

Article

An Explanatory Model Approach for the Spatial Distribution of Free-Floating Carsharing Bookings: A Case-Study of German Cities

Johannes Müller ^{1,*} , Gonçalo Homem de Almeida Correia ²  and Klaus Bogenberger ³ 

¹ Institute for Traffic, Transport and Regional Planning, University of the Federal Armed Forces Munich, Werner-Heisenberg-Weg 39, 85579 Neubiberg, Germany

² Department of Transport & Planning, TU Delft, Stevinweg 1, 2628 CN Delft, The Netherlands; g.correia@tudelft.nl

³ Institute for Traffic, Transport and Regional Planning, University of the Federal Armed Forces Munich, Werner-Heisenberg-Weg 39, 85579 Neubiberg, Germany; klaus.bogenberger@unibw.de

* Correspondence: johannes.mueller@unibw.de

Received: 17 June 2017; Accepted: 14 July 2017; Published: 24 July 2017

Abstract: When the first free-floating carsharing operators launched their business, they did not know if it would be profitable. They often started in highly populated cities without performing extensive target group analysis, and were less concerned about fleet management. Usually, there are two main datasets that can be used to find areas that would have a high demand for free-floating carsharing: booking data, for measuring the actual demand; and land use and census data for describing the activities performed in different areas in a city. In this paper, we aim to use this information to help predict the demand of free-floating carsharing systems. We use booking data provided by DriveNow for Berlin in 2014 and contextual information about the type of activity each neighborhood has. Using Berlin as a case study, we apply a negative binomial statistical model to explain the number of bookings. From the results, we conclude that free-floating carsharing is predominantly successful in areas with more affluent citizens who are open to trying new and sustainable technologies. Other important determinants that result in a high number of carsharing bookings are the area's centrality and parking lot availability. The statistical model for Berlin was then transferred to Munich and Cologne, two other cities in Germany with similar population sizes. A comparison between the estimated demand categories and actual bookings shows satisfying results, but also non-negligible local conditions influencing the spatial demand for bookings.

Keywords: carsharing demand analysis; spatial analysis; free-floating carsharing; negative binomial model; booking data

1. Introduction

Free-floating carsharing (FFCS) has become a new mode of transportation in big European and American cities. Users do not need to return their carsharing vehicle to their original location, but can start and finish their trip at any parking lot within the operating area. This flexibility has made FFCS a more attractive option over traditional round-trip models. The aim of this paper is to find significant characteristics of those city districts with a high number of FFCS bookings. The study will also help to identify a typical user of these mobility services for Germany, the case-study country.

Operators usually adopt a piecemeal approach when launching FFCS in new cities. According to interviews with the fleet management teams, the initial operating area is not strictly defined. Instead, operators focus on the city centre and gradually integrate new peripheral districts. A definition of promising districts could, therefore, help the operator decide whether or not to include new city

districts into its operating area. With this, customers can then enjoy FFCS services over a wider area. Operators can also shorten vehicle idle times by planning the system to match the existing market demand.

Although spatial analysis is important, assessing the FFCS customer profile is also important. While there is usually no specific information about users, the external census data can be used as a proxy to understand the customer. As Seign demonstrated in his doctoral thesis[1], bookings normally start in areas where users live. Therefore, census data can potentially reveal insights into the average carsharing user. This implies that the work may be considered as a study to ascertain the typical profile of an FFCS user even if external data is used to arrive at this description. Operators can potentially benefit from such customer analyses: they can roll out targeted advertising campaigns and special offers; or cooperate with businesses with similar customer profiles.

One way of understanding the customer may be through surveys, but the information gathered tends to be unsuitable for the purposes of ascertaining the average carsharing profile. While demographic information can be gathered, the only way to establish meaningful relationships between such data would be to ask respondents for their zipcodes. The frequency of FFCS trips may also be difficult to estimate in a survey, which is crucial to differentiating between customer groups. While one might be able to directly ask the respondent how often they use FFCS services and offer specific frequencies for them to select, such self-reporting may not represent actual booking behaviour. It is also difficult to get a representative sample of participants since there is no way to confirm that the sample size will have the same proportions of customer groups and their total number of trips.

To circumvent these problems, this paper focuses on observing actual carsharing trips. Courtesy of the FFCS operator DriveNow, we are able to obtain booking data in Berlin from January to December 2014. This booking data contains essential information such as the start and end locations of each trip, which allows us to illustrate FFCS demand on a spatial level. We then aggregated the booking start points over the grid of external census data that shows the population living in each neighbourhood, and their main activities. Regression models are then used to find the explanatory variables for the target variable: the number of carsharing trip starts per district.

This regression model is not only useful for finding explanatory variables in one city, but could be potentially applied to other cities where one can assume similar customer behaviours. In a sense, then, this paper's results are not only useful for fine-tuning the operational area of an FFCS service in a case study city, but can also help define potential operational areas for cities that do not currently have such service. Thus, to assess whether the model is transferable to other cities, we used booking and external data from Munich and Cologne, to validate the regression model used in Berlin.

This paper starts by reviewing the literature related to the modeling of carsharing demand. A detailed description of all datasets used follows in Section 3. The negative-binomial model is introduced afterwards in Section 4 before the results are presented and interpreted in Section 5. The paper ends by drawing several general conclusions that can be concluded from this research.

2. Literature Review

In the early 2000s, the term smart city shaped the vision for many cities around the globe [2–4]. Information and communications technology (ICT) have also influenced the mobility sector and has made shared mobility services more feasible. Carsharing and bicycle sharing are considered essential contributions to smart mobility solutions [5–8].

Carsharing services have also evolved with the rise of the mobile internet. Fixed vehicle stations have been rendered obsolete by mobile positioning systems that can provide the location of every vehicle in a city i.e., the so-called free-floating carsharing systems. Customers have found this type of carsharing more attractive because returning to the vehicle pick-up location is no longer mandatory, which made carsharing services serve a wider range of purposes other than the usual round trips.

In most work about carsharing demand modeling, the expected demand is obtained by accessing and reading the FFCS operator's API (application programming interface). The interface is commonly

used by smartphone applications and websites to provide the current distribution of available cars in the fleet. Such booking data, however, should be treated with caution. The civity study by Brockmeyer et al. [9] used this method to collect the booking data of FFCS operators in Berlin. Since they could only observe the (non-)availability of a vehicle on the map, they could not distinguish between service and customer trips. Instead of recording the amount of time a vehicle was used at around 3–4 h per day, which is demonstrated by Lenz and Bogenberger [10], they observed a time of 62 minutes. This implies that API data may be full of errors. Weigele, a co-author of the civity study, later assumed that there were some errors in the methodology, such as overestimating the assumed booking time [11]. Other studies like [12] took this data to measure the influence of points of interest (POI) on the number of bookings. The authors aggregated these datasets over a base grid consisting of squares with an edge length of 100 meters. In their chosen zero-inflated Poisson regression model, the bookings were taken as the dependent variable and the density of the several POIs as the independent variable. The zero-inflated model design excluded those cells that did not show any bookings, such as parks and other parking-prohibited areas. The significant variables with a positive influence on the number of bookings were, for example, bars, restaurants, the airport and areas where residents earn less than 500 EUR per month. A negative correlation was observed in regions with a highly educated population. Lenz and Bogenberger [10] also identified through customer surveys that the project WiMobil had well-educated men averaging 33 years old as typical users.

