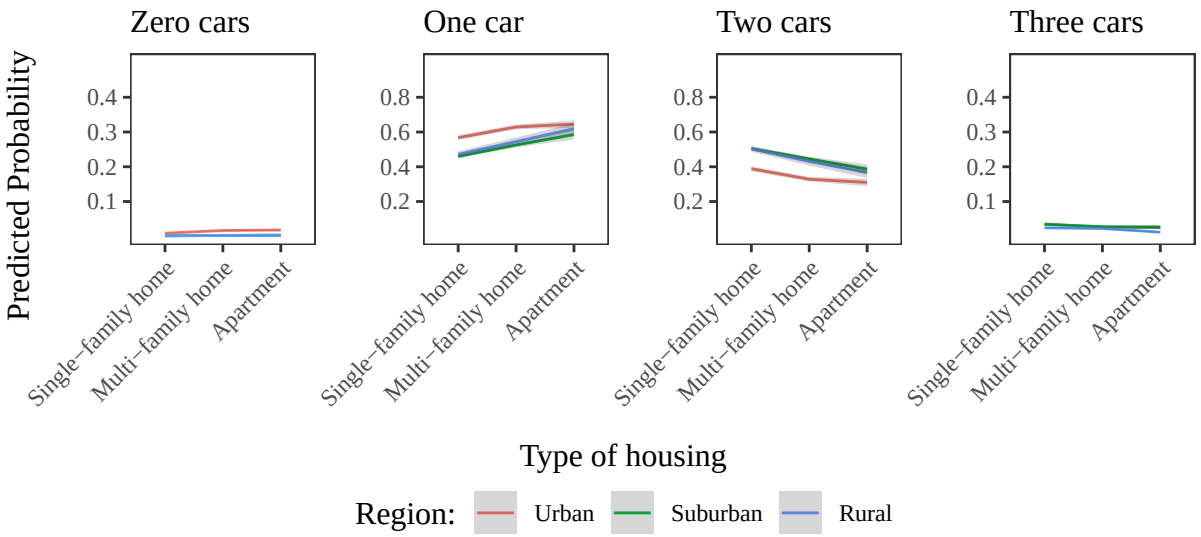
Appendix 21: Complete effect plots for the interaction between the variables ”carsharing” and ”region”

Appendix 22: Complete effect plots for the interaction between the variables ”bikes” and ”region”



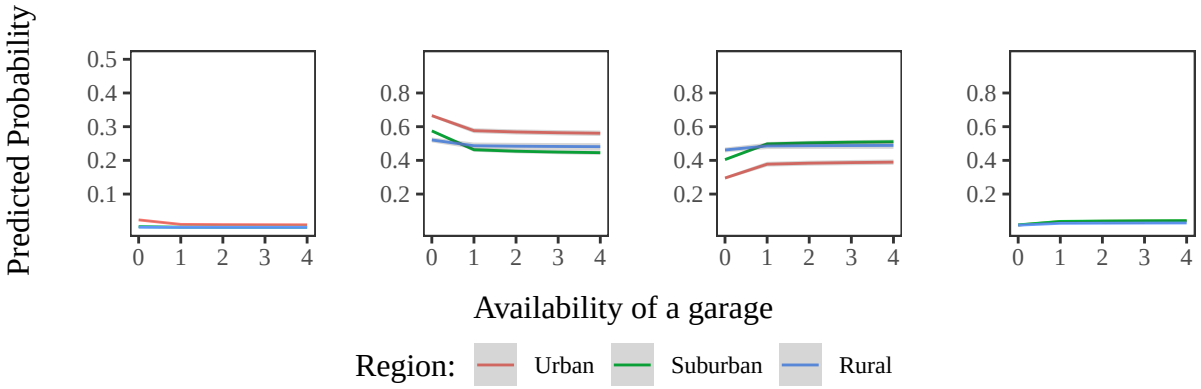
Appendix 22: Complete effect plots for the interaction between the variables ”bikes” and ”region”

Appendix 23: Complete effect plots for the interaction between the variables ”housing” and ”region”



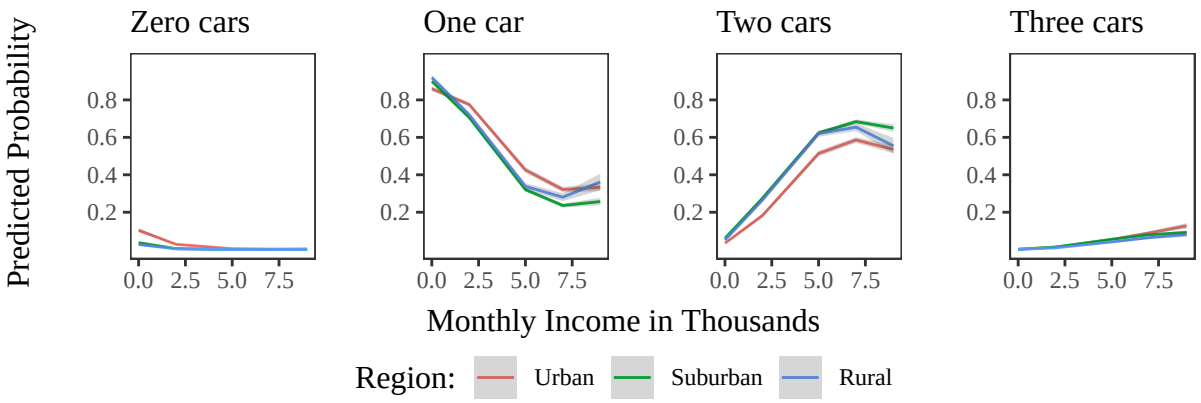
Appendix 23: Complete effect plots for the interaction between the variables ”housing” and ”region”

