



Article

How Personal Accessibility and Frequency of Travel Affect Ownership Decisions on Mobility Resources

Vaclav Plevka ^{1,*}, Paola Astegiano ¹, Willem Himpe ¹ , Chris Tampère ¹
and Martina Vandebroek ^{2,3} 

¹ L-Mob Leuven Mobility Research Centre, KU Leuven, B-3000 Leuven, Belgium; paola.astegiano@kuleuven.be (P.A.); willem.himpe@kuleuven.be (W.H.); chris.tampere@kuleuven.be (C.T.)

² Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium; martina.vandebroek@kuleuven.be

³ Leuven Statistics Research Centre, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

* Correspondence: vaclav.plevka@kuleuven.be

Received: 15 December 2017; Accepted: 13 March 2018; Published: 21 March 2018



Abstract: This paper presents a mobility-resource ownership model. The model captures inter-related personal mobility decisions: which transport mode (out of those available to a decision-maker) to use for a particular trip and which mobility resources (e.g., car, bicycle, transit season ticket or a combination) should the decision-maker own to enable the most “appropriate” set of transport modes. Importantly, the mobility decisions are not evaluated only for a single trip or a single day. In fact, for each decision-maker, an entire set of trips, observed over multiple days, is evaluated. We call this personal accessibility to travel. We present a two-step discrete choice model that includes both mode choice and ownership decisions. The model is estimated based on household travel survey data from Germany. This paper also investigates the simulation of travel times for non-chosen modes that are required as an input. The estimation results show significant effects of the personal accessibility and travel frequency on mobility-resource ownership decisions. To further validate the estimation, the forecasting and sensitivity analysis of the model for different scenarios is evaluated. The proposed model offers an efficient solution to situations when the impact of transport sustainability measures on mobility behaviour needs to be plausibly predicted.

Keywords: mobility; ownership decisions; non-chosen alternatives

1. Introduction

Transport supply shapes urban areas and influences the quality of life of its inhabitants. Together with other factors, such as socio-economic background or built environment characteristics, the provision of transport supply determines which transport modes people will use. Hence, policymakers have been leveraging the transport supply to achieve more sustainable urban forms. It is not always obvious exactly how the transport supply impacts modal split. For example, to what extent might a city that suffers from car-traffic congestion mitigate the problem by installing a new parking scheme? This could make it more expensive for drivers to park their cars in the city. Or should the city enhance its public transportation supply? Answering these questions requires serious research effort. One of the challenges is that the mode choice model needs to account not only for the impact of the new parking scheme or improved public transit quality, but also for the mobility resource constraints. Such constraints are associated with everyday choices of transport modes, which are subject to the mobility resources available. For example, people who own a car might still use it irrespective of increased car-related costs. They might consider, for example, a public transit season ticket as a disadvantageous investment because of the expenses they already have (car amortization,

insurance, service etc.). This example illustrates how mobility decisions are interrelated: which mobility resource I will acquire (taking into account the travel activities I typically do), and which mode I will use (from those available to me given my resources). This can change for different travel activities.

This study elaborates on a model that predicts mobility-resource ownership. The model accommodates interactions between long- and short-term mobility decisions, which represent mobility-resource and transport mode decisions, respectively. In contrast to many existing mobility-resource ownership models, which are based on socio-demographic and built-environment characteristics only, this model is sensitive to supply provision and personal mobility needs. The model is useful to predict interrelated long- and short-term demand changes induced by modifications to existing supply. It has the potential to also be adapted to predicting first estimates of potential market shares of newly introduced mobility options. An example of this is a new shared taxi service that will be installed, for which no socio-demographic preferences can be known yet. In this example, the monetary and time cost utility components can be compared amongst already existing alternatives and then rationally extrapolated to modal share estimates using the model.

In order to test the model on real-world data, the study explores methods for simulating the (dis-)utility (in this case: travel times) of alternatives that were not chosen and hence were not directly observed in our survey data.

1.1. Research Questions

The study develops a model that predicts mobility-resource ownership, depending on personal accessibility to travel. Personal accessibility to travel (or simply, personal accessibility) is defined and quantified in this paper as: *“the total utility an individual incurs for trip making that is required to participate in a personal set of activities (called PAS—see the definition in Section 2.1).”* As in this definition, accessibility thus depends on the characteristics of a specific person (as it is defined on his/her personal set of activities), we distinguish it from more traditional definitions of accessibility by adding the adjective ‘personal.’ Moreover, personal accessibility depends on a person’s environment: it especially depends on the locations of activities in the personal set, and on the supply characteristics (e.g., speed, price, convenience) offered by different transport modes and services. An ownership model built on personal accessibility would hence predict market shares that are a function of both the population characteristics and the supply design. This is of particular interest to many authorities and mobility service suppliers. For sustainability or economic reasons, these entities wish to know the impact of potential supply changes (e.g., policy measures, investments, effective operations) on mobility-resource ownership in a specific region and population. The first research question is thus:

- How to formulate a model that predicts mobility-resource ownership while taking into account personal accessibility?

It is assumed that mobility-resource (e.g., private vehicle, bicycle, public transport ticket, car-sharing program) ownership decisions and daily transport mode choices are inter-related. Long-term decisions, such as choosing the most appropriate mobility resources, determine the access to related transport mode(s). Short-term decisions such as choosing the most convenient transport mode might reinforce the position of particular mobility resource(s) in personal holdings. For example, buying a car enables an individual to cover her entire travel activity space by car and, in turn, using the car frequently stimulates her to keep the car. It also makes her likely to buy a new car when facing a decision to update her mobility resources. Le Vine [1] has formalized such a functional relationship using the maximum random utility of the discrete choice model. His model assumes rational decision-makers develop their mobility-resources ownership decisions by balancing the utilities projected over the short- and long-term. This study elaborates on this model, and its development is presented in Section 2.

Once a theoretical structure for the model has been defined, the following question regarding ownership factors arises:

- What are the mobility-related variables influencing mobility-resources ownership decisions?

Section 2 reports on the explanatory variables used in this study and frames them in the context of the related literature about mobility-resource ownership. The model proposed in the first part of this study considers two determining factors that need to be further specified. First, there are the fixed costs of acquiring mobility resources and transport modes. Second, the personal accessibility requires specification of the personal travel patterns and of the utility components considered for each trip in the pattern.

1.2. Contributions

This study continues the development and application of a mobility-resource ownership model structure originally proposed by Le Vine [1]. Unlike in the original model, here the travel utility component is decomposed into two terms. One of the travel utility terms is connected to personal accessibility, and the other term is associated with frequency of travel activities. Such a refined model structure offers new insights into the ownership decisions on mobility resources. Applying the updated model to real-world household travel survey data from Germany, it appears that models for the correlation in unobserved factors among mobility alternatives must be chosen with caution. This study extensively reports on the estimation process that guides different discrete choice models. As a result, a set of models producing coherent predictions with satisfactory accuracy is found. Another important contribution of this study lies in simulating input data required by the model. Such simulation is required, as for model estimation the modeller needs to dispose of values for the utilities of the non-chosen alternative modes for trips with missing geospatial references, as well as utilities of non-chosen sets of mobility resources.

The main contributions of this study can be summarized as follows:

- The empirical analysis confirms the significant influence of personal accessibility to travel patterns on ownership decisions on mobility resources.
- The travel patterns that influence mobility-resource ownership can be approximated by a subset of trips observed during multiple days. In the model-estimation process, the importance of trips is determined by how these trip subsets are perceived by the individuals (while deciding on mobility-resource ownership).
- The study provides evidence that a household travel survey that lacks geospatial referencing and non-chosen alternatives is sufficient for modelling ownership decisions on mobility resources. Using data simulation techniques, the missing records are created and added to the travel survey. In addition, it appears that the quality of simulated travel data is critical for the model's calibration and use.

The rest of this article is organized as follows: Section 2 presents the methodological foundation of the model and includes factors influencing ownership decisions and a detailed description of the original model; Section 3 introduces a household travel survey used for empirical analysis and also presents the simulation of data required by the model; Section 4 reports on the actual empirical analysis and involves the model estimation and subsequently the prediction and sensitivity analysis obtained by the fitted models.

2. Mobility-Resource Ownership Model

This section introduces the Perceived Activity Set (PAS), which represents the travel patterns of individuals. Additionally, Le Vine's original model is presented in detail. Finally, the model refinements proposed in this study are introduced.

2.1. Factors Influencing Transport-Ownership Decisions

The existing body of research on transport ownership has recognized the critical influence of personal or household socio-economic background on mobility-resource ownership. It is likely that

wealthy people tend to own more cars when compared to socially or economically deprived people. In particular, the influence of income on vehicle ownership has been used successfully in vehicle ownership modelling for decades [2]. Although such a description is valid [3], it is not exhaustive. This is evident from the example shown in Figure 1 that displays the distribution of household cars among different income groups observed in the German Mobility Panel (MOP). The high-income group tends to own more cars. Yet, the distribution suggests that also more than half of the low-income households (58%) own at least one car. The example indicates the existence of other motives, beyond socio-economic factors, that influence ownership. Therefore, it is useful to account for alternative factors in mobility-resource ownership models.