The first analysis of FFCS bookings was done by Kortum and Machemehl in 2012 [13]. The evaluated data of car2go in Austin showed a high acceptance and use of the system in areas with a high population and household density. A high percentage of citizens between 20 and 39 years old, as well as students or government workers, had also a positive effect on the number of bookings. The last factor could be explained by the fact that many government agencies reduced their own fleet of cars, and provided their employees discounted rates for FFCS.

Most of the literature that analyzes user groups of carsharing systems are related to station-based carsharing systems. A study from De Lorimier and El-Geneidy [14] for Montréal's station-based *communauto* tried to explain varying booking demands. The authors applied a multilevel regression analysis and showed that vehicle age, the concentration of users within a specific geographic region, and the vicinity of stations are important factors for high vehicle usage. Applying an analogous model for a station-based system in Seoul, Kang et al. [15] identified a high density of business offices and a high density of people aged between 20 and 30 to be positively correlated with carsharing demand. However, for understanding and predicting the use of FFCS, it is necessary to create a more comprehensive customer profile. A classic way to characterize typical customers and their mobility behavior is to use surveys, which can help find attributes of an average user or groups of users who are more inclined to carshare. Among other studies, Cervero's characterizations of station-based carsharing users from 2001 [16] and 2002 [17] are among the most well-known early works in carsharing demand research. In his surveys, more than 62% of the carsharing users were female, and the average yearly income of the users was about 50,000\$ which is an over-average income. The study also found that the carsharing system was mainly used during the afternoon for non-work purposes. It also noted, interestingly, that one-third of carsharing users lived alone, and every fourth shared their home with non-related adults. Cervero called them the "non-traditional" households [17].

Morency et al. also identified gender and age as characteristics having a significant impact on carsharing behaviour [18]. They also found out that user behavior in the previous four months directly influences the current frequency of usage. Kawgan-Kagan, focusing on gender, revealed that female early adopters generally show a higher affinity for bikes and a lower open-mindedness towards new technologies in comparison to male users [19].

In another study by Celsor and Millard-Ball [20], the authors emphasize the importance of the users' neighborhoods. They distilled the results from other researchers and listed four factors: parking pressure, the ability to live without a car, high population density and the mix of uses of a district.

Stillwater et al. analyzed the dependency of carsharing on public transport. Whereas a neighborhood with a light rail station had a positive impact on the demand of carsharing, regional rail availability decreased the number of bookings [21]. An overview of relevant studies from 1989–2013 on carsharing target groups was put together by Hinkeldein et al. in [22], pp. 182–186. The listed research analyzed factors like mobility-related attitudes, lifestyle, family status and leisure activities. A literature review about general approaches to model carsharing demand was published by Jorge and Correia [23].

3. Description of the Case Study Data

3.1. Booking Data

FFCS operator DriveNow provided the booking data used by this work. DriveNow started in Munich and Berlin in 2011. It now provides carsharing fleets in several European and American cities. Users register once as members, then pay a time-based fare for every trip taken (around 0.30 EUR per minute). The reservation process is designed for spontaneous trips: at the time of analysis, reserving a vehicle was free for the first 15 minutes; and then 0.10 EUR per minute thereafter.

As there are no stations or parking lots reserved for free-floating carsharing, customers use a smartphone or the operator's website to look for available vehicles. The position of the vehicle at each start and end of a trip is saved in GPS coordinates in the booking dataset.

Every trip in Berlin in the year 2014 has been included in this dataset. The dataset has also been anonymized, which makes direct links between socio-demographic data and specific trips impossible.

There are several types of FFCS trips: private, business, service in which it is either a regular trip by a private customer, a business that has contracted the service or for maintenance. This paper only takes non-service trips into account. Trips that appear to be caused by erroneous data logging were also excluded; a trip is skipped if the average speed is theoretically more than 200 km/h or if the booking begins after the end of the booking. Bookings with missing or null-values in one of the coordinate cells were also eliminated. As there were a large number of bookings, our data analyses were not affected by these deletions. The number of bookings cannot be stated in this paper due to the agreement made with the company.

3.2. Explanatory Variables

The booking data represents the observed demand for FFCS on a highly detailed spatial level. We will also try to explain the demand patterns by using data that characterises the different neighbourhoods in Berlin.

We took four groups of variables into consideration:

- Census data;
- Election behavior;
- Density of points of interest (POI);
- Centrality.

3.2.1. Census Data

The available set of census data was collected in 2012 by the geo infas institute. They provide information for German cities with different levels of precision. The grid of the present data is the so-called "district grid". The size of a district is comparable to a block in U.S. cities with a length of 400–500 m. The number of citizens is on average 500–800, but can vary greatly. The business area of the FFCS operator contains around 1863 districts in Berlin with a mean area of 0.18 sqkm. The data not only contains socio-demographic indicators, but also variables that describe how the space is used. The following are the variables that we were able to get:

- Residents data: % gender, % age (categories), % foreigners;

- Household data: % households with 1, 2, 3 or more children, purchasing power of households in average (index), % single, % yuppies (young urban professionals), % DINKS (double income no kids), rent (per sqm), automobile density, quality of buildings, social class;
- Number of companies: # services, # hotels, # wholesale markets, # clinics, # administrative offices, # retail, # manufacturers, # insurances, # mechanics;

In addition to these variables, the factors “street length” and “area size” are considered in the models. The street length is meant to be a proxy for the number of public parking lots. Only the OSM street types “primary”, “secondary”, “tertiary”, “residential” and “living street” are selected for this purpose. All streets of those types intersecting a cell were summed up and built the new variable. The area size is important since districts can differ greatly in size, so standardization may be needed.

3.2.2. Election Behavior

The data above describes the distribution of the different population groups according to their sociodemographic characteristics. It is already known that attitudes play an important role in explaining the propensity to use specific modes of transportation [24]. To measure the general social environment in a district, we look at the political attitude of the citizens, using it as a proxy to determine open-mindedness towards new mobility options. Even if political parties are not elected by a homogeneous group of people, the preference for a particular party may allow us to draw conclusions about whether voters are conservative or open-minded. By including the election behavior in the dataset of explanatory variables, we are assuming that this general attitude is also reflected in their mobility behavior and their attitudes towards new technologies.

The election of the national parliament, “Bundestag”, in October 2013 offers the best dataset to measure social milieu because it was not swayed by local issues. Barnett and Casper defined the human social environment as the “physical surrounding, social relationship and cultural milieu within which defined groups of people function and interact” [25]. The components are inter alia the government and the political attitude.