Appendix 24: Complete effect plots for the interaction between the variables ”garage” and ”region”



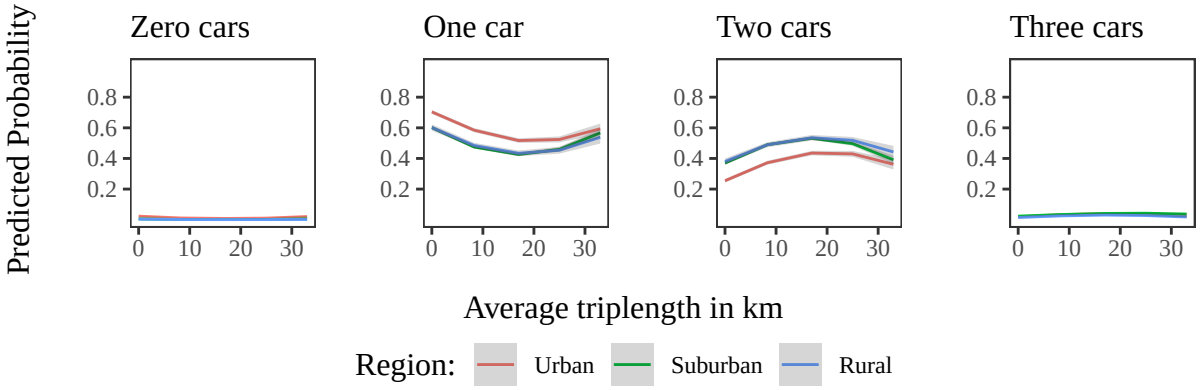
Appendix 24: Complete effect plots for the interaction between the variables ”garage” and ”region”

Appendix 25: Complete effect plots for the interaction between the variables ”income” and ”region”



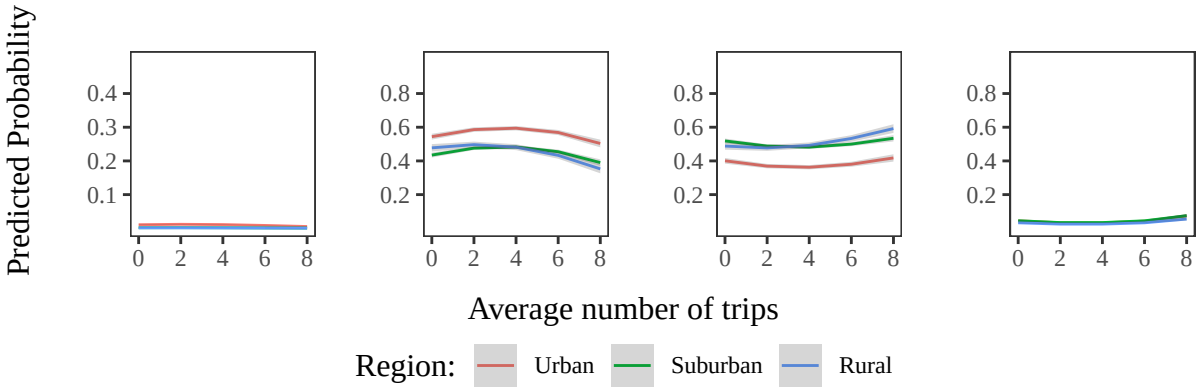
Appendix 25: Complete effect plots for the interaction between the variables ”income” and ”region”

Appendix 26: Complete effect plots for the interaction between the variables ”triplength.av” and ”region”



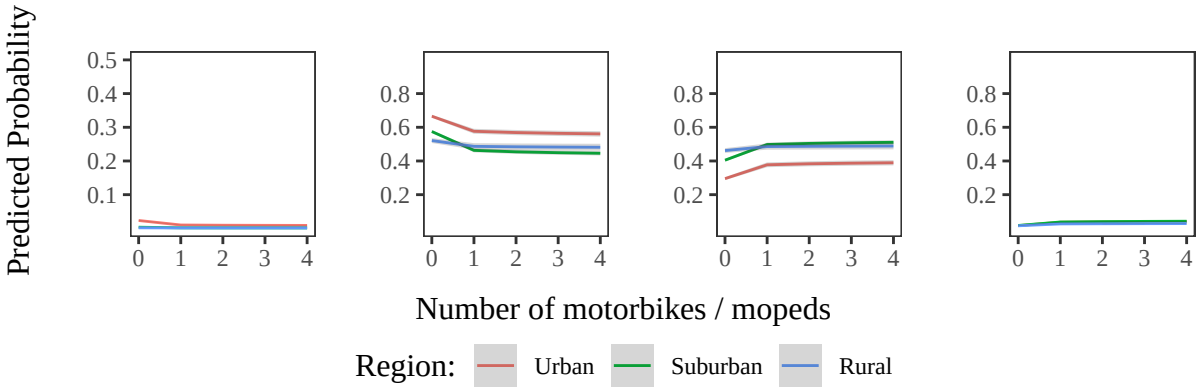
Appendix 26: Complete effect plots for the interaction between the variables ”triplength.av” and ”region”

Appendix 27: Complete effect plots for the interaction between the variables ”trips.av” and ”region”