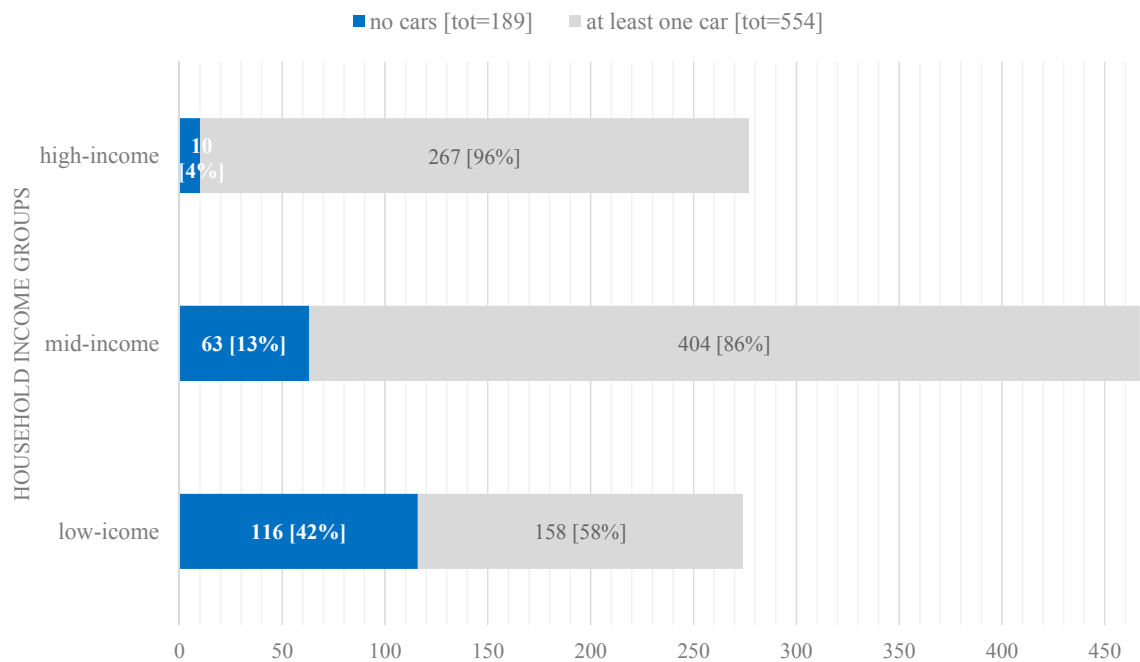


Figure 1. Distribution of income groups by car ownership.

It is reasonable to assume that car ownership is higher in areas with limited provisions of car-alternative services. Therefore, recent attention has been focused on the effect of introducing new or modified mobility services. According to Fagnant, Kockelman, and Bansal [4], shared automated vehicle programs might provide a viable alternative to private-car ownership and lead to a reduction in car ownership. Yang et al. [5] showed that car ownership increases car use, however the study did not detect any significant impact on ownership based on commuting distance or time. The authors accounted for the results by referring to the specific conditions of the urban areas under study. According to the study of Mulalic, Pilegaard, and Rouwendal [6] the extension of the metro system in Copenhagen led to a slight decrease in car ownership. At the same time, the city centre experienced a substantial increase in residents. Their study also hints at another important aspect of ownership decisions. They are typically developed during a longer period, similar to household relocation decisions, and therefore they can be referred to as strategic decisions. The significance of the aforementioned studies is in the way they accommodate the factors that influence ownership. In contrast to the vast body of research, which projects the socio-economic background exogenously on transport ownership, the studies about supply provision factors also take into account the actual usage of transport modes. The latter was proved in the study of Simma and Axhausen [7], which explored the influence of public transport (PT) season tickets and car acquisitions on mode choice. It concluded that the commitment to a certain transport mode promotes usage of this mode. Commins and Nolan [8] also support the claim that transport supply characteristics (costs, travel times) affect vehicle ownership.

Le Vine [1] advanced the latter approach and explored the link between using and owning mobility resources given a certain degree of supply provision. This led to the development of the concept: Perceived Activity Set (PAS). The PAS is defined as, “the array of activities which persons view (at a particular point in their life) as encompassing their potential travel needs when developing decisions that structurally affect their accessibility.” The decisions refer to, for example, buying a car or relocating a household. Under the PAS concept, people are assumed to be aware of their most common travel activities (i.e., commuting to work, weekly shopping and regular leisure activities) as well as characteristics like travel time and out-of-pocket costs in the different mobility-resource ownership scenarios. Additionally, Le Vine noted the importance of PAS travel activities perceived by people while making ownership decisions. For example, mandatory travel activities, such as commuting, might be perceived by people as more determinable than random, optional activities. If a bicycle is considered as the most convenient transport mode for commuting, then people likely include the bicycle in the set of owned mobility resources. Hypothetically, this may even reinforce the bicycle as the mode of transport for other travel activities. Le Vine et al. [1] has estimated the importance of different travel activity purpose types (e.g., to work, leisure, shopping). Astegiano et al. [9] also used the PAS concept, but the importance of a particular travel type was set a priori, being proportional to their frequency. The current study also employs the PAS concept and the importance is estimated for the different travel activity types.

It is worth noting that the concept has some limitations. For example, the data needed for PAS would ideally encapsulate all relevant activities considered by individuals while considering mobility-ownership decisions. However, this approach would be extremely expensive and surely beyond the capacities of any household travel survey (as a common source of travel activity data). A reasonable compromise is to use time-constrained, multiple-day travel observations. The assumption is that the multiple-day observations sufficiently represent the personal travel activity space, as seen by people while considering ownership decisions on mobility resources. In the future, similar concerns might be eliminated with big data and new methods of managing data (e.g., integrating data from several resources).

2.2. Mobility-Resource Ownership Model Utility Functions

The model predicts people’s subset selection from a finite set of mobility resources, which is referred to as Restricted Choice Set (RCS). Each RCS choice alternative d involves a combination of mobility resources R_d and transport modes M_d enabled by these resources. An example of RCS is portrayed in Table 1. People face a decision between two RCS alternatives labelled as RCS_C and RCS_{CP} . The first RCS includes a car as mobility resource that enables a car ride (transport mode by car). Additionally, people may also use PT, but they must pay per trip for a ticket. The other alternative, RCS_{CP} , registers the car and the PT season ticket as mobility resources. These resources enable the car mode, and unlike the preceding RCS, they grant free access to PT. Both RCS include the mode by foot. Note that Le Vine used a similar choice set structure. He referred to it as mobility resource “portfolios.” To avoid confusion with finance and investment terminology, the RCS term is used in this research. The term RCS refers to the fact that the modal choice for daily trip making is restricted by the subset of modes enabled by the chosen RCS.

Table 1. An example of choice set structure with two restricted choice set (RCS) alternatives.

Choice Alternative d	Enabled Mobility Resources R_d	Enabled Transport Modes M_d
RCS_C	Car	By car (driver), by PT (paid), by foot
RCS_{CP}	Car, PT season ticket	By car (driver), by PT (free), by foot

Used acronyms: PT . . . public transport.

The model structure introduced by Le Vine et al. [1] quantifies the utility of a certain RCS alternative perceived by people as the sum of two components. The first component encapsulates non-travel utility associated with owning mobility resources belonging to the given RCS. The second term encapsulates the total travel utility. The model structure is defined as:

$$U_d^i = V_d^{i,\text{non-travel}} + V_d^{i,\text{travel}} + \varepsilon_d^i \quad (1)$$

which can be further specified as:

$$U_d^i = \sum_{r=0, r \in R_d}^R V_r^{i,\text{non-travel}} + \sum_{j=1}^{J^i} \gamma_{ji} \cdot \frac{1}{\lambda^{\text{travel}}} \cdot \ln \sum_{m=1, m \in M_d}^M e^{(V_{m,j}^{i,\text{travel}} \cdot \lambda^{\text{travel}})} + \varepsilon_d^i \quad (2)$$

The term $\sum_{r=0, r \in R_d}^R V_r^{i,\text{non-travel}}$ accounts for the total utility that arises from particular travel resources belonging to the RCS. Among others, it may involve fixed costs like acquisition or registration costs, maintenance expenses or insurance. The term $\ln \sum_{m=1, m \in M_d}^M e^{(V_{m,j}^{i,\text{travel}} \cdot \lambda^{\text{travel}})}$ quantifies the “optimal” transport mode utility perceived by people. For this, it is assumed that the effect of the cost on the travel utility obtained by mode m is heterogeneous and follows a Gumbel distribution with variance λ^{travel} . The expected maximum perceived utility (EMPU) among the available modes enabled by the RCS can then be written analytically in closed form as the given log-sum of systematic utilities per mode. The term $\sum_{j=1}^{J^i} \gamma_{ji} \cdot (\dots)$ enumerates all PAS activities of an individual and accounts for the importance of the j -th activity by the term γ_{ji} . The notation $r \in R_d$ and $m \in M_d$ refers to the mobility resources and transport modes, respectively, that were enabled by the RCS choice alternative d . Finally, ε_d^i is an error term that captures the influence of all unobserved factors affecting RCS choice.