The polling districts do not correspond to the grid of census land use data. The district grid from infas is therefore taken as a basic layer. The election results that are assigned to a cell of the district grid are averaged by the results from those polling districts intersecting the cell of the district grid (see Figure 1). To make the election data transferable to other cities, neither absolute, nor percentage results are taken into account. Instead, the difference of the percentage result of the constituency and the percentage result of the polling district is evaluated.

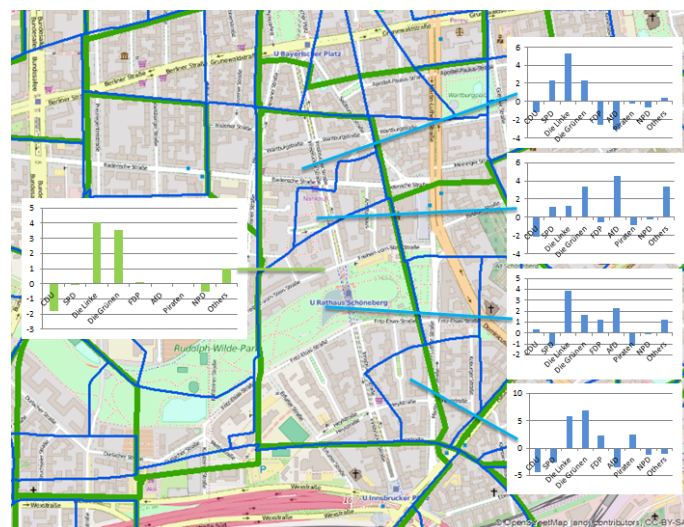


Figure 1. Elections results for the census data grid (green) averaged by those polling districts (blue) intersecting the particular cell.

3.2.3. Density of POIs

It is the purpose of the study to assess the influence of specific POI on the demand for FFCS. The analyzed groups of POI are places to go out (e.g., bars, restaurants, cinemas), places for non-locals (touristy attractions, accommodations), places for daily use (e.g., ATMs, banks) and spots for transferring to another transport mode (taxi stand, bus stop, subway/suburban train station). These POI are considered as absolute numbers.

3.2.4. Centrality

The last group of variables consists of just two measurements: the distance of each cell centroid of the district grid to the district center; and the distance from the same point to the city center. These variables are indicators for the relative position of the district in the city.

4. Count Data Modeling

One goal of this paper is to explain the booking frequency in a district. The standard instrument for studying the influence of several factors in an outcome is a standard regression model. Since the dependent variable is a count variable, it makes sense to use general linear models (GLM) instead of linear regression models as done in earlier studies with smaller datasets [26]. We performed different modeling approaches and present in this paper the negative binomial model.

The density function of the negative binomial model is given by:

$$f(y_i|\mu_i, \theta) = P(Y_i = y_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) \cdot y_i!} \cdot \frac{\mu_i^{y_i} \cdot \theta^\theta}{(\mu_i + \theta)^{y_i + \theta}} \forall i = 1, \dots, p$$

whereby in our case

Y_i the number of bookings (is the FFCS booking starts from the dataset),

i the number of the cell of the district grid,

p total number of cells in the district grid (Berlin: 1863),

θ dispersion parameter.

μ_i is the predictor defined as:

$$\mu_i = \exp(x_i' \beta) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_q x_{iq}), i \in \{1, \dots, p\}$$

with:

$x_i = x_{i1}, \dots, x_{iq}$ values of the explanatory variables in the i -th district,

β_j parameter for the j -th explanatory variable,

j the number of the explanatory variable,

q the total number of explanatory variables (here: 354).

To decide if the j -th explanatory variable contributes essentially to the explanation of the number of FFCS bookings, the respective coefficient β_j is tested for its significance. The Wald statistic is used for testing if the omission of the variable (i.e., setting the respective parameter β_j to zero) would lead to a significantly different output.

However, before checking the significance of a variable, it is necessary to see if it is already represented by other factors. The redundancy of a variable results from a high collinearity of a factor with another one. One option to quantify this redundancy is therefore to consider the variance inflation factor (VIF) defined as:

$$VIF(\beta_j) = \frac{1}{1 - R_j^2}$$

R_j^2 is the R^2 in a linear model containing x_j as the explanatory variable only. O'Brien proposed in [27] to indicate multicollinearity $VIF(\beta_j) > 5$ or $VIF(\beta_j) > 10$. The chosen threshold for this work

is 7.5. There are some approaches for designing a measurement for the fitness of the GLMs. The R^2 can only be computed in a linear regression model, so most coefficients of determination work with the likelihood function. McFadden's R^2 is chosen for the present analysis [28] and defined by:

$$R_{McFadden}^2 = 1 - \frac{\log l_1}{\log_2} \in [0, 1)$$

with l_1 being the likelihood of the model with explanatory variables and l_2 the likelihood of the null model.

5. Results and Interpretation

The application of the negative binomial model starts with the variable selection. The first selection criterion is the non-redundancy of the variables. This reduced set of variables is used as explanatory factors for each GLM model, whereby the significance of the variable is a further selection step.

5.1. Redundancy Check

Variables that are highly correlated to other ones do not provide an added value to the model. It is thus first checked if the variables are non-redundant. The VIF is at the beginning calculated for a model with all of the variables in the saturated model. The variables with the highest VIF value greater than 7.5 are iteratively omitted [29]. Some examples of the redundant variables are:

- population: age distribution: several age groups (∞)
- number of companies: several specifications (∞)
- economy: purchasing power for consumption, for daily goods (∞)
- vehicles: registered vehicles, automobiles, motorcycles, private vehicles (∞)
- households: double income no kids (384.3491), households, yuppies (62.91652), population, affinity for leasing (index) (8.571255)
- election results: CDU (Conservatives), second vote (169.109), Die Grünen (Greens), second vote (22.05291), first vote (10.75966), Die Linke (far-left), second vote (15.70378), NPD (far-right), second vote (13.6863), FDP (Liberals), second vote (8.09295)

5.2. Significance Check

The 108 non-redundant variables are the first basic dataset for the regression model. We distinguish at first two NB models: Model I includes all non-redundant variables with 95% significance; Model II consists just of the very highly significant, namely 99.9% significant variables (Table 1).

The McFadden's R^2 values of all models are not very different, which means that the highly significant variables have the most explanatory power. As explained in [30], a R^2 value between 0.2 and 0.4 is usually a good result as this is not measured in the same scale as that of a linear regression. The models nevertheless have a quite disappointing value and would mean an explanation rate of around 30% of the data. We therefore want to discuss a third regression model that skips the redundancy check and contains all variables from the four datasets. Model III consists of all 95% significant, but partly redundant variables listed in Table 2. The fit of the model is not significantly better however.

There are some reasons that might cause the quite low pseudo R^2 values. One usual explanation is a poor choice of statistical model. This is not likely since the residual plot in Figure 2 shows a normal distribution of the residuals with just some outliers. The residual plots of the Poisson and quasi-Poisson model approaches were worse when these models were experimented with. The NB model could moreover factor in the indicated overdispersion of the data.