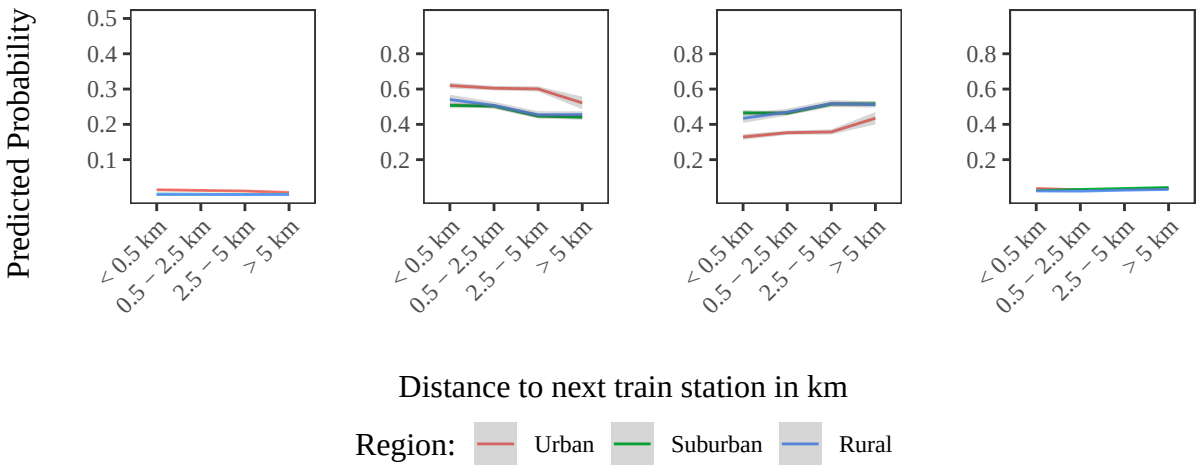
Appendix 27: Complete effect plots for the interaction between the variables ”trips.av” and ”region”

Appendix 28: Complete effect plots for the interaction between the variables ”motorbikes” and ”region”



Appendix 28: Complete effect plots for the interaction between the variables ”motorbikes” and ”region”

Appendix 29: Complete effect plots for the interaction between the variables ”train” and ”region”



Appendix 29: Complete effect plots for the interaction between the variables ”trainf” and ”region”

Bibliography

Akoglu, Haldun (2018): User's guide to correlation coefficients, in: Turkish Journal of Emergency Medicine, Vol. 18, No. 3, pp. 91–93.

Anowar, Sabreena / Eluru, Naveen / Miranda-Moreno, Luis F. (2014): Alternative modeling approaches used for examining automobile ownership: a comprehensive review, in: Transport Reviews, Vol. 34, No. 4, pp. 441–473.

Ao, Yibin / Yang, Dujuan / Chen, Chuan / Wang, Yan (2019): Exploring the effects of the rural built environment on household car ownership after controlling for preference and attitude: Evidence from Sichuan, China, in: Journal of Transport Geography, Vol. 74, pp. 24–36.

Banister, David (2007): Cities, Mobility, and Climate Change, in: Journal of Industrial Ecology, Vol. 11, No. 2, pp. 7–10.

Begg, Colin B. / Gray, Robert (1984): Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions, in: Biometrika, Vol. 71, No. 1, pp. 11–18.

Ben-Akiva, Moshe / Lerman, Steven R. (1985): Discrete Choice Analysis : Theory and Application to Travel Demand, Cambridge et al. 1985.

Bento, Antonio / Cropper, Maureen / Mobarak, Ahmed Mushfiq / Vinha, Katja (2005): The Effects of Urban Spatial Structure on Travel Demand in the United States, in: Review of Economics & Statistics, Vol. 87, No. 3, pp. 466–478.

Berry, William / Golder, Matt / Milton, Daniel (2012): Improving tests of theories positing interaction, in: The Journal of Politics, Vol. 74, No. 3, pp. 653–671.

Best, Henning / Lanzendorf, Martin (2005): Division of labour and gender differences in metropolitan car use: An empirical study in Cologne, Germany, in: Journal of Transport Geography, Vol. 13, No. 2, pp. 109–121.

Bhat, Chandra / Guo, Jessica (2007): A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels, in: Transportation Research Part B: Methodological, Vol. 41, pp. 506–526.

Bhat, Chandra R. / Pulugurta, Vamsi (1998): A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions, in: Transportation Research: Part B, Vol. 32, No. 1, pp. 61–75.

Borra, Simone / Di Ciaccio, Agostino (2010): Measuring the prediction error: A comparison of cross-validation, bootstrap and covariance penalty methods, in: Computational Statistics & Data Analysis, Vol. 54, No. 12, pp. 2976–2989.

Breiman, Leo (2001): Random Forests, in: Machine Language, Vol. 45, No. 1, pp. 5–32.

Breiman, Leo (2007): Manual Setting Up, Using, And Understanding Random Forests V4.0, URL:
https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf, 10/12/2020, 02.53 p.m.

Breiman, Leo / Cutler, Adele (2020): Random Forests, URL:
https://www.stat.berkeley.edu/~breiman/RandomForests/cc_manual.htm, 3/12/2020, 05.51 p.m.

Brodersen, Kay Henning / Ong, Cheng Soon / Stephan, Klaas Enno / Buhmann, Joachim M. (2010): The Balanced Accuracy and Its Posterior Distribution, in: 20th International Conference on Pattern Recognition, ed. Institute of Electrical and Electronics Engineers, California et al., 2010, pp. 3121–3124.

Buehler, Ralph / Kunert, Uwe (2010): Determinanten und Perspektiven des Verkehrsverhaltens: USA und Deutschland im Vergleich, in: Internationales Verkehrswesen, Vol. 62, No. 6, pp. 16–21.

Buehler, Ralph / Pucher, John (2009): Sustainable Transport that Works: Lessons from Germany, in: World Transport Policy and Practice, Vol. 15, No. 1, pp. 13–46.