The key assumption here is that people choose RCS by taking into account the trade-off between paying for mobility resources and getting access to a certain quality for the considered PAS activities. For example, people who pay for more mobility resources may utilize broader (more diverse) and possibly higher-utility travel options for the activities in their PAS.

2.3. Choice Model Structure

The RCS choice model will also call the “second step” model. (The “first step” is the travel mode choice for every trip in the PAS, which is embedded in the utility function of the second step.) The second step involves the estimation of the non-travel costs, travel frequency and the personal accessibility to PAS parameters (explained in Section 2.4). The latter parameter associates with the variable $A_{p,d}^i$ (see the description in Section 2.4.2) that is given as the average travel cost over a set of PAS activities. For each activity, the most “useful” mode (the optimal travel alternative) should be chosen out of the ones enabled by the RCS. Hence, a mode choice is embedded in $A_{p,d}^i$ and the parameters of which (coefficients of variables affecting the mode choice of each PAS trip) must be estimated first. Thus, [10] used an estimation process split into two steps: the parameters of travel utility component are estimated in the first step; and subsequently, in the second step, the RCS choice model parameters are estimated.

2.4. Mobility-Resource Ownership Model Development

So far, the study has mostly reported on the original model introduced by Le Vine et al. (2013). The following Sections 2.4.1 and 2.4.2 present the contributions of this paper related to the model. (The order of subsections follows the steps of the model-estimation process.)

2.4.1. The First Step: Specification of the Mode Choice Model for the PAS (the Short-Term Decisions)

As explained in Section 2.3., the RCS model (second step) requires quantification of the EMPU of modal choice for PAS trips. Thus, we first need a definition of utility and an estimation of the

parameters therein. Following Le Vine, we do this in the first step, which considers mode choice only. At this point, we assume RCS per individual to be exogenously given and observe their modal choice for all trips contained in the PAS data of our population.

We propose a utility definition in the following form, driven by only travel-time cost (tt) and out-of-pocket cost (OPC):

$$V_m^{i,\text{travel}} = \beta^{\text{tt}*} \cdot \text{tt}_{m,i}^i + \beta^{\text{OPC}} \cdot \text{OPC}_{m,i}^i + \text{ASC}_m \quad (3)$$

Upon estimation of the coefficients $\beta^{\text{tt}*}$, β^{OPC} , and ASC_m , Le Vine et al. [1] assumed that the transport mode choices for the PAS are independent. Astegiano et al. [9] relaxed the assumption, considering the fact the PAS data involve repeated transport mode choices of the same individual (panel data) and hence perception errors are correlated. Therefore, a mixed logit (MXL) model was introduced, accounting for correlations in error term and relaxing the independence of assumed irrelevant alternatives considered by the Multinomial Logit (MNL). Also, in this paper, the MXL structure is used for the transport mode choice model.

Adopting MXL means that the travel-time parameter $\beta^{\text{tt}*} \sim N(\mu^{\text{tt}}, \sigma^{\text{tt}})$ is assumed to vary across the population where the mean μ^{tt} and the standard deviation σ^{tt} are two estimated parameters. Thus, the travel-time parameter accounts for a random taste variation and for correlation in unobserved factors over time for a given individual. The out-of-pocket cost parameter is constant across the population.

2.4.2. The Second Step: Specification of the RCS Choice Model (the Long-Term Decisions)

We know from the previous step the mode choice model parameter estimates (the results of the first step), which can be used for computation of the EMPU in modes for trips in one's PAS. Therefore, we can now proceed with specifying the utility U_d^i that drives the second step of the combined choice model: the RCS choice.

We choose to split the travel utility term $V_m^{i,\text{travel}}$ of Equation (1) for the utility U_d^i in two terms: the average personal accessibility to PAS provided by transport modes $A_{p,d}^i$, and the frequency of travel PAS activities N_d^i . Both terms are weighted by the respective parameters γ_p and β_d^N . To measure the effects of the two parameters independently of each other, the personal accessibility to PAS is averaged dividing by the number of trips N_p :

$$A_{p,d}^i = \frac{\sum_{j^i=1}^{J^i} \ln \sum_{m=1}^M \sum_{m \in M_d} e^{(V_{m,i}^{i,\text{travel}})}}{N_p} \quad (4)$$

We further define that $A_{p,d}^i$ is specified for five pre-defined travel activity purposes: p: work, shop, leisure, escort, and other; the details are presented in Section 3. For each individual i , the personal accessibility is evaluated for his or her unique PAS, which is composed of $j^i = 1 \dots J^i$ travel activities. This reflects the utility incurred by the travel modes belonging to RCS d . The final specification of the RCS utility function used in this research is shown in Equation (5). The frequency of travel activities does not vary among RCS, but only among individuals. Hence, including both an alternative specific constant and the frequency of travel activities would result in model misspecification (two degrees of freedom for defining one constant value). Therefore, only the latter is used.

$$U_d^i = \beta^{\text{FC}} \sum_{r=1}^R \text{FC}_r^i + \beta_d^N \cdot N_d^i + \sum_{p=1}^P \left(\gamma_p \cdot A_{p,d}^i \right) + \epsilon_d^i \quad (5)$$

Explicitly recognizing the number of PAS travel activities and the personal accessibility to PAS in Equation (5), compared to the original utility definition in [1,9] shown as Equation (2), allows for making a more refined distinction between choice situations. Let us consider two individuals,

one typically accomplishes many short-distance trips, and the other does only a few long-distance trips. For example, individual i1 with a PAS consisting of 10 activities given by $tt_{1,1,\dots,10}^{i1} = \{5, 10, 10, 10, 5, 5, 15, 5, 5, 10\}$ where the values denote travel times in minutes provided by the optimal travel mode m1. Individual i2 has a PAS given by $tt_{2,1,2}^{i2} = \{40, 40\}$ using the optimal mode m2. For both, the out-of-pocket cost is zero and all activities are of the same type. According to the new utility definition, the effect of the personal accessibility to PAS will be different for both individuals—it will be more significant for i2. Although it is not possible to exactly interpret the effects of the number of trips (recall that the parameter also captures the effect of alternative specific constant), it is likely that its effect will also be different for the two individuals. It appears that the new utility definition delivers some additional meaningful insights into the ownership decisions on mobility resources.

In the model specification of Equation (5), not only the systematic utility terms differ from the original definition by Le Vine. We also make different assumptions on the perturbation ε_d^i . The original study used the MNL for the RCS choice (the second step) assuming that the ε_d^i of the RCS alternatives are mutually independent [1]. [9] estimated a cross-nested logit (CNL) and a nested generalized extreme values (NGEV) model for the second step. The CNL captures the correlations of both, the travel and the non-travel utility perception among RCS alternatives having the same mobility resources. NGEV generalizes the CNL model, using a multilevel, cross-nested correlation structure. A more detailed description of NGEV is provided in [11]. The need for a complex correlation structure arises from the characteristics of the RCS. Each RCS combines the properties of mobility resources and transport modes common to the different alternatives. The proposed modifications were also tested in this study. However, with the dataset used in this paper, it was not possible to find a coherent CNL or NGEV model for this study. (The model parameters were found insignificant.) On the other hand, it turned out that a relatively simple nested logit (NL) can be estimated with all significant model and utility parameters and with expected signs. Therefore, only the results of the NL will be presented. The estimation process is described in Sections 4.1.1–4.1.3 in detail.

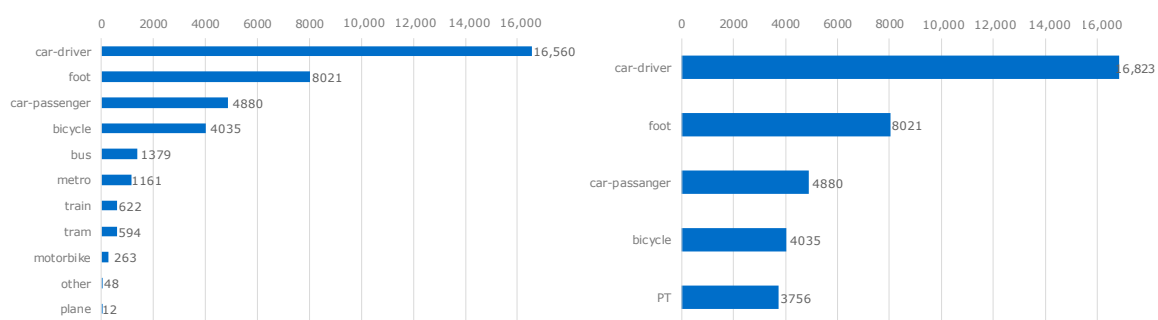
3. Data

The data used in this research comes from the MOP [12] that consists of panel observations collected continuously from 1994 and stratified over the German states. The following MOP data files from the year 2008–2009 were used:

- The Personal file containing the socio-economic information on 1783 individuals.
- The Household file containing the background information on 1062 households of individuals from the Personal file.
- The Trip file containing 43,029 travel activities recorded over a course of 7 days by the personal file individuals.
- The Vehicle file containing car information such as type, fuelling, or age.