Table 1. Non-redundant variables with at least 95% (99.9% (bold)) significance and the according model results. ‘*’, ‘**’ and ‘***’ indicate 95%, 99% and 99.9% significance, respectively.

Variable Name	Estimate	Std. Error	Significance
Intercept	8.92	2.70	***
Census			
population, percentage, foreigners	-4.212×10^{-3}	1.574×10^{-3}	**
household size, average	4.981×10^{-1}	1.609×10^{-1}	**
population, density (per sqkm)	-2.977×10^{-5}	2.117×10^{-6}	***
population, percentage, 03–05 years	-2.294×10^{-1}	6.804×10^{-2}	***
population, percentage, 10–14 years	1.205×10^{-1}	4.465×10^{-2}	**
# companies: retail, medium	3.254×10^{-2}	1.453×10^{-2}	*
# companies: hotels, big	-3.681×10^{-1}	1.756×10^{-1}	*
# companies: density (per sqkm)	5.676×10^{-5}	2.818×10^{-5}	*
# companies: insurances, big	3.126×10^{-1}	1.367×10^{-1}	*
buildings, percentage, poor quality	-4.604×10^{-3}	1.007×10^{-3}	***
buildings, percentage, very high quality	-2.879×10^{-3}	8.699×10^{-4}	***
# buildings	-2.063×10^{-3}	7.436×10^{-4}	**
households, percentage, >=3 persons	-2.956×10^{-2}	7.228×10^{-3}	***
# households, net income: 900–1500 EUR	5.914×10^{-3}	6.555×10^{-4}	***
# households, characteristic: down-to-earth	-7.626×10^{-4}	2.503×10^{-4}	**
households, social class: low	-2.869×10^{-3}	9.675×10^{-4}	**
registered vehicles, index	-4.778×10^{-3}	8.988×10^{-4}	***
population, profit-oriented, index	-6.026×10^{-3}	2.672×10^{-3}	*
population, telecommunication type: analogous	-1.007×10^{-2}	4.890×10^{-3}	*
rent	1.377×10^{-1}	2.158×10^{-2}	***
automobiles, percentage, private	-3.374×10^{-3}	1.720×10^{-3}	*
street length	2.400×10^{-4}	1.943×10^{-5}	***
district size	2.979×10^{-7}	1.266×10^{-7}	*
Election			
SPD (socialists), 2nd vote	-2.807×10^{-2}	9.913×10^{-3}	**
Piratenpartei (Pirates), 1st vote	-5.645×10^{-2}	2.219×10^{-2}	*
Piratenpartei (Pirates), 2nd vote	5.311×10^{-2}	1.943×10^{-2}	**
NPD (far-right), 1st vote	-2.043×10^{-1}	3.856×10^{-2}	***
POI			
Bars	2.895×10^{-2}	9.387×10^{-3}	**
Centrality			
Distance to the city center	-1.036×10^{-4}	1.563×10^{-5}	***
Model I (95 % significant, non-redundant variables)			
Null deviance: 6057.5 on 1862 degrees of freedom			
Residual deviance: 1979.5 on 1833 degrees of freedom			
AIC: 27,328			
Theta: 3.0955			
Std. Err.: 0.0980			
McFadden: 0.07701861			
Model II (99.9 % significant, non-redundant variables)			
Null deviance: 5350.8 on 1862 degrees of freedom			
Residual deviance: 1988.4 on 1851 degrees of freedom			
AIC: 27,545			
Theta: 2.7314			
Std. Err.: 0.0858			
McFadden 0.06846484			

Table 2. Variables (incl. redundant) with at least 95% significance and the according model results. ‘*’, ‘**’ and ‘***’ indicate 95%, 99% and 99.9% significance, respectively.

Variable Name	Estimate	Std. Error	Significance
Intercept	2.439	1.809	
Census			
household size, average	−1.057	3.952×10^{-1}	**
population, density (per sqkm)	-3.052×10^{-5}	2.168×10^{-6}	***
population, percentage, 00–14 years	2.063×10^{-1}	9.289×10^{-2}	*
population, percentage, 06–09 years	-4.411×10^{-1}	1.447×10^{-1}	**
population, percentage, 25–29 years	-1.517×10^{-1}	7.317×10^{-2}	*
population, percentage, 25–49 years	1.694×10^{-1}	6.931×10^{-2}	*
population, percentage, 35–39 years	-3.386×10^{-1}	8.212×10^{-2}	***
population, percentage, 40–44 years	-3.635×10^{-1}	1.107×10^{-1}	**
# companies: government agencies, big	-3.566×10^{-1}	1.555×10^{-1}	*
# companies, density (per sqkm)	5.850×10^{-5}	2.951×10^{-5}	*
buildings, absolute	-1.699×10^{-3}	8.526×10^{-4}	*
households, percentage, 1 person	3.800×10^{-2}	1.209×10^{-2}	**
households, percentage, 2 persons	4.744×10^{-2}	1.832×10^{-2}	**
# households, characteristic: active middle class	4.266×10^{-3}	1.999×10^{-3}	*
vehicles, density, index	-3.246×10^{-3}	9.370×10^{-4}	***
liquid purchasing power, per person	-4.581×10^4	1.979×10^4	*
decisive criteria for purchasing: brand, as index	6.431×10^{-3}	2.677×10^{-3}	*
population, affinity to environment, index	5.675×10^{-3}	2.009×10^{-3}	**
automobiles, density per inhabitant	-2.075×10^{-1}	8.051×10^{-2}	**
automobiles, percentage, commercial	3.794×10^{-3}	1.717×10^{-3}	*
Distance: airport	1.796×10^{-5}	6.665×10^{-6}	**
Distance: long-distance train station	-4.689×10^{-5}	1.649×10^{-5}	**
Distance: motorway	-2.939×10^{-5}	1.377×10^{-5}	*
street length	2.767×10^{-4}	2.075×10^{-5}	***
district area	3.602×10^{-7}	1.351×10^{-7}	**
Election			
CDU (Conservatives), 2nd vote	6.598×10^{-2}	2.116×10^{-2}	**
SPD (Socialist), 1st vote	2.708×10^{-2}	9.623×10^{-3}	**
FDP (Liberals), 2nd vote	1.568×10^{-1}	2.808×10^{-2}	***
Die Gruenen (Greens), 2nd vote	8.129×10^{-2}	1.791×10^{-2}	***
Piratenpartei (Pirates), 1st vote	7.201×10^{-2}	2.072×10^{-2}	***
Piratenpartei (Pirates), 2nd vote	7.835×10^{-2}	2.756×10^{-2}	**
AfD (right-populists), 1st vote	1.142×10^{-1}	3.915×10^{-2}	**
NPD (far-right), 1st vote	-1.594×10^{-1}	8.038×10^{-2}	*
POI			
Bars	2.241×10^{-2}	9.731×10^{-3}	*
Centrality			
Distance to the city center	-9.840×10^{-5}	1.923×10^{-5}	***
Model III (95% significant variables)			
Null deviance: 5664.6 on 1862 degrees of freedom			
Residual deviance: 1984.1 on 1827 degrees of freedom			
AIC: 27,477			
Theta: 2.8930			
Std. Err.: 0.0912			
Mcfadden 0.07240754			

Another simple reason for the pseudo values could be that there are important variables missing in the model or some variables like the street length do not represent the supposed influence of parking lots in a sufficient way. We assume that some effects appear locally and could not be represented through the very general set of variables. For instances, the number of bars might be a plausible influence variable on the number of FFCS bookings. However, if there are several bars within

a pedestrian precinct at one place and a similar number of bars with available infrastructure for cars, the booking demand is supposed to be different.