Bund-Länder Demografie Portal (2020): Zukünftige regionale Bevölkerungsentwicklung, URL: https://www.demografie-portal.de/DE/Fakten/bevoelkerungsentwicklung-regional-zukunft.html, 13/12/2020, 11.45 a.m.

Bundesministerium für Verkehr und digitale Infrastruktur (2020): Regionalstatistische Raumtypologie, URL: https://www.bmvi.de/SharedDocs/DE/Artikel/G/regionalstatistische-raumtypologie.html, 13/12/2020, 09.43 a.m.

Bundesministerium für Verkehr und Infrastruktur (2019a): Mobilität in Deutschland - Kurzreport, URL: http://www.mobilitaet-in-deutschland.de/pdf/infas_Mobilitaet_in_Deutschland_2017_Kurzreport.pdf, 13/12/2020, 09.31 p.m.

Bundesministerium für Verkehr und Infrastruktur (2019b): Mobilität in Deutschland - Methodenbericht, URL: http://www.mobilitaet-in-deutschland.de/pdf/MiD2017_Methodenbericht.pdf, 13/12/2020, 02.14 p.m.

Bundesministerium für Verkehr und Infrastruktur (2019c): Mobilität in Deutschland Nutzerhandbuch: Dokumentation Raumvariablen, URL: https://daten.clearingstelle-verkehr.de/279/52/MiD2017_SonstigeRaumvariablen.pdf, 13/12/2020, 11.13 a.m.

Bundesverband Carsharing (2020): Aktuelle Zahlen und Fakten zum CarSharing in Deutschland, URL: https://carsharing.de/alles-ueber-carsharing/carsharing-zahlen/aktuelle-zahlen-fakten-zum-carsharing-deutschland, 28/12/2020, 01.11 p.m.

Cantarella, Giulio Erberto / De Luca, Stefano (2005): Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models, in: Transportation Research Part C: Emerging Technologies, Vol. 13, No. 2, pp. 121–155.

Caulfield, Brian (2012): An examination of the factors that impact upon multiple vehicle ownership: The case of Dublin, Ireland, in: Transport Policy, Vol. 19, No. 1, pp. 132–138.

Charpentier, Arthur (2015): 'Variable Importance Plot' and Variable Selection, URL: https://www.r-bloggers.com/2015/06/variable-importance-plot-and-variable-selection/, 10/12/2020, 01.08 p.m.

Cheng, Long / Chen, Xuewu / De Vos, Jonas / Lai, Xinjun / Witlox, Frank (2019): Applying a random forest method approach to model travel mode choice behavior, in: Travel Behaviour and Society, Vol. 14, pp. 1–10.

Cheng, Simon / Long, Jeremy Scott (2007): Testing for IIA in the Multinomial Logit Model, in: Sociological Methods & Research, Vol. 35, No. 4, pp. 583–600.

Choudhary, Ravi / Vasudevan, Vinod (2017): Study of vehicle ownership for urban and rural households in India, in: Journal of Transport Geography, Vol. 58, pp. 52–58.

Chu, You-Lian (2002): Automobile ownership analysis using ordered probit models, in: Transportation Research Record, Vol. 1805, No. 1, pp. 60–67.

Cortese, Giuliana (2020): How to use statistical models and methods for clinical prediction, in: Annals of Translational Medicine, Vol. 8, No. 4, pp. 1–3.

Cortina, Jose M. (1993): Interaction, Nonlinearity, and Multicollinearity: implications for Multiple Regression, in: Journal of Management, Vol. 19, No. 4, pp. 915–922.

Dalal, Dev K. / Zickar, Michael J. (2012): Some Common Myths About Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression, in: Organizational Research Methods, Vol. 15, No. 3, pp. 339–362.

Dargay, Joyce M. (2002): Determinants of car ownership in rural and urban areas: a pseudo-panel analysis, in: Transportation Research Part E: Logistics and Transportation Review, Vol. 38, No. 5, pp. 351–366.

De Jong, Gerard / Fox, James / Daly, Andrew / Pieters, Marits / Smit, Remko (2004): Comparison of car ownership models, in: Transport Reviews, Vol. 24, No. 4, pp. 379–408.

De Wolff, Peter (1938): The Demand for Passenger Cars in the United States, in: Econometrica: Journal of the Econometric Society, Vol. 6, No. 2, pp. 113–129.

Eisenmann, Christine / Buehler, Ralph (2018): Are cars used differently in Germany than in California? Findings from annual car-use profiles, in: Journal of Transport Geography, Vol. 69, pp. 171–180.

Eluru, Naveen / Bhat, Chandra R. (2007): A joint econometric analysis of seat belt use and crash-related injury severity, in: Accident Analysis and Prevention, Vol. 39, No. 5, pp. 1037–1049.

European Environment Agency (2011): Passenger car ownership in the EEA, URL: https://www.eea.europa.eu/data-and-maps/figures/passenger-car-ownership-in-the-eea, 3/12/2020, 10.21 a.m.

Fleiss, Joseph L / Levin, Bruce / Paik, Myunghee Cho (2003): Statistical methods for rates and proportions, Third Edition, Hoboken 2003.

Fox, John (1987): Effect Displays for Generalized Linear Models, in: Sociological Methodology, Vol. 17, pp. 347–361.

Fox, John / Hong, Jangman (2009): Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package, in: Journal of Statistical Software, Vol. 32, No. 1, pp. 1–24.

Fox, John / Monette, Georges (1992): Generalized Collinearity Diagnostics, in: Journal of the American Statistical Association, Vol. 87, No. 417, pp. 178–183.

Frondel, Manuel / Vance, Colin (2018): Drivers' response to fuel taxes and efficiency standards: evidence from Germany, in: Transportation, Vol. 45, No. 3, pp. 989–1001.