In order to guarantee comparability to future work in which we intend to estimate similar mobility-resource ownership models on a household level, we retained only the households with all eligible household members participating in the travel survey. This reduced the sample to 902 households. For the sake of significance of the parameter estimation, we avoided segments in the data with insufficient observations by aggregating the transport modes and trip purposes according to the schemes shown in Figures 2 and 3, respectively. The trips done by the modes of transport “other” and “plane” were removed. The trips reported as by “bus,” “tram,” “train,” and “metro” were aggregated to the transport mode PT. Merging urban, regional, and inter-regional types of PT into one transport mode may oversimplify the level of service provided by the particular PT types. Therefore, the interpretation of results for the PT transport mode must be taken with caution. However, this simplification maintains the size of the choice set moderate. Next, the travel activities “to home,” “to second home,” “back to hotel,” and “same origin and destination” were discarded because they represent journeys correlated with the first leg of trip chains. Therefore, their contribution to the explanation of ownership decisions

on mobility resources is already accounted for in the first leg and we remove them to avoid redundancy. It is worth noting that such assumption might be considered in the current data set without any shared mobility options. In applications where some forms of shared mobility are available, it is desirable to evaluate the entire trip chain, allowing for more complex personal mode choice tactics. The travel activities “pick-up/drop-off someone” and “other” were aggregated to the single category “other.” Finally, the RCS alternatives were formed using the three mobility resources: bicycle, PT season ticket, and car. For the latter, car and motorcycle were considered to be one mobility resource. The RCS used in this research is explained in Section 3.1.



used acronyms: PT ... public transport

Figure 2. Transport mode distribution before (left panel) and after (right panel) the aggregation.

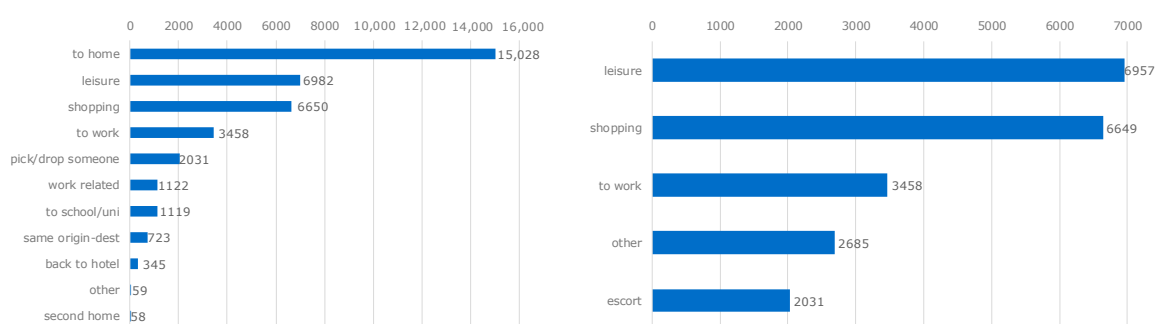


Figure 3. Trip purpose distribution before (left panel) and after (right panel) the aggregation.

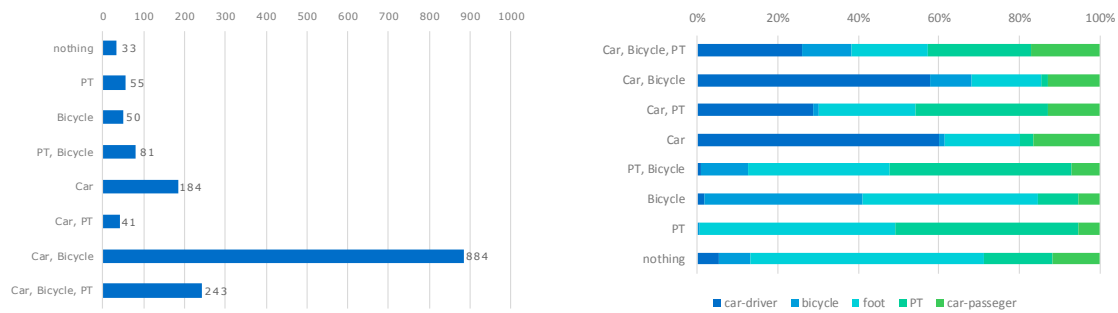
3.1. Data Limitations

One of the main reasons for using the MOP data is that the travel survey involves multiple day observations that are required by the PAS. However, there exist a few concerns limiting the use of the data in this research.

First, unlike the PT season ticket or bicycle ownership which were reported at the personal level, the car ownership was reported at the household level. It is reasonable to assume that cars are considered as a household rather than individual property. However, this is inconsistent with the model that was defined at the individual level. It was decided to consider the vehicle ownership as if it had been reported by individual survey participants. Therefore, the total number of vehicles reported in the survey is higher than in reality. A possible solution may involve the modelling of ownership decisions on mobility resources at the household level, taking into consideration the household consumption patterns and related interpersonal (but intra-household) strategies. The resolution of this concern is beyond the scope of this study.

Second, inspecting the full factorial RCS scheme displayed in Figure 4 (left panel), which consists of eight unique alternatives, it turned out that the car-free RCS were significantly underrepresented. This considerably hinders the estimation process. Therefore, it was decided to keep only the RCS alternatives with a car. The final design of choice set formed by the four RCS alternatives is shown in

Table 2. Note that three transport modes are common to all RCS alternatives: by car (driver), by car (passenger), and on foot. An individual without a PT season ticket (RCS_C , RCS_{CB}) may still opt for a PT ride. In such a case, an out-of-pocket cost of 2.5 EUR is charged for PT trips less than or equal to 60 min and 5 EUR for PT trip(s) exceeding 60 min. The concatenation of PT journeys was also taken into consideration while computing the PT journey costs. Note that the PT pricing parameters were chosen arbitrarily, representing the average PT fares in Germany).



used acronyms: PT ... public transport

Figure 4. RCS market shares (left panel) and RCS modal split (right panel).

Table 2. Overview of mobility resources and transport modes forming the RCS used in this research.

Choice Altr. d	Enabled Mobility Resources R_d	Enabled Transport Modes M_d
RCS_C	Car	By car (driver), by car (passenger), by foot, by PT (paid)
RCS_{CB}	Car, Bicycle	By car (driver), by car (passenger), by foot, by PT (paid), by bicycle
RCS_{CP}	Car, PT season ticket	By car (driver), by car (passenger), by foot, by PT (free)
RCS_{CPB}	Car, Bicycle, PT season ticket	By car (driver), by car (passenger), by foot, by bicycle, by PT (free)

Used acronyms: PT ... public transport.

Finally, the MOP contains only the observed travel records. However, the model also requires the non-chosen alternatives information. Section 3.2 explains how this problem was resolved in this study.

3.2. Simulating the Non-Chosen Alternatives

Although the lack of non-chosen alternatives data is a common problem for the revealed-preferences-based discrete choice experiments, little is known about reliable methods of inferring the unobserved data. [1] used a web-scraping technique that extracts the requested data in batch from websites, for example, travel times from a PT operator website. [9] augmented the observed travel records by the non-observed travel times downloaded through Google API. In both cases, the original surveys contained the geographical references. Due to privacy reasons, the MOP stores the travel records without any geographical information. Other researchers in similar situations rely on simulation of the non-observed values using statistical models. [13] forecasted the non-observed travel behaviour data (mode choice rates) by a multinomial logit with parameters obtained by Bayesian inference. An auxiliary calibration sample was used to extract the Bayesian priors. The non-observed travel times were estimated from the posterior distribution that exploits the priors and the observed survey data.

In this study, the non-observed travel time values $\hat{t}_{nki}^{i,b}$ are simulated using a linear regression model (LRM). Simulated travel times are used for the non-chosen alternatives while reported values are employed for the chosen alternatives. This asymmetry may have affected the parameter estimation. Further validation could be performed by using simulated data for all alternatives, which would indeed achieve symmetry, though at a loss in precision for the chosen alternatives.

The basic idea is to predict the travel time given the distance, the mode, and the region. For each region b , the LRM regresses the non-observed mode travel times on the observed travel distance $d_{m,j}^{i,b}$ for the given record:

$$\hat{t}_{n_k}^{i,b} = d_{m_j}^{i,b} \cdot a_{1,n}^b + a_{0,n}^b \quad (6)$$

where index m_j denotes transport mode m chosen for journey j while n_k is related to the unobserved journey k performed with alternative mode n . In a given region, LRM parameter $a_{1,n}^b$ and constant $a_{0,n}^b$ are estimated for each transport mode separately assuming the constant speed. Note that the regional characteristics refer to one of sixteen states (regions) in Germany originally recorded in the MOP. For each of 16 states, five different LRM (one for each transport mode) was estimated. The main advantage of this approach is the simplicity of the simulation process that also ensures straightforward interpretation of results. While a linear function appeared to sufficiently approximate the relationship between travel time and distance for bicycle and walk, the same relationship for the PT and car transport mode may be non-linear. In such cases, the simulation results must be used with caution, and other modelling techniques may be required. The observed and simulated values of travel time for the considered transport modes are visualized in Figure 5.