The results of our modeling approach show that precise forecasts with our chosen datasets are not possible. Nevertheless, we are able to find trends and general positive and negative tendencies of spatial characteristics that can be demonstrated by the significance and signs of the coefficients.

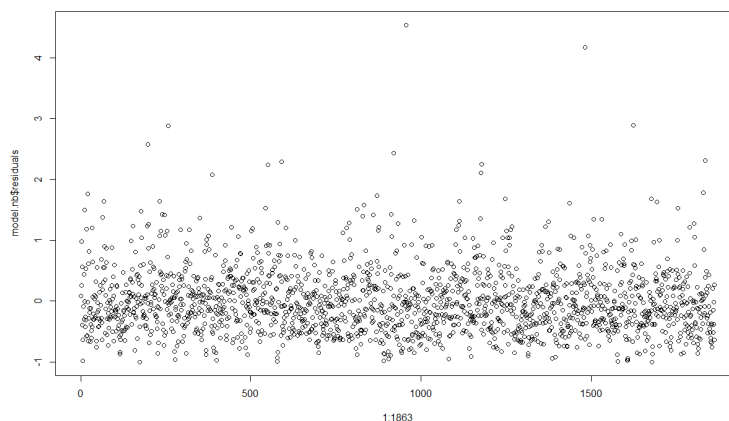


Figure 2. Residual plot of the NB regression model.

5.3. Interpretation of the Variables' Effect

The interpretation is focused on the models with only non-redundant, significant variables (Model I). The variable selection process helped to focus on the factors that can explain the demand in the best way.

The different scales of the variables do not allow for a comparison between them; hence, the interpretation is focused on the sign of the estimate in Table 2. We assume that because of the redundancy selection process done at the beginning of the analysis, these variables are also representing other similar ones from the original bigger set. The interpretation aims therefore at finding categories of significant variables.

There are some explanatory variables that are obviously related to mobility behavior. The average number of private cars and the index for registered vehicles describe the affinity of citizens in the district towards private car ownership, thus representing what we call the type of car user. The greater the percentage of people who own a vehicle, the less is the frequency of FFCS bookings.

Considered on a city level, the private car density also indicates, like rent, the centrality of an area. Berlin's citizens tend not to own a car in central districts. Rents are also higher in these areas, and the sign of this variable is, therefore, positive. The distance to the city center having a coefficient with a negative sign can be assigned to the centrality category, as well. A high density of bars and companies in general is positive for the FFCS demand. The absolute number of buildings, however, has a negative influence, which may be caused by the fact that in dense areas, the absolute number is lower, but the number of units per building is significantly higher than in the periphery.

The rent in a district represents, as well, a certain measure of the attractiveness of a place, but also how much money the residents in this area can afford to pay for living in it. Thus, the variable is also a representative of the financial situation of the users. FFCS is a means of transport that is not affordable for every social class. A 10-min trip is as expensive as an inner city ride by public transport. Customers of flexible carsharing should not be too price sensitive since they value convenience. A high number of households of people from a low social class is therefore negatively influencing the demand. A too profit-oriented population is the other extreme and reduces the number of booking starts, as well.

The street length is not difficult to interpret either. The variable was inserted into the model as a proxy for parking opportunities. The positive sign shows that public parking is probably essential for a high number of bookings and more relevant than the size of a district.

A political party that turns out to be non-redundant and very significant is the far-right party NPD. Voters for this party are assumed to come from a very conservative milieu who tend to refuse the usage of new modes of transport. The negative sign thus indicates the low open-mindedness of citizens in these districts, which are recognizable in their negative attitude towards FFCS carsharing. The percentage of foreigners and the affinity towards analogous telephones can also be interpreted as traditional households, which do not positively influence the FFCS demand.

The age variable (03–05 years old population) and the household size form a category that can best be characterized by the expression family situation. The factors represent the percentage of young families in a district. Because of the fact that the birth of a child is still a reason most parents buy a car (thereby altering mobility behavior completely), the variables have a natural negative impact on the number of carsharing bookings. This may play a role: baby seats are not part of the standard equipment, and only backless booster seats are available on some rental vehicles. Despite all efforts of the FFCS provider, some customers take the equipment away from the car's trunk. The variable 10–14 years describes families in a different situation. They are usually financially better situated and may use FFCS as a substitute for one's own second car.

The rest of the variables are not always easy to categorize. Surprisingly, the residents' density has a negative impact on the booking frequency. Intuitively, more people in an area would mean more potential customers. It is likely that the density has to be considered as a compensation for other variables in the model. Another reason is that a high density of citizens is positive for the demand of FFCS vehicles, but can only be satisfied if sufficient parking space is available. This is often not the reality in central districts. Especially in districts with many old buildings, it is common that many possibilities of curb parking do not exist.

An interesting fact is that at least one variable of these categories also appears in Model II. This indicates that a limitation of the model to the highly significant variables makes sense.

Some variables (such as the votes for the Greens) surprisingly do not appear in Model I or II because they were redundant. To ensure that we do not neglect any important group of variables in our interpretation, we consider also the at least 95% significant variables of the model containing all variables (Model III, Table 2).

It is nearly only census data variables that prove to be redundant. The reason for this is that many variables appear in many specifications. For instance, the age groups for male and female are clearly correlated with the age groups in total, and the indices are related to the corresponding variable in EUR. Quite surprising is the fact that the voting results of the Bundestag election are for some parties already expressed by other variables and therefore turn out to be redundant.

A look at the variables of Model III shows that the parties are mostly significant, and the Greens have as expected a positive estimate.

There are also more age variables appearing in the list. Again, residents with very young children (up to 10 years) or between 35 and 44 years have a negative influence, while households with one or two persons have a positive impact. A higher density of registered vehicles appears again to be non-favorable for FFCS. Centrality proves also in this model to be an important influencing factor.

We can thus conclude that there are six important variable categories found to have a statistically-significant influence on the spatial distribution of the demand for FFCS in all three models. These were

- Open-mindedness;
- Type of car user;
- Financial situation;
- Centrality;
- Parking availability;

- Family situation.

Some of the variables have already appeared to be significant in a linear regression model applied to other smaller datasets [26]. The results of these models are however more reliable due to a better model fit.