Gelman, Andrew / Hill, Jennifer (2006): Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge 2006.

Gray, J. Brian / Woodall, William H. (1994): The Maximum Size of Standardized and Internally Studentized Residuals in Regression Analysis, in: The American Statistician, Vol. 48, No. 2, pp. 111–113.

Ha, Tran Vinh / Asada, Takumi / Arimura, Mikiharu (2019): Determination of the influence factors on household vehicle ownership patterns in Phnom Penh using statistical and machine learning methods, in: Journal of Transport Geography, Vol. 78, pp. 70–86.

El-Habil, Abdalla (2012): An Application on Multinomial Logistic Regression Model, in: Pakistan Journal of Statistics and Operation Research, Vol. 8, No. 2, pp. 271–291.

Hagenauer, Julian / Helbich, Marco (2017): A comparative study of machine learning classifiers for modeling travel mode choice, in: Expert Systems with Applications, Vol. 78, pp. 273–282.

Hägler, Max (2020): Die Zahl der Autos in Deutschland steigt weiter, URL: https://www.sueddeutsche.de/wirtschaft/auto-deutschland-anzahl-staedte-1.4940232, 17/12/2020, 02.13 p.m.

Hair, Joseph F. / Black, William C. / Babin, Barry J. / Anderson, Rolph E. (2010): Multivariate Data Analysis: Pearson New International Edition, Seventh edition, Prentice Hall 2010.

Hashemi, Ray R. / Le Blanc, Louis A. / Rucks, Conway T. / Shearry, Angela (1995): A Neural Network for Transportation Safety Modeling, in: Expert Systems with Applications, Vol. 9, No. 3, pp. 247–256.

Hauke, Jan / Kossowski, Tomasz (2011): Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data, in: Quaestiones Geographicae, Vol. 30, No. 2, pp. 87–93.

Hensher, David A. / Ton, Tu T. (2000): A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice, in: Transportation Research Part E: Logistics and Transportation Review, Vol. 36, No. 3, pp. 155–172.

Hoaglin, David C. / Iglewicz, Boris (1987): Fine-Tuning Some Resistant Rules for Outlier Labeling, in: Journal of the American Statistical Association, Vol. 82, No. 400, pp. 1147–1149.

Hu, The-Wei (1972): The fitting of log-regression equation when some observations in the regressand are zero or negative, in: Metroeconomica, Vol. 24, No. 1, pp. 86–90.

Ivan, John N. / Sethi, Vaneet (1998): Data Fusion of Fixed Detector and Probe Vehicle Data for Incident Detection, in: Computer-Aided Civil and Infrastructure Engineering, Vol. 13, No. 5, pp. 329–337.

Jakobsson, Niklas / Gnann, Till / Plötz, Patrick / Sprei, Frances / Karlsson, Sten (2016): Are multi-car households better suited for battery electric vehicles? – Driving patterns and economics in Sweden and Germany, in: Transportation Research Part C: Emerging Technologies, Vol. 65, pp. 1–15.

Jiang, Yang / Gu, Peiqin / Chen, Yulin / He, Dongquan / Mao, Qizhi (2017): Influence of land use and street characteristics on car ownership and use: Evidence from Jinan, China, in: Transportation Research Part D: Transport and Environment, Vol. 52, pp. 518–534.

Kain, John F. / Beesley, Me (1965): Forecasting car ownership and use, in: Urban Studies, Vol. 2, No. 2, pp. 163–185.

Karlaftis, Matthew G. / Vlahogianni, Eleni I. (2011): Statistical methods versus neural networks in transportation research: Differences, similarities and some insights, in: Transportation Research Part C: Emerging Technologies, Vol. 19, No. 3, pp. 387–399.

Keller, Rose / Vance, Colin (2013): Landscape pattern and car use: Linking household data with satellite imagery, Working Paper, Jacobs University Bremen 2013.

Khan, Sarosh I. / Ritchie, Stephen G. (1998): Statistical and neural classifiers to detect traffic operational problems on urban arterials, in: Transportation Research Part C Emerging Technologies, Vol. 6, No. 5-6, pp. 291–314.

Kim, Daejin / Park, Yujin / Ko, Joonho (2019): Factors underlying vehicle ownership reduction among carsharing users: A repeated cross-sectional analysis, in: Transportation Research Part D: Transport and Environment, Vol. 76, pp. 123–137.

Kim, Ji-Hyun (2009): Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap, in: Computational Statistics & Data Analysis, Vol. 53, No. 11, pp. 3735–3745.

Kitamura, Ryuichi / Bunch, David S. (1990): Heterogeneity and State Dependence in Household Car Ownership: A Panel Analysis Using Ordered-Response Probit Models with Error Component, Working Paper, University of California Transportation Center 1990.

Klaffke, Martin (2018): Generationen-Management, URL: https://wirtschaftslexikon.gabler.de/definition/generationen-management-99636/version-328745, 26/11/2020, 05.54 p.m.

Knittel, Christopher R. / Murphy, Elizabeth (2019): Generational trends in vehicle ownership and use: Are millennials any different?, Working Paper, National Bureau of Economic Research 2019.

Kuhnimhof, Tobias / Buehler, Ralph / Wirtz, Matthias / Kalinowska, Dominika (2012): Travel trends among young adults in Germany: increasing multimodality and declining car use for men, in: Journal of Transport Geography, Vol. 24, No. C, pp. 443–450.

Lafond, Daniel / Roberge-Valliéres, Benoît / Vachon, François / Tremblay, Sébastien (2017): Judgment analysis in a dynamic multitask environment: Capturing nonlinear policies using decision trees, in: Journal of Cognitive Engineering and Decision Making, Vol. 11, No. 2, pp. 122–135.