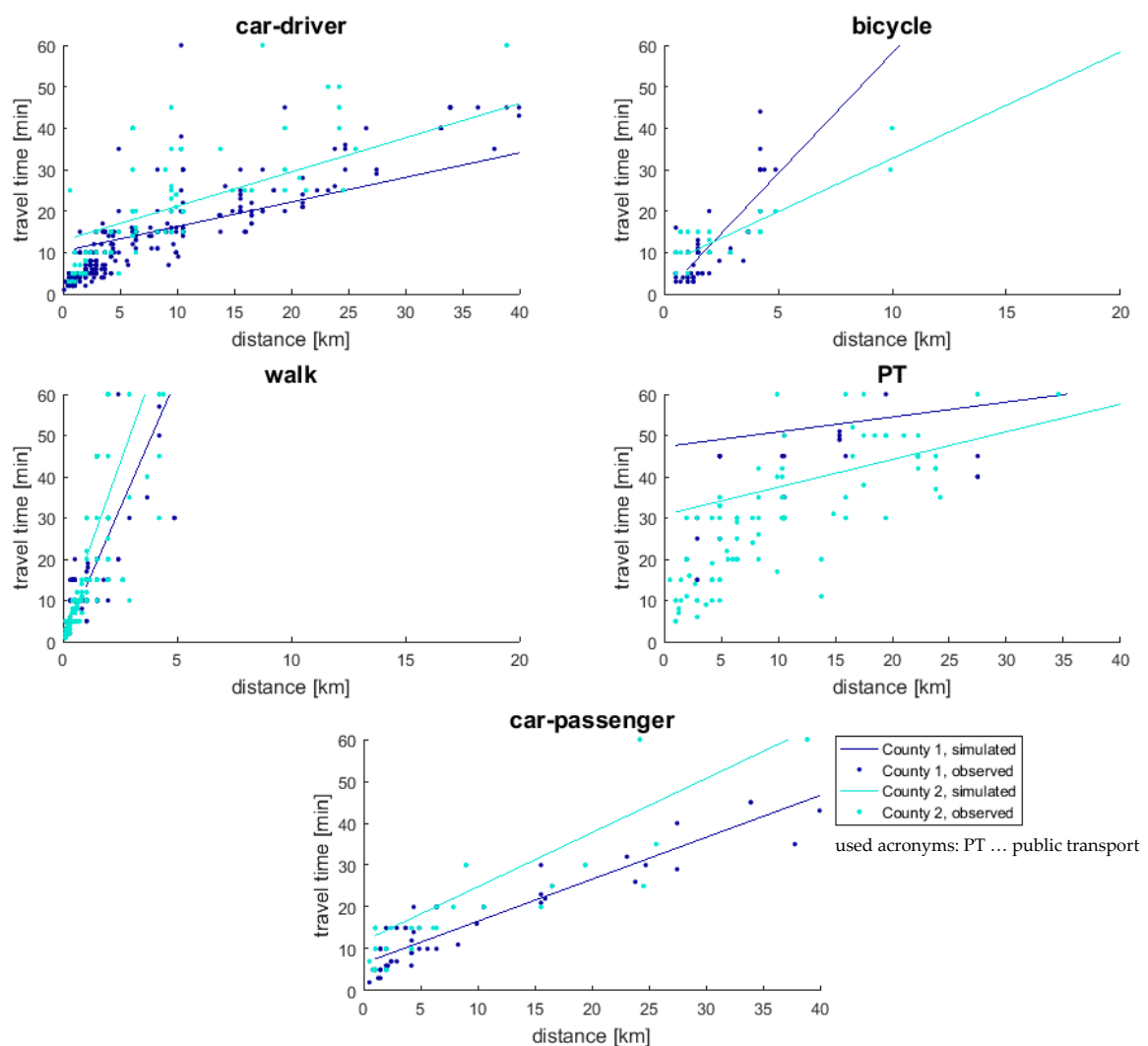


Figure 5. Observed and simulated travel times for different transport modes and regions (only 2 of 16 regions reproduced here in order to not overload graphs).

Figure 5 also shows the distribution of trips over the travel distance. It can be concluded that the short distance trips are more frequent than the long-distance trips. However, the LRM still attempts to capture the less frequent observations placed further away from zero. As a result, the long-distance observations might significantly leverage the slope of the function. Such influence could be especially harmful in the presence of outlying records, which lie outside the pattern suggested by the majority of data points. Table 3 presents mean and standard deviation (SD), and their robust counterpart, median and median absolute deviation (MAD) of the observed travel times for the considered transport modes. The substantial differences between the values of mean and median, and similarly the differences between the values of SD and MAD suggest that there indeed exist influential outliers. For example, the lower median values compared to the mean values suggests that the travel times for all modes are right-skewed.

Table 3. Statistics of the observed travel times for the considered transport modes.

Measure	Observed Travel Time				
	Car-Driver	Bicycle	Walk	PT	Car-Passenger
Mean	18.16	40.81	149.73	38.97	19.98
Median	13.43	18.18	51.58	33.92	16.15
SD	25.03	119.94	459.48	26.56	25.34
MAD	2.13	8.83	36.77	2.70	3.07

Used acronyms: SD ... standard deviation, MAD ... median absolute deviation, PT ... public transport.

In order to overcome the problem of outliers, a robust linear regression model (RLRM) [14,15] was used. In contrast to the LRM, the effect of outliers is downsized by the RLRM. Both models were assessed by computing the relative mean absolute error (RMAE) and root mean squared error (RMSE) for the pairs of simulated and observed travel times. The results are presented in Table 4. While the RLRM has more accurate results according to the RMAE, the LRM delivers better results when inspecting the RMSE. Such opposing conclusions can be explained by investigating the contribution of particular errors. It is likely that the RLRM improves the errors of observations closer to zero ignoring the influence of the longer distance observations. Finally, the RLRM outperforms the LRM according to the RMAE, which is more sensitive to the short distance residuals. Because most of the values lie close to zero, the results suggest that the RLRM delivers more suitable results. This claim is further examined in the empirical analysis presented in Section 4 where all models are developed for the data simulated by LRM and RLRM.

Table 4. Evaluation of linear regression model (LRM) and robust linear regression model (RLRM) for different transport modes.

Statistics	Model	Car-Driver	Bicycle	Walk	PT	Car-Passenger
RMAE	LRM	0.66	0.52	0.63	0.57	0.73
	RLRM	0.41	0.38	0.45	0.44	0.43
RMSE	LRM	11.37	9.47	20.37	27.91	16.38
	RLRM	14.64	11.06	23.06	42.59	32.41

Used acronyms: RMAE ... relative mean absolute error, RMSE ... root mean squared error, PT ... public transport.

4. Use Case: Predicting Mobility Resource Market Shares

4.1. Estimation

In this step, the transport mode choice model defined by Equation (3) is estimated using the MOP. Tables 5 and 6 report the results for the MXL model including the non-robust and robust simulated travel times, respectively. Hereafter, we abbreviate travel times predictions by the robust (or non-robust) estimation method by the term “robust travel times (or non-robust travel times).”

Table 5. Mixed logit (MXL) model using the non-robust travel times and observed data.

		Value	Std Error	p-Value
Utility Parameters	ASC transport mode car	−0.842	0.034	0.000
	ASC transport mode bicycle	−1.890	0.037	0.000
	ASC transport mode walk	0 ^f		
	ASC transport mode PT	−0.784	0.057	0.000
	Travel time [min]	−0.120	0.005	0.000
	Variance of travel time [min]	0.128	0.005	0.000
	Out-of-pocket cost [EUR/journey]	−0.164	0.031	0.000
Model Statistics	Adjusted ρ^2	0.320		
	Final log-likelihood	−11,316.893		
	Number of observations	19,201.000		

f . . . fixed value.

Table 6. MXL using the robust travel times and observed data.

		Value	Std Error	p-Value
Utility Parameters	ASC transport mode car	−0.368	0.032	0.000
	ASC transport mode bicycle	−1.570	0.035	0.000
	ASC transport mode walk	0 ^f		
	ASC transport mode PT	−0.744	0.051	0.000
	Travel time [min]	−0.080	0.004	0.000
	Variance of travel time [min]	0.101	0.004	0.000
	Out-of-pocket cost [EUR/journey]	−0.139	0.139	0.000
Model Statistics	Adjusted ρ^2	0.250		
	Final log-likelihood	−12,775.901		
	Number of observations	19,201.000		

f . . . fixed value.

The results obtained by the two models are consistent: the parameters have the expected negative signs, all of them are statistically significant, and convergence was reached. The model estimated on the non-robust travel times achieved higher Adjusted ρ^2 compared to the model using the robust travel times. One possible explanation might be due to the impact of (less-frequent) long distance observations, which tend to be less accurately modelled by the RLRM. Having everything else equal, the ASC suggests that walk is always the most preferred transport mode, and contrarily bicycle is the least preferred transport mode.