An important question remains as if these categories, which primarily characterize each district, can also be used to describe the typical customers of FFCS. As has been said already, the authors see the study from Seign as a reason for transferring socio-demographic characteristics of the district to the users. This conclusion is confirmed by the findings of Mueller et al. [31] who present the results of surveys from onboard units of the vehicles in which users were asked to tell the purpose of their trip: Most of the costumers use carsharing for their trip back home.

5.4. Transfer of the Berlin Model to Munich and Cologne

In the following, the estimated NB model is applied to the cities of Munich and Cologne to assess the usability and transferability for other cities in order to predict potential hot spots for FFCS. It is also applied for Berlin itself to show the performance of the model in predicting its own estimation data. As mentioned above, all three considered models do not have enough explanatory power to be used as precise forecast models for the absolute number of bookings, which would be required to solve some operational problems of the carsharing operator [32–34]. It is also not possible to estimate the exact number of bookings in another city since the fleet size and the number of customers influences the absolute booking frequency. Rather, the model could be useful for predicting booking hot spots. Categorizing the prediction of trips and observed bookings is hence a necessary step. The validation of the negative binomial model is done by applying it to other cities and comparing the results with observed booking data by distinguishing five categories between low demand and high demand. The result is presented in two ways: Figures 3–5 show maps with the observed booking demand. This is simply calculated by aggregating the position of trip starts over the district grid distinguished into five categories. The maps below show those results for Model II on the right, and for a better comparison between observed and predicted categories, a difference plot is mapped on the left.

The other results are quantitative. Table 3 shows the rate of correctly predicted districts (zero), as well as the rate of the underestimated (negative values) and overestimated ones (positive values) for each city.

Figure 3 and Table 3 show the results of the model application for Berlin, which are, as expected, very good. Underestimated districts in the difference plot are colored green, overestimated regions in red. A correct prediction of the frequency category leads to yellow-colored cells. The model works more than satisfactory. More than 45% is predicted correctly, and over 85% has only a deviation of ± 1 .

Table 3. Quantitative evaluation of the differences between predicted and observed categories for the cities of Berlin, Munich and Cologne. The values are the percentage of cells with the respective difference between the predicted and observed category.

	−4	−3	−2	−1	0	1	2	3	4
Berlin	0.00	0.82	5.83	21.41	45.48	18.74	6.54	1.09	0.11
Munich	2.11	4.02	8.85	17.30	31.29	21.23	10.97	3.72	0.50
Cologne	0.16	1.63	7.48	20.00	37.24	20.81	9.43	2.60	0.65

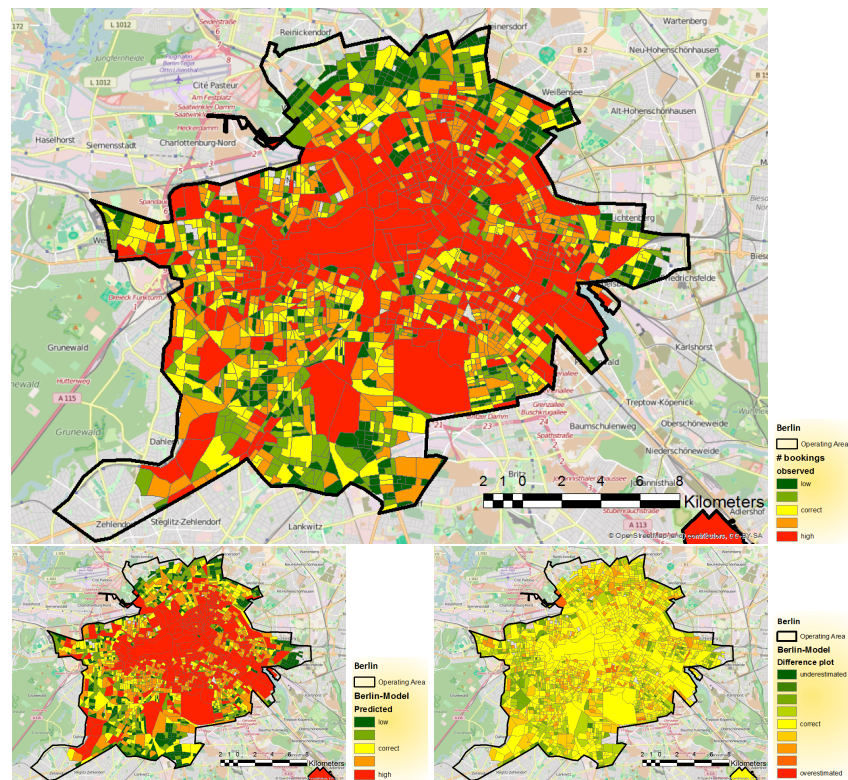


Figure 3. Maps of Berlin showing the observed (top) and predicted (bottom, left) categories for the number of bookings, as well as the difference plot (bottom, right).

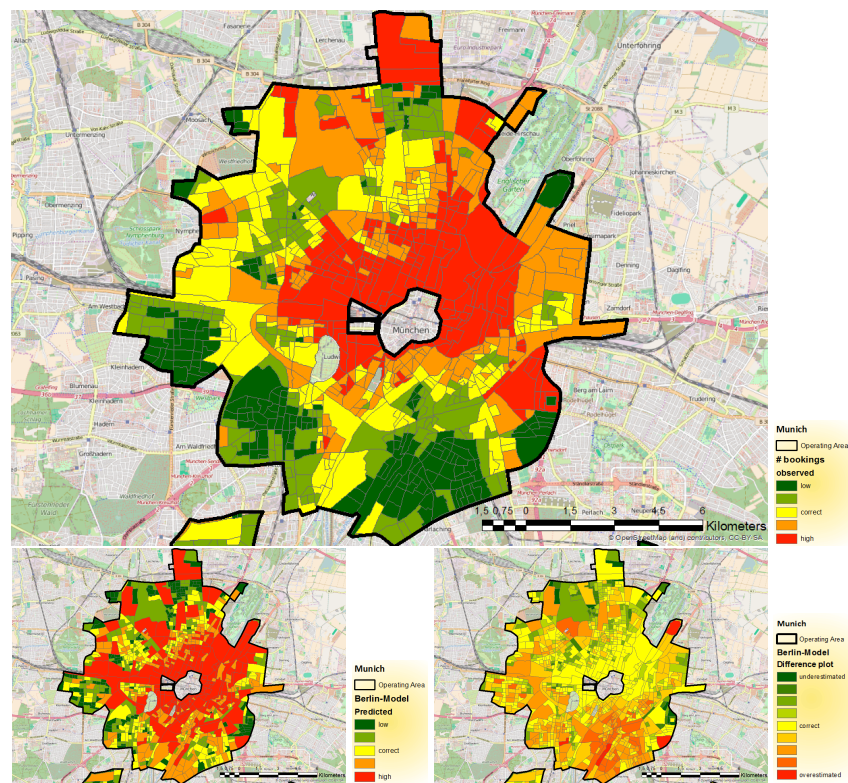


Figure 4. Maps of Munich showing the observed (top) and predicted (bottom, left) categories for the number of bookings, as well as the difference plot (bottom, right).