Lai, Xin / Liu, Liu (2018): A simple test procedure in standardizing the power of Hosmer–Lemeshow test in large data sets, in: Journal of Statistical Computation and Simulation, Vol. 88, No. 13, pp. 2463–2472.

Landis, J. Richard / Koch, Gary G. (1977): The Measurement of Observer Agreement for Categorical Data, in: Biometrics, Vol. 33, No. 1, pp. 159–174.

Lavieri, Patricia S. / Garikapati, Venu M. / Bhat, Chandra R. / Pendyala, Ram M. (2017): Investigation of Heterogeneity in Vehicle Ownership and Usage for the Millennial Generation, in:

Transportation Research Record: Journal of the Transportation Research Board, Vol. 2664, No. 1, pp. 91–99.

Leeper, Thomas J. (2018): Interpreting Regression Results using Average Marginal Effects with R's margins, URL:
https : / / cran . r - project . org / web / packages / margins / vignettes / TechnicalDetails . pdf, 28/11/2020, 08.41 p.m.

Letmathe, Peter / Suares, Maria (2017): A consumer-oriented total cost of ownership model for different vehicle types in Germany, in: Transportation Research Part D: Transport and Environment, Vol. 57, pp. 314–335.

Li, Jieping / Walker, Joan L. / Srinivasan, Sumeeta / Anderson, William P. (2010): Modeling Private Car Ownership in China: Investigation of Urban Form Impact Across Megacities, in: Transportation Research Record, Vol. 2193, No. 1, pp. 76–84.

Lingras, Pawan / Adamo, Mario (1996): Average and peak traffic volumes: neural nets, regression, factor approaches, in: Journal of Computing in Civil Engineering, Vol. 10, No. 4, pp. 300–306.

Long, Jeremy Scott / Freese, Jeremy (2006): Regression models for categorical dependent variables using Stata, Second edition, College Station 2006.

Ma, Jie / Ye, Xin (2019): Modeling Household Vehicle Ownership in Emerging Economies, in: Journal of the Indian Institute of Science, Vol. 94, No. 4, pp. 647–671.

Ma, Jie / Ye, Xin / Shi, Cheng (2018): Development of Multivariate Ordered Probit Model to Understand Household Vehicle Ownership Behavior in Xiaoshan District of Hangzhou, China, in: Sustainability, Vol. 10, No. 10, pp. 1–17.

Manning, Christopher D. / Raghavan, Prabhakar / Schütze, Hinrich (2008): Introduction to Information Retrieval, Cambridge 2008.

Matas, Anna / Raymond, Josep-LLuis (2008): Changes in the structure of car ownership in Spain, in: Transportation Research Part A: Policy and Practice, Vol. 42, No. 1, pp. 187–202.

McFadden, Daniel (1974): Conditional logit analysis of qualitative choice behavior, in: Frontiers in Econometrics, ed. Paul Zarembka, New York et al., 1974, pp. 105–142.

McFadden, Daniel (1977): Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some recent developments, Working Paper, Yale University 1977.

Mcfadden, John / Yang, Wen-Tai / Durrans, S. (2001): Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways, in: Transportation Research Record: Journal of the Transportation Research Board, Vol. 1751, No. 1, pp. 9–17.

Menard, Scott (2004): Six Approaches to Calculating Standardized Logistic Regression Coefficients, in: The American Statistician, Vol. 58, No. 3, pp. 218–223.

Menard, Scott (2011): Standards for Standardized Logistic Regression Coefficients, in: Social Forces, Vol. 89, No. 4, pp. 1409–1428.

Mize, Trenton D. (2019): Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects, in: Sociological Science, Vol. 6, No. 4, pp. 81–117.

Mohammadian, Abolfazl / Miller, Eric (2002): Nested Logit Models and Artificial Neural Networks for Predicting Household Automobile Choices: Comparison of Performance, in:

Transportation Research Record: Journal of the Transportation Research Board, Vol. 1807, No. 1, pp. 92–100.

Mokhtarian, Patricia / Cao, Xinyu / Handy, Susan L. (2009): Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings, in: Transport Reviews, Vol. 29, No. 3, pp. 359–395.

Nieuwenhuijsen, Mark / Khreis, Haneen (2019): Transport and health, in: Urban Health, ed. Sandro Galea / Catherine Ettman / David Vlahov, New York, 2019, pp. 52–58.

Nijkamp, Peter / Reggiani, Aura / Tritapepe, Tommaso (1996): Modelling inter-urban transport flows in Italy: A comparison between neural network analysis and logit analysis, in: Transportation Research Part C: Emerging Technologies, Vol. 4, No. 6, pp. 323–338.

Norton, Edward C. / Wang, Hua / Ai, Chunrong (2004): Computing Interaction Effects and Standard Errors in Logit and Probit Models, in: The Stata Journal, Vol. 4, No. 2, pp. 154–167.

O'Herlihy, C. St. J. (1965): Demand for Cars in Great Britain, in: Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 14, No. 2-3, pp. 162–195.

Ortúzar, Juan De Dios / Willumsen, Luis (2011): Modelling transport, Fourth edition, London et al. 2011.

Oshiro, Thais Mayumi / Perez, Pedro Santoro / Baranauskas, José Augusto (2012): How Many Trees in a Random Forest?, in: Machine Learning and Data Mining in Pattern Recognition, ed. Petra Perner, Berlin et al., 2012, pp. 154–168.

Peng, Joanne / Lee, Kuk / Ingersoll, Gary (2002): An Introduction to Logistic Regression Analysis and Reporting, in: The Journal of Educational Research, Vol. 96, No. 1, pp. 3–14.