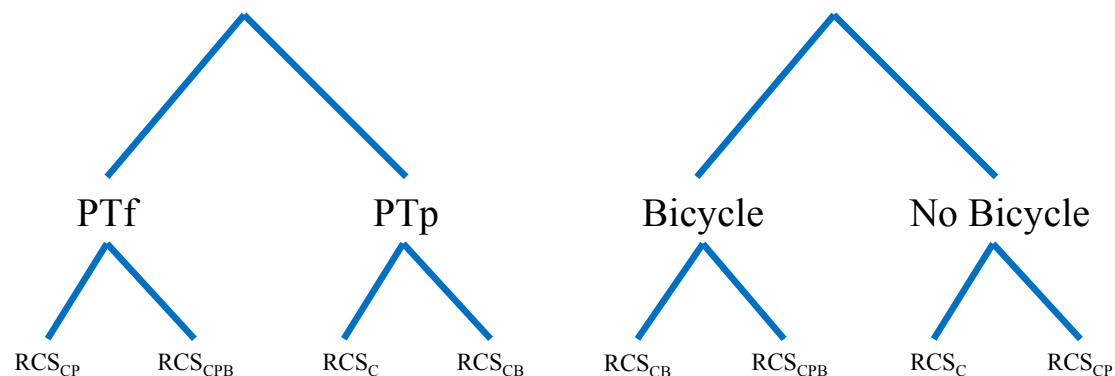
4.1.1. Nested logit NL Estimation Results

In the second step, the RCS choice model defined by Equation (5) is estimated. As explained in Section 2.4.2, the second step model uses the individual parameters of the preceding step to calculate the $A_{p,d}^i$. Two candidate nest structures (Figure 6), which could reflect the correlations among the alternatives sharing the same mobility resources, were tested. The first model groups together the RCS alternatives with (PTf nest) and without (PTp nest) the PT season ticket ownership. The second model groups together the RCS alternatives with (Bicycle nest) and without (No Bicycle nest) the bicycle ownership.

Using the non-robust simulated input, the estimation of the model failed to reach convergence with either of the proposed nest structures. The results are shown in Appendix A, Tables A1 and A2. Using the robust simulated data, the two models did converge. However, the PT nest model has its nest parameter equal to one suggesting that the nest structure is meaningless. The results can be found in Appendix A, Table A3. The estimation of PT nest model failed even when starting from the results of the successfully estimated Bicycle nest model. The most likely explanation of why the model with PT nest structure failed (even if estimated on the robust travel times) is that the utility of including a public transport season ticket in the RCS is not considered by decision-makers. This is reasonable because the survey participants registered a relatively small number of travel activities by PT. It turns

out that only the bicycle nest model estimated on the robust input delivers reliable results which are presented in Table 7. Therefore, only this model qualified for subsequent analysis.

As expected, the fixed costs parameter is negative, indicating that people prefer less expensive mobility resources. An increasing number of travel activities mostly disfavour the alternative with the broadest set of mobility resources. In line with expectation, the important parameters associated with the quality of access are all positive (the computed personal accessibility values according to Equation (4) are negative). Among the importance parameters, the leisure activities are perceived the most decisive. The personal accessibility to leisure activities is valued much higher than the other considered types of PAS activities.



used acronyms: PTf ... Public Transport Season Ticket nest (free access), PTp ... Public Transport Season Single ticket (paid access), RCS_C ... restricted choice set (car), RCS_{CB} ... restricted choice set (car, bicycle), RCS_{Cp} ... restricted choice set (car, public transport), RCS_{CPB} ... restricted choice set (car, public transport, bicycle)

Figure 6. Nested logit (NL) model nest structure recognizing the public transport (PT) season ticket ownership (**left panel**); and the bicycle ownership (**right panel**).

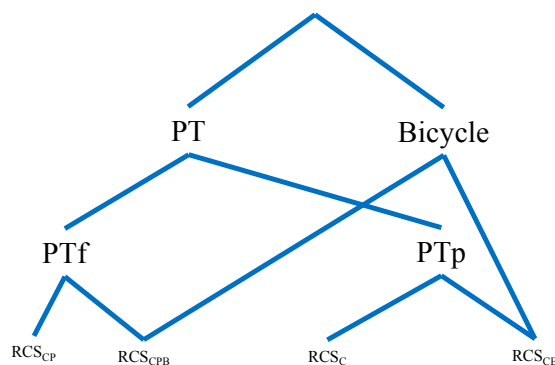
Table 7. NL with the bicycle nest and the robust dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.017	0.003	0.000
	Travel activity frequency RCS (car)	0 ^f		
	Travel activity frequency RCS (car, bicycle)	−0.065	0.008	0.000
	Travel activity frequency RCS (car, PT)	−0.009	0.012	0.000
	Travel activity frequency RCS (car, bicycle, PT)	−0.115	0.011	0.000
	Importance of ‘escort’ PAS activities	0 ^f		
	Importance of ‘leisure’ PAS activities	63.600	6.830	0.000
	Importance of ‘other’ PAS activities	41.400	7.250	0.000
	Importance of ‘shopping’ PAS activities	55.000	6.250	0.000
	Importance of ‘work’ PAS activities	34.600	7.160	0.000
Model Paramters	Nest ‘bicycle’	1 ^f		
	Nest ‘no bicycle’	3.880	0.394	0.000
Model Staitistics	Adjusted ρ^2	0.690		
	Final log-likelihood	−564.428		
	Number of observations	1352.000		

f ... fixed value.

4.1.2. Nested Generalized Extreme Values NGEV Estimation Results

The nested correlation structure used here is shown in Figure 7. The design essentially accounts for the correlations in the unobserved term for the alternatives with the bicycle and PT season ticket ownership. The latter is further divided into the alternatives correlating with the PT access.



used acronyms: PTf ... Public Transport Season Ticket nest (free access), PTp ... Public Transport Season Single ticket (paid access), RCS_C ... restricted choice set (car), RCS_{CB} ... restricted choice set (car, bicycle), RCS_{CP} ... restricted choice set (car, public transport), RCS_{CPB} ... restricted choice set (car, public transport, bicycle)

Figure 7. Nested generalized extreme values NGEV structure used in this study with two intermediate nest levels. The upper nest level recognizes the correlations in unobserved term among the alternatives sharing either bicycle or PT mobility resources. The lower nest level recognizes the correlations in unobserved term among the alternatives granting either free (PTf) or paid (PTp) access.

The NGEV models were estimated for the non-robust and robust simulated datasets. The estimation results are shown in Appendix A, Tables A4 and A5. In both cases, the nest parameters associated with the PT nest were insignificant. Hence, the nesting structure could not be estimated. In conclusion, only the model with the simpler nested structure presented in Table 7 should be used.

4.1.3. Estimation Summary

For the subsequent prediction and sensitivity analysis, the results in Table 7 will be used. The estimation results showed that walk is the most preferred transport mode (*ceteris paribus*). One possible explanation might lie in the trip distance distribution (discussed in Section 3.2). There, it appeared that a substantial amount of travel activity records was trips shorter than or equal to 10 min; other modes may have little added values over walking for such short trips indeed. Additionally, the transport mode choice model accommodates only the influence of travel time and out-of-pocket cost. The alternative specific constant covers factors influencing the transport mode choice (e.g., parking space availability, travel comfort, travel time reliability) not covered by other attributes in the model.

The NL results based on the complete dataset suggest that leisure and shop activities have a stronger influence on the RCS choice than the other type of activities. Importantly, the model has all importance parameters significant. It can thus be concluded that the estimation results confirm the role of personal accessibility to PAS in the ownership decisions on mobility resources of people. The importance of simulating robust travel times is clear from the estimation results as only these simulated data led to valid model parameters.

4.2. Prediction

The successfully estimated model was used to predict the RCS market shares (the second modelling step). The prediction is based on the same datasets as the estimation. The objective of such exercise is primarily to validate the estimated models and assess their predictive power.

In order to inspect the quality of predictions, the predicted probabilities of choosing the observed RCS alternatives were calculated for each individual. Figure 8 shows the distribution of the predicted choice probabilities of the chosen alternative. As suggested in the study of [16], a distribution with its centre located on the right-hand side (negatively skewed) suggests that the truly observed RCS choices were mostly predicted with high probabilities.

Next, the RCS market shares were predicted. Knowing the probabilities of choosing the RCS alternatives for each individual, an RCS market share is equal to the mean of these probabilities for the given RCS over the population. Alternatively, it is possible to simulate the market shares for each individual separately using the probabilities of choosing the RCS alternatives. The advantage of the latter approach is that the miss-matches (the differences between predicted and observed market shares) can be calculated for each alternative. Hence, the capacity of the model to predict specific choice set alternatives can be assessed. The simulation randomly samples the RCS choices using a very large number of draws (exactly 100,000) per each individual. The draws are hereafter reported in a contingency table where rows represent the observed choices and columns represent the predicted choices. Each diagonal cell displays a correct prediction, that is, a model predicts the same RCS as it was observed. On the contrary, off-diagonal cells show the mismatches.

Table 8 presents such contingency table. The distributions were scaled down to the original sample size (the market shares were divided by the number of draws). Overall the model shows good concentration on the diagonal of correct predictions (confirming the results shown in Figure 8). Inspecting the distribution, it appears that the model had difficulties in predicting RCS_{CP} . This is likely due to the low number of RCS_{CP} observations. Moreover, one recognizes certain false-associations among the alternatives. For example, the model is prone to predict RCS_{CB} instead of truly observed RCS_C . Similarly, the models predicted RCS_{CBP} instead of RCS_{CB} . This is likely caused by the fact that the commitment to the RCS alternatives with a bicycle is marginal comparing to the provided quality of access. That is: given the low fixed cost and low travel times that a bicycle provides for some trips, one would rationally expect more users to own it in their RCS than what is actually observed. In reality, however, the quality of access by bicycle might be (negatively) influenced by the factors other than travel time alone and which were not considered in the study, for example, weather conditions.