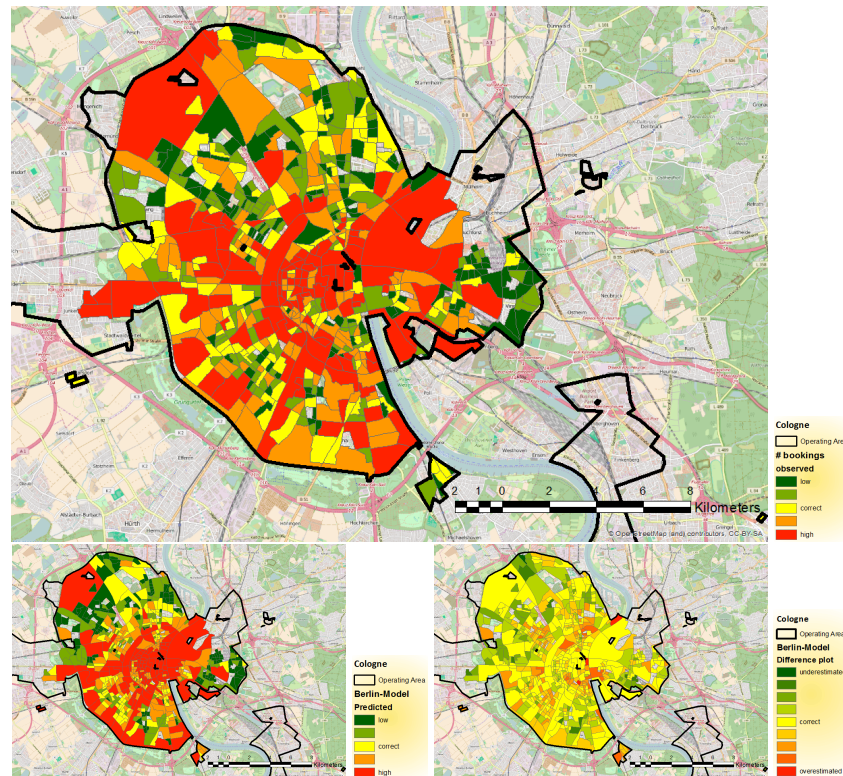


Figure 5. Maps of Cologne showing the observed (top) and predicted (bottom, left) categories for the number of bookings as well as the difference plot (bottom, right).

The observed data for Munich (Figure 4, Table 3) also show a strong centrality. Some northern parts also have a high demand, whereas southern regions show fewer bookings. The southern districts are overestimated, whereas the area around the BMW headquarters in the north is slightly underestimated. This is a good example of an additional local effect that is unpredictable by transposing the model of another city. Nearly 70% of the cells are classified with a good accuracy, and around 30% of the cells are categorized in the right category.

The city of Cologne also obtains a prediction by Model II (Figure 5, Table 3). The model just fails in some northern parts of Cologne: 37% of all districts are predicted correctly; 78% have just a slight deviation.

The validation of the models by applying them to other cities shows in general satisfying results. Even if the number of variables is reduced to the very significant ones, the NB model can be used as an excellent instrument to explain and predict hot spots of FFCS demand. The success can easily be observed by looking at each difference plot.

Nevertheless, there are local effects that affect the demand for bookings and are not represented in the model. These are for instance an over-average popularity of carsharing at a company. The BMW headquarters is an example, but also other companies may have special agreements with the carsharing operator. Peripheral areas appear to be more likely unpredictable, as well. The POIs outside of the operating area are a possible influence for this effect. Furthermore, some inner city areas vary slightly from the model prediction, which could be caused by local parking restrictions.

6. Conclusions

The purpose of this paper was to use booking data from the FFCS operator DriveNow to model and explain the spatial demand for carsharing cars by means of a negative binomial model. The chosen data are useful in explaining most of the hot spots in a city as is visible from the transfer of the Berlin model to Munich and Cologne. However, a prediction model that is just based on land use data, the

political election behavior, POIs and information about centrality is not sufficient for a precise forecast of the absolute number of bookings. The pseudo R^2 value of around 0.07 in all three considered negative binomial models can be interpreted as an explanation of booking data by around 30%.

The different models all contain similar categories. These are describing either the residents or the spatial environment of the area. A moderate or good financial situation of the residents has a positive effect on the carsharing demand since the convenience of this new means of transport is more expensive than public transport. It is also positive if the kind of car user is neither the traditional private car owner nor a denier of motorized individual transport, but someone who has an affinity to rental cars. Since FFCS is still a new technology, a high percentage of open-minded residents in an area measured by an index or by non-voters for far-right parties has a positive effect. FFCS is at the moment not attractive enough for families with small children. Singles or couples with no kids or children above 10 years have a generally more positive attitude towards carsharing.

Two spatial effects become apparent in our study. Central areas are higher demanded than peripheral districts, and the availability of parking lots is crucial since the vehicles need curb parking spaces.

The model for Berlin was transferred to Cologne and Munich. Booking hot spots were successfully predicted by the model, but there are additional local effects in each city that make the demand locally biased. These are for instance special agreements between the carsharing operators and companies, specific parking restrictions especially in the inner city or effects from outside of the operating area that are not considered in the model.

The scale of the parameters of the model could be improved. Since the dependent variables are categorical, metric or interval scaled, it is helpful to equalize the variables on one level and make the effect of a factor better comparable with others. We propose moreover to leave the idea of non-flexible classic modeling and to try machine learning algorithms. Supervised machine learning can be applied by using the number of bookings as labels and the dependent variables as input. The output is expected not to be different in its interpretation, but the results could provide a better understanding of the influencing effects.

Acknowledgments: The authors of this work would like to thank the company DriveNow for sharing the data of their daily bookings. We would also like to thank the following municipalities for the free use of their geographic data: Amt für Statistik Berlin-Brandenburg, Landeshauptstadt München, Statistikamt Nord, Stadt Köln. This work would not have been realized without the funding by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety for the project WiMobil.