Potoglou, Dimitris / Kanaroglou, Pavlos (2008a): Disaggregate Demand Analyses for Conventional and Alternative Fueled Automobiles: A Review, in: International Journal of Sustainable Transportation, Vol. 2, No. 4, pp. 234–259.

Potoglou, Dimitris / Kanaroglou, Pavlos (2008b): Modelling car ownership in urban areas: a case study of Hamilton, Canada, in: Journal of Transport Geography, Vol. 16, No. 1, pp. 42–54.

Reggiani, Aura / Tritapepe, Tommaso (2000): Neural networks and logit models applied to commuters' mobility in the metropolitan area of Milan, in: Neural Networks in Transport Applications, ed. Veli Himanen / Peter Nijkamp / Aura Reggiani / Juha Raitio, London, 2019, pp. 111–29.

Ritter, Nolan / Vance, Colin (2013): Do fewer people mean fewer cars? Population decline and car ownership in Germany, in: Transportation Research Part A: Policy and Practice, Vol. 50, pp. 74–85.

Roorda, Matthew / Mohammadian, Abolfazl / Miller, Eric (2000): Toronto Area Car Ownership Study: A Retrospective Interview and Its Applications, in: Transportation Research Record, Vol. 1719, No. 1, pp. 69–76.

Ryan, James / Han, Gregory (1999): Vehicle-Ownership Model Using Family Structure and Accessibility Application to Honolulu, Hawaii, in: Transportation Research Record, Vol. 1676, No. 1, pp. 1–10.

Sanko, Nobuhiro / Dissanayake, Dilum / Kurauchi, Shinya / Maesoba, Hiroaki / Yamamoto, Toshiyuki / Morikawa, Takayuki (2012): Household car and motorcycle ownership in Bangkok and Kuala Lumpur in comparison with Nagoya, in: Transportmetrica A: Transport Science, Vol. 10, No. 3, pp. 187–213.

Schiller, Preston / Bruun, Eric Christian / Kenworthy, Jeffrey (2010): An Introduction to Sustainable Transportation: Policy, Planning and Implementation, London et al. 2010.

Schintler, Laurie / Olurotimi, Oluseyi (2019): Neural Networks as Adaptive Logit Models, in: Neural Networks in Transport Applications, ed. Veli Himanen / Peter Nijkamp / Aura Reggiani / Juha Raitio, London, 2019, pp. 131–150.

Soltani, Ali (2005): Exploring the impacts of built environments on vehicle ownership, in: Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5, pp. 2151–2163.

Starkweather, Jon / Moske, Amanda Kay (2011): Multinomial Logistic Regression, URL: https://it.unt.edu/sites/default/files/mlr_jds_aug2011.pdf, 28/11/2020, 03.05 p.m.

Statistics Canada (2015): Employment patterns of families with children, URL: https://www150.statcan.gc.ca/n1/pub/75-006-x/2015001/article/14202-eng.htm, 7/1/2021, 02.21 p.m.

Statistisches Bundesamt (2019): Qualität der Arbeit - Eltern, die Teilzeit arbeiten, URL: https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-3/eltern-teilzeitarbeit.html, 7/1/2021, 02.13 p.m.

Statistisches Bundesamt (2020): Pressemitteilung Nr. N 055 vom 11. September 2020, URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/09/PD20_N055_461.html, 17/12/2020, 02.13 p.m.

Stoltzfus, Jill C. (2011): Logistic Regression: A Brief Primer, in: Academic Emergency Medicine, Vol. 18, No. 10, pp. 1099–1104.

Strobl, Carolin / Boulesteix, Anne-Laure / Zeileis, Achim / Hothorn, Torsten (2007): Bias in random forest variable importance measures: Illustrations, sources and a solution, in: BMC Bioinformatics, Vol. 8, No. 1, pp. 1–25.

Sun, Yi-lin / Susilo, Yusak / Waygood, Owen / Wang, Dian-hai (2014): Detangling the impacts of age, residential locations and household lifecycle in car usage and ownership in the Osaka metropolitan area, Japan, in: Journal of Zhejiang University Science A(Applied Physics & Engineering), Vol. 15, No. 7, pp. 517–528.

Tanner, John Curnow (1958): An analysis of increases in motor vehicles in Great Britain and the United States, Working Paper, Road Research Laboratory Harmondsworth 1958.

Tanner, John Curnow (1978): Long-term forecasting of vehicle ownership and road traffic, in: Journal of the Royal Statistical Society: Series A (General), Vol. 141, No. 1, pp. 14–41.

Van Asch, Vincent (2013): Macro-and micro-averaged evaluation measures [[BASIC DRAFT]], Working Paper, University of Antwerp 2013.

Vance, Colin / Hedel, Ralf (2008): On the Link Between Urban Form and Automobile Use: Evidence from German Survey Data, in: Land Economics, Vol. 84, No. 1, pp. 51–65.

Weinberger, Rachel / Goetzke, Frank (2010): Unpacking Preference: How Previous Experience Affects Auto Ownership in the United States, in: Urban Studies, Vol. 47, No. 10, pp. 2111–2128.

Whelan, Gerard (2007): Modelling Car Ownership in Great Britain, in: Transportation Research: Part A: Policy and Practice, Vol. 41, No. 3, pp. 205–219.

Williams, Richard (2015): Interpreting Interaction Effects; Interaction Effects and Centering, URL:
https://www3.nd.edu/~rwilliam/stats2/l53.pdf, 10/11/2020, 05.44 p.m.

Wittwer, Rico / Hubrich, Stefan (2016): What happens beneath the surface? Evidence and insights into changes in urban travel behaviour in Germany, in: Transportation Research Procedia, Vol. 14, pp. 4304–4313.