Table 8. Contingency tables of predicted market shares of the robust NL (right panel).

		Predicted				SUM (observ.)
		RCS_C	RCS_{CB}	RCS_{CP}	RCS_{CBP}	
Observed	RCS_C	111.4	57.1	3.0	12.5	184.0
	RCS_{CB}	50.9	695.7	1.3	136.1	884.0
	RCS_{CP}	2.4	4.1	25.1	9.4	41.0
	RCS_{CBP}	3.2	91.7	16.7	131.4	243.0
SUM (predict.)		167.9	848.7	46.1	289.4	1352.0

Used acronyms: RCS_C ... restricted choice set (car), RCS_{CB} ... restricted choice set (car, bicycle), RCS_{CP} ... restricted choice set (car, public transport), RCS_{CBP} ... restricted choice set (car, public transport, bicycle).

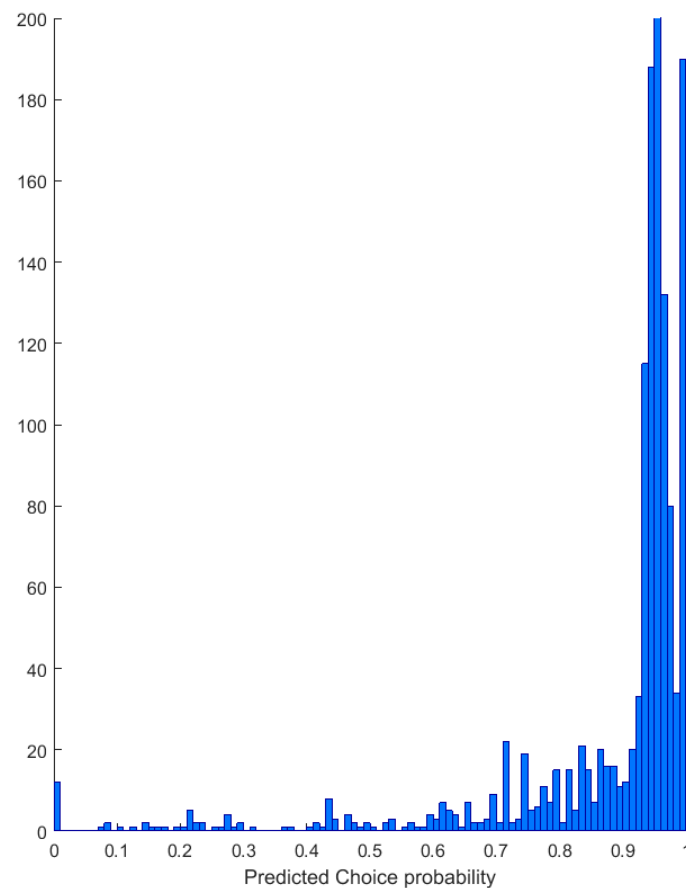


Figure 8. Distribution of the predicted choice probabilities of the chosen RCS alternative for the robust NL.

In order to improve the accuracy of prediction and receive additional insights into ownership decisions on mobility resources, the current analysis could in future be extended by distinguishing population market segments [16] or by including socio-demographic variables as a complement to the cost-related attributes considered in this paper.

4.3. Sensitivity Analysis

The aims of the sensitivity analysis here are two-fold. First, it provides an additional validation level. A valid and coherent model should reasonably respond to input changes, for example, the market shares should move in the expected direction and with an appropriate magnitude. Second, it manifests the fundamental objective of the modelling exercise: the ability to forecast the mobility resource market shares corresponding to supply modifications.

The sensitivity analysis scenarios were chosen to investigate the model's response to changes in travel and non-travel related factors. Four different scenarios were considered:

- “tt50”: the PT travel times were reduced by half.
- “tt200”: the PT travel times were doubled.
- “FC50”: the bicycle fixed costs were reduced by half.
- “FC200”: the bicycle fixed costs were doubled.

For the first two scenarios, the PT travel times were modified according to a given scenario, and the terms $A_{p,d}^i$ were re-calculated using the corresponding first model parameters. Using the corresponding second step model parameters and newly calculated terms $A_{p,d}^i$, the predicted choice probabilities and RCS market shares were produced. For the latter two scenarios, the datasets with the

modified bicycle fixed costs were used together with the corresponding second step model parameters to produce the predicted choice probabilities and RCS market shares.

Table 9 presents the differences in market shares predicted by the NL model, first for the reference case and then for the four scenarios. The first row of the table shows the reference case, displaying the difference between the predicted (without any modifications) and the observed market shares. The values indicate the overall quality of modelled predictions on the reference data. In particular, the smaller the differences are, the more accurately the models replicate the reality. Note that the shown differences correspond to the values presented in Table 8. The explanation of market share modifications for the four scenarios captured in the remaining rows of the table is straightforward. Changing a property of RCS in a certain direction affects the market shares of other RCS without such property (or having that property only as a secondary) inversely. For example, it is reasonable to assume that for alternatives RCS_{CP} and RCS_{CBP} the PT transport mode stands—along with car—for an important (primary) transport mode, recall Figure 4. Hence, it was expected to observe a gain in market shares for these two alternatives and a drop for RCS_C and RCS_{CB} if the PT mode became twice faster. Inversely, the market shares for RCS_C and RCS_{CB} expanded with slower PT travel times. Such reactions are recorded in the second and third row of Table 9. Next, a bicycle is available only to the RCS_{CB} and RCS_{CBP} . Therefore, it was expected that the discount in the bicycle fixed cost stimulates the growth of RCS_{CB} and RCS_{CBP} . Contrarily, the rise in the bicycle fixed cost reduced the demand for RCS_C and RCS_{CP} . These effects are shown in the fourth and fifth row of Table 9. The model sensitivity to the changes of bicycle fixed cost is rather marginal. A likely explanation is that fixed cost of a bicycle is low anyhow (in particular, it corresponds approximately to the price of a single PT season ticket), hence the provided incentives were too small to yield sufficient leverage. As such, the models predicted adequately what would happen due to the changes.

Table 9. Differences in market shares after supply modifications predicted by the NL model. The values in brackets show the relative changes in market shares.

Scenario	Robust Data			
	RCS_C	RCS_{CB}	RCS_{CP}	RCS_{CBP}
Simulation - Observed	−16.2 (−1.2%)	−35.3 (−2.6%)	5 (0.4%)	46.4 (3.4%)
tt50 - Simulation	−2.4 (−0.2%)	−82.8 (−6.1%)	27.7 (2%)	57.5 (4.3%)
tt200 - Simulation	22.6 (1.7%)	77.9 (5.8%)	−34.1 (−2.5%)	−66.4 (−4.9%)
FC50 - Simulation	−1.7 (−0.1%)	1.5 (0.1%)	−0.5 (0%)	0.6 (0%)
FC200 - Simulation	3.3 (0.2%)	−3.1 (−0.2%)	0.9 (0.1%)	−1.2 (−0.1%)

Used acronyms: RCS_C ... restricted choice set (car), RCS_{CB} ... restricted choice set (car, bicycle), RCS_{CP} ... restricted choice set (car, public transport), RCS_{CBP} ... restricted choice set (car, public transport, bicycle).

5. Conclusions

The presented study examined the ownership decisions on mobility resources using a two-step discrete choice model applied to the household travel survey conducted in Germany.

Prior to operationalizing the model, it was necessary to obtain non-observed (non-chosen) travel times, which serve as an input to estimation of the model parameters. Unfortunately, the observed travel behaviour data completely lacks the geospatial referencing. The proposed method simulates the non-observed data via a linear regression model (LRM) relating trip distance to travel time for each mode, adjusted to the local conditions (different LRMs were estimated for different regions). LRM offers a simple, intuitive, easily applicable technique for generating travel times without the need for accurate geospatial information. A detailed look at the travel time distributions for different transport modes revealed the presence of outlying values. To suppress the impact of outliers, a robust linear regression estimation method (RLRM) was employed. The positive effect of RLRM on the simulated dataset was manifested throughout the process of model estimation and use. Although the results obtained by the RLRM were positive, it can be argued that the assumption of constant speed is

not entirely plausible. It might be especially problematic if the model would be implemented on an urban level. (The relationship between travel time and distance is likely to be nonlinear there.) In the future, it is worth investigating other nonlinear models for simulating the non-observed travel times. Alternatively, future work could focus on how much better prediction performance of the model can be if precise activity locations are available, and hence how much effort is justified for acquiring such (privacy-sensitive) data. One step further would be to endogenise even the location of the activities, as people might decide to do the same activity elsewhere depending on the mobility resources that are available (e.g., shopping closer to home if no car is available anymore).

The mixed logit model (MXL) applied to the transport mode choices (the first modelling step) was successfully estimated. This supports the refinement to the original Le Vine's model proposed in the study of Astegiano et al. (2016) that was motivated by the presence of correlations in the error term over repeated choices by the same individual. Surprisingly, using the robust estimation method for travel times for estimating the first step model showed a worse explanation of the variations in people's transport mode choices compared to the models built on the non-robust datasets. The most likely cause of lower values ρ^2 is due to the impact of (less-frequent) long-distance observations, which tend to be less accurately modelled by the RLRM. The model found that walking is the most preferred mode of transport (*ceteris paribus*). This is a reasonable finding, considering the substantial number of short-distance trips reported in the survey and the limited set of factors in this study influencing the transport mode choice.

The estimation process continued with searching for a valid second-step model. Exploring various nested and cross-nested model structures, the only valid model found was the nested logit model (NL) with the nests grouping the alternatives with and without a bicycle. The model was estimated to have all parameters with the expected signs and ρ^2 of 0.69. Inspecting the utility parameters, it turned out that leisure and shop activities are the most decisive travel activities in PAS. Having significant parameters for all the importance and frequency of PAS activities corroborates the proposed utility function. The estimation manifested the positive effect of using the RLRM since all attempts to estimate the model on the non-robust datasets failed to obtain consistent and valid results.

The next step was to investigate how well the valid model could predict market shares of mobility-resource portfolios or "restricted choice sets" (RCS). A detailed analysis was provided using prediction distributions and contingency tables. The model was found to satisfactorily predict the aggregated market shares with only small deviations from the truly observed choices. The NL produced an average relative error in RCS market shares of 1.2%. The combination of ownership of a car and a public transport seasonal ticket was predicted somewhat less accurately. This is probably due to the low number of observations of this combination and because all forms of public transport were aggregated into one generic class. The model's predictive power changes with the used input data. Using the RLRM instead of the non-robust LRM yields a better prediction of truly observed choices.

The model was subsequently tested for sensitivity under various scenarios, investigating whether it is possible to plausibly predict RCS shares when the supply characteristics are modified. Two different changes were tested. The first was a change that affects the daily attractiveness of an option, hence acting directly on the short-time horizon (i.e., travel time of public transport) and through this, indirectly on the long-term horizon. The second was a change of fixed acquisition cost (i.e., the purchase and maintenance costs for bicycles), hence acting directly on the strategic long-term horizon. Both types of changes yielded plausible predictions by the NL model. Less attractive RCS models indeed lost market shares in favour of RCS models that improved because of the change. Moreover, the magnitude of the market shift was reasonably in proportion to that of the input change.

The contribution of this study has been to confirm the modelling concept proposed initially by [1]. Furthermore, this study has contributed to an elaborated utility function that delivers additional insights into ownership decisions on mobility resources. The empirical results show that personal accessibility and frequency of travel activities have significant effects on ownership decisions on mobility resources. The proposed choice model structure overcomes systematic biases (unobserved

factors over the repeated choice and among the nested structures) influencing the modelled transport mode choice decisions. In addition, this research proposed and validated data simulation methods for studies where the travel times of non-chosen alternatives are lacking in the data. The findings of this study are relevant for practice because they allow for a way to relate the development of transport services and policies to the true mobility needs of individuals.

The central topic of this research was to recognize the role of personal accessibility (here represented by the PAS) on ownership by isolating its impact. However, for a realistic analysis, other influencing factors, namely individual socio-economic background or spatial characteristics of households, should be taken into consideration as well. It is worth noting that the modelling framework can be extended to account for such effects. It is recommended to explicitly investigate the associations of personal accessibility and exogenous factors in future studies.

Next, the definition of personal accessibility to PAS travel activity types (e.g., to work, leisure, shopping) that we used in this study can be considered far from exhaustive. In fact, one might find many possible dimensions of PAS relevant for describing the personal accessibility. For example, some early attempts (not presented in this paper) with the model involved the accessibility to short- and long-distance PAS travel activities. The idea was that the short- and long-distance PAS travel activities encompass different characteristics (e.g., short-distance travel activities occurred mostly during working days and on regular basis). Accordingly, the requirements on the mobility resources associated with the respective PAS travel activities are different. So far, our experiment with accessibility to the short- and long-distance PAS travel activities failed because it was not possible to estimate the model with significant utility parameters. However, one might think about other meaningful dimensions such as journey constraints, scheduling and the like.

Another improvement of the PAS definition could be directed in revising the value of the travel time. In its current capacity, the model does implicitly consider congestion effects, albeit to a marginal extent, by considering travel times in the network as part of the utility of travel. Naturally, the utility of selecting a specific mode (and its relevance in the ownership portfolio) will be influenced by the mode's travel times on the network. Choosing a different utility function, which explicitly considers congestion by means of a dedicated parameter, is always possible and will not affect the validity or functionality of the overall approach. Naturally, choosing a behaviourally richer utility function formulation also implies the need for more precise, richer data, which was not available through the course of this study.

Besides the definition of variables included in PAS, the other important issue is the elasticity of PAS. In this study, PAS was assumed to remain constant. However, it is plausible that people update their travel patterns while changing their RCS. For example, a PT service on an influential connection (the one that is perceived to be important) is improved. This may lead some of the users to add a PT seasonal ticket to their RCS. These people may then also utilize PT for other activities in their PAS. Until now, the example depicts the cases treated in this research. However, what if these users start doing more (or less) journeys, or change locations of activities because they adapt to PT accessibility? To the best knowledge of the authors, the understanding of how travel patterns vary with mobility-resource ownership is very limited. It is interesting to consult the empirical findings from the UbiGo project, a pilot Mobility as-a-service (MaaS) project targeting at Gothenburg residents in the city centre. Sochor, Strömberg, and Karlsson [17] found a tendency of people to overestimate their prospective usage of transport modes. The UbiGo users were asked to register their expected utilization of transport for an upcoming month. For example, the users subscribed for 2220 days/month of PT service but utilized 1920 days/month. The difference was even more pronounced for a car-sharing service that was utilized 620 h/month out of subscribed for 904 h/month. It is intriguing to see the difference between the expectations and the actual usage. A better understanding of PAS elasticity helps to improve the predictive power of the mobility-ownership model. More importantly as presented in [17], the operators running innovative mobility services such as MaaS will capitalize on better knowledge about prospective travel patterns because they can tailor their operations to the current and prospective mobility needs of users.

More research is needed to better understand the decision level of the household on using and owning mobility resources. This research simplified the way resources are utilized within households by modelling the related choices on a personal basis. Future extensions of the model should take into account household dynamics and how individual mobility needs are conditional to inner-household interactions. For example, the model should account for efficient utilization of resources within the household, for example, sharing of a vehicle by multiple household members, or the use of one member's resource to escort another. Analysing the role of household members leads to a related research question: what is a decision-maker level appropriate for this type of study? Does it make sense to examine mobility resource markets at person level instead of household level?

Finally, the model estimation and prediction exhibited a dependency on data quality. Therefore, it is suggested to explore more systematically other data simulation techniques applicable when geospatial references are missing and non-chosen alternatives are not observed.

Author Contributions: V.P. contributed by conceiving and designing the study. He also analysed, interpreted the data and wrote the paper; P.A. contributed in the designing of the modelling part and in writing the paper; W.H. contributed with data interpretation; C.T. supervised the research and contributed in data interpretation and in the manuscript evaluation process; M.V. contributed to the review process of the manuscript.

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

List of Notations

U_d^i	Utility of RCS alternative d perceived by individual i
$V_d^{i, \text{non-travel}}$	Non-travel utility component associated with RCS alternative d perceived by individual i
$V_d^{i, \text{travel}}$	Travel utility component associated with RCS alternative d perceived by individual i
ε_d^i	Utility error term for individual i
λ^{travel}	Gumbel-distributed variance of the utility error term
$V_r^{i, \text{non-travel}}$	Non-travel utility associated with mobility resource r perceived by individual i
$V_{m_j}^{i, \text{travel}}$	Travel utility by mode m on travel activity j perceived by individual i
$r = (0, 1, \dots, R)$	Mobility resources
R_d	Set of mobility resources enabled by RCS alternative d
γ_j	Importance of travel activity j belonging to the PAS of individual i
$j_i = (1, \dots, J_i)$	Travel activities included in the PAS of individual i
$m = (1, \dots, M)$	Transport modes
M_d	Set of transport modes enabled by RCS alternative d
$\beta^{\text{tt}*} \sim N(\mu^{\text{tt}}, \sigma^{\text{tt}})$	The travel-time parameter with mean μ^{tt} and the standard deviation σ^{tt}
β^{OPC}	Out-of-pocket cost parameter
$tt_{m_j}^i$	Travel time in [minutes] for mode m on PAS travel activity j by individual i
$\text{OPC}_{m_j}^i$	Out-of-pocket cost in [EUR/journey] for mode m on PAS travel activity j by individual i
ASC_m	Alternative specific constant for transport mode m
β^{FC}	Fixed cost parameter
γ_p	Importance parameter associated with p-type PAS activities
p	PAS travel activity types defined by the trip purpose (to work, leisure, shopping, escort, other)
β_d^N	Travel activity frequency parameter associated with RCS alternative d
FC_d^i	Fixed costs of RCS alternative d in [EUR/week] perceived by individual i
$A_{p,d}^i$	The personal accessibility to p-type PAS activities provided by transport modes enabled by RCS alternative d perceived by individual i