Author Contributions: Johannes Müller performed all analyses in this manuscript and produced all figures and tables. Gonçalo Homem de Almeida Correia supported the literature review and proofread carefully the text. Klaus Bogenberger was responsible for the data acquisition and gave substantial thematic guidance and advice.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AfD	Alternative für Deutschland
AIC	Akaike's Information Criterion
API	Application Programming Interface
CDU	Christlich Demokratische Union Deutschlands
FDP	Freie Demokratische Partei
FFCS	free-floating carsharing
ICT	information and communication technology
NB	negative-binomial
NPD	Nationaldemokratische Partei Deutschlands
SPD	Sozialdemokratische Partei Deutschlands
VIF	variance inflation factor
POI	point of interest

References

1. Seign, R. Model-Based Design of Free-Floating Carsharing Systems. Doctoral Thesis, University of the Federal Armed Forces Munich, Neubiberg, Germany, 2015.
2. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518.
3. Neirotti, P.; De Marco, A.; Cagliano, A.C.; Mangano, G.; Scorrano, F. Current trends in Smart City initiatives: Some stylised facts. *Cities* **2014**, *38*, 25–36.
4. Chourabi, H.; Nam, T.; Walker, S.; Gil-Garcia, J.R.; Mellouli, S.; Nahon, K.; Pardo, T.A.; Scholl, H.J. Understanding smart cities: An integrative framework. In Proceedings of the IEEE 2012 45th Hawaii International Conference on System Science (HICSS 2012), Maui, HI, USA, 4–7 January 2012; pp. 2289–2297.
5. Midgley, P. The role of smart bike-sharing systems in urban mobility. *Journeys* **2009**, *2*, 23–31.
6. Benevolo, C.; Dameri, R.P.; D’Auria, B. Smart mobility in smart city. In *Empowering Organizations*; Springer: Berlin, Germany, 2016; pp. 13–28.
7. Pinna, F.; Masala, F.; Garau, C. Urban Policies and Mobility Trends in Italian Smart Cities. *Sustainability* **2017**, *9*, 494.
8. Garau, C.; Masala, F.; Pinna, F. Cagliari and smart urban mobility: Analysis and comparison. *Cities* **2016**, *56*, 35–46.
9. Brockmeyer, F.; Frohwerk, S.; Weigele, S. Free-Floating-Carsharing: Urbane Mobilitaet im Umbruch-Herausforderungen und Chancen fuer den oeffentlichen Verkehr/Free-Floating-Carsharing: Change of urban mobility. *NAHVERKEHR* **2014**, *32*, 13–18. (In German)
10. Lenz, B.; Bogenberger, K. WiMobil—Wirkung von E-CarSharing-Systemen auf Mobilität und Umwelt in Urbanen Räumen. Halbzeitkonferenz zur Nutzung von E-CarSharing-Systemen am Beispiel von car2go, DriveNow und Flinkster, 2014. Available online: http://www.erneuerbar-mobil.de/sites/default/files/2016-10/Abschlussbericht_WiMobil.pdf (accessed on 17 June 2017).
11. Rat für nachhaltige Entwicklung. *Nutzt Carsharing? Warum zwei Studien zu Unterschiedlichen Ergebnissen Kommen*; Rat für nachhaltige Entwicklung: Berlin, Germany, 2015. (In German)
12. Wagner, S.; Brandt, T.; Neumann, D. Data analytics in free-floating carsharing: Evidence from the city of Berlin. In Proceedings of the IEEE 2015 48th Hawaii International Conference on System Sciences (HICSS 2015), Grand Hyatt, HI, USA, 5–8 January 2015; pp. 897–907.
13. Kortum, K.; Machemehl, R.B. *Free-Floating Carsharing Systems: Innovations in Membership Prediction, Mode Share, and Vehicle Allocation Optimization Methodologies*; Technical report; Southwest Region University Transportation Center, Center for Transportation Research, University of Texas at Austin: Austin, TX, USA, 2012.
14. De Lorimier, A.; El-Geneidy, A.M. Understanding the factors affecting vehicle usage and availability in carsharing networks: A case study of Communauto carsharing system from Montréal, Canada. *Int. J. Sustain. Transp.* **2013**, *7*, 35–51.
15. Kang, J.; Hwang, K.; Park, S. Finding Factors that Influence Carsharing Usage: Case Study in Seoul. *Sustainability* **2016**, *8*, 709.
16. Cervero, R. City CarShare: First-year travel demand impacts. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1839*, 159–166.
17. Cervero, R.; Tsai, Y. San Francisco City CarShare: Second-Year Travel Demand and Car Ownership Impacts. In Proceedings of the Transportation Research Board 2004 Annual Meeting, Washington, DC, USA, 12–16 January 2003.
18. Morency, C.; Habib, K.M.N.; Grasset, V.; Islam, M.T. Understanding members’ carsharing (activity) persistency by using econometric model. *J. Adv. Transp.* **2012**, *46*, 26–38.
19. Kawgan-Kagan, I. Early adopters of carsharing with and without BEVs with respect to gender preferences. *Eur. Transp. Res. Rev.* **2015**, *7*, 33.
20. Celsor, C.; Millard-Ball, A. Where does carsharing work?: Using geographic information systems to assess market potential. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *1992*, 61–69.
21. Stillwater, T.; Mokhtarian, P.L.; Shaheen, S. *Carsharing and the Built Environment: A GIS-Based Study of One US Operator*; UC Davis: Davis, CA, USA, 2008.

22. Hinkeldein, D.; Schoenduwe, R.; Graff, A.; Hoffmann, C. Who Would Use Integrated Sustainable Mobility Services—And Why? In *Sustainable Urban Transport*; Emerald Group Publishing Limited: Bingley, UK, 2015; pp. 177–203.
23. Jorge, D.; Correia, G. Carsharing systems demand estimation and defined operations: A literature review. *Eur. J. Transp. Infrastruct. Res.* **2013**, *13*, 201–220.
24. De Almeida Correia, G.H.; de Abreu e Silva, J.; Viegas, J.M. Using latent attitudinal variables estimated through a structural equations model for understanding carpooling propensity. *Transp. Plan. Technol.* **2013**, *36*, 499–519.
25. Barnett, E.; Halverson, J. Local increases in coronary heart disease mortality among blacks and whites in the United States, 1985–1995. *Am. J. Public Health* **2001**, *91*, 1499–1506.
26. Schmöller, S.; Weikl, S.; Müller, J.; Bogenberger, K. Empirical analysis of free-floating carsharing usage: The Munich and Berlin case. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 34–51.
27. O'Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690.
28. McFadden, D. *Conditional Logit Analysis of Qualitative Choice Behavior in Frontiers in Econometrics*; Academic Press: Cambridge, MA, USA, 1974; pp. 105–142.
29. Beck, M. Collinearity and Stepwise VIF Selection. Available online: <https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/> (accessed on 15 April 2016).
30. Lee, D. A comparison of choice-based landscape preference models between british and korean visitors to national parks. *Life Sci. J.* **2013**, *10*, 2028–2036.
31. Mueller, J.; Schmoeller, S.; Giesel, F. Identifying Users and Use of (Electric-) Free-Floating Carsharing in Berlin and Munich. In Proceedings of the 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC 2015), Canary Islands, Spain, 15–18 September 2015; pp. 2568–2573.
32. Jorge, D.; Correia, G.H.; Barnhart, C. Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1667–1675.
33. Jorge, D.; Barnhart, C.; de Almeida Correia, G.H. Assessing the viability of enabling a round-trip carsharing system to accept one-way trips: Application to Logan Airport in Boston. *Transp. Res. Part C Emerg. Technol.* **2015**, *56*, 359–372.
34. De Almeida Correia, G.H.; Antunes, A.P. Optimization approach to depot location and trip selection in one-way carsharing systems. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 233–247.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).