Wong, Ka Io (2013): An Analysis of Car and Motorcycle Ownership in Macao, in: International Journal of Sustainable Transportation, Vol. 7, No. 3, pp. 204–225.

Yagi, Michiyuki / Managi, Shunsuke (2016): Demographic determinants of car ownership in Japan, Working Paper, Kobe University 2016.

Yamamoto, Toshiyuki (2009): Comparative analysis of household car, motorcycle and bicycle ownership between Osaka metropolitan area, Japan and Kuala Lumpur, Malaysia, in: Transportation, Vol. 36, No. 3, pp. 351–366.

Zegras, Christopher (2010): The Built Environment and Motor Vehicle Ownership and Use: Evidence from Santiago de Chile, in: Urban Studies, Vol. 47, No. 8, pp. 1793–1817.

Zhang, Yunlong / Xie, Yuanchang (2008): Travel Mode Choice Modeling with Support Vector Machines, in: Transportation Research Record: Journal of the Transportation Research Board, Vol. 2076, No. 1, pp. 141–150.

Zhang, Zhao / Jin, Wen / Jiang, Hai / Xie, Qianyan / Shen, Wei / Han, Weijian (2017): Modeling heterogeneous vehicle ownership in China: A case study based on the Chinese national survey, in: Transport Policy, Vol. 54, pp. 11–20.

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. Further rights of reproduction and usage, however, are not granted here.

This paper was not previously presented to another examination board and has not been published.

Vanessa Angela Wanner

Adress:	Tegernseer Landstraße 129 81539 München
Phone:	0176/64986108
E-Mail:	vanessa.wanner@freenet.de
Date of Birth:	14/09/1995
Nationality:	Germany

Working Experience

Jul 2020 – today **Working Student IT Project Management**
BayWa AG, Munich

- Multiprojectcontrolling and -reporting
- Qualitymanagement inside the project management process and within the relevant software (CanDo, Jira, Confluence)

Nov 2018 – Dec 2019 **Working Student Business Intelligence and Risk Model Solutions**
MEAG MUNICH ERGO AssetManagement, Munich

- Tasks in Linux and MS Azure DevOps in the context of code organization, creation of querys and dashboard changes.
- Support in the target-actual analysis of IT projects in the enterprise-wide market risk system
- Regular performance analysis of relevant systems using SQL queries and VBA tools, as well as operative support in user management and software development within the scrum process

Jan 2018 – Jul 2018 **Intern Controlling for Hardware Development Cost**
Bayerische Motoren Werke AG, Munich

- Planning of development costs for preliminary models within the SAP-based controlling system
- Independent execution and organization of cross-functional workshops for the assessment of development costs on the basis of technical premises
- Support in developing measures within the plan-target convergence process of vehicle projects
- Independent development of a database concept to enhance estimation-transparency across the development process
- Advancement of VBA tools

Mar 2017 – Jun 2017 **Intern Restructuring and Valuation**
Baker Tilly Germany, Frankfurt am Main

- Independent monitoring and processing of the daily cashflow, as well as creation of target/actual performance comparison
- Analysis and validation of budgets and other company data to be monitored in the context of restructuring reports
- Conduction of peer group analysis and calculation of company betas, margins and multiples with the aid of Bloomberg

Mar 2016 – Apr 2016 **Intern Group Strategy and M&A**
MAN Diesel and Turbo, Augsburg

- Conduction of market analyses and visualizations of future market developments in the area of ship engines
- Analysis and creation of competitor and acquisition profiles
- Support in a strategy project concerning future development opportunities of the company

Education

Oct 2018 – now	**Technology and Management (M.Sc.)** Technical University of Munich, Munich, Germany Main Focus: Information Technology / Accounting and Finance Master thesis: “Understanding car ownership in Germany – Findings from using statistical methods and machine learning”
Oct 2014 – Apr 2018	**Global Business Management (B.Sc.)** University of Augsburg, Augsburg, Germany Main Focus: International Finance Grade Point Average: 1.31
Sep 2016 – Feb 2017	**Semester abroad in the context of the Erasmus Scholarship** Univesidad de Cádiz, Cádiz, Spain Grade Point Average: 1.03
Mar 2014	**A-Levels** Geschwister-Scholl-Gymnasium, Ludwigshafen am Rhein, Germany Grade Point Average: 1.2

International Experience and Volunteer Experience

Aug 2015 – Oct 2015	**Projectleader „Workshops in elderly homes” in the context of voluntary work in children- and elderly homes** Mensajeros de la Paz, Buenos Aires, Argentina
May 2014 – Aug 2014	**Camp Counselor at an American summer camp** YMCA Camp Lakewood, Missouri, USA
Oct 2014 – Sep 2016	**AIESEC Augsburg** Student agency for internships abroad • Teamleader in the department „Outgoing Exchange“ • Support in the finance department
Oct 2014 **–** Feb 2015	**Christian-Dierig-Haus Augsburg** Suport in the care of dementia patients

Additional Qualifications

Sep 2020 – Nov 2020	**Data Science Bootcamp with Python** TechLabs
Nov 2017 – Aug 2018	**Android Basics Nanodegree: App Development with JAVA** Udacity
IT	**Advanced Knowledge**: Python, R, Java, SQL, VBA, Financial Modeling, Bloomberg, Matlab, Microsoft Office, Git, Jira **Basic Knowledge:** Swift, HTML, Google Cloud Server, Spark
Language Proficiency	**German:** Native **English:** Business fluent **Spanish:** Intermediate
Awards	Deutschlandstipendium of the Technical University Munich Finalist in the CFA Research Challenge Germany in 2019/2020

Munich, 14. Januar 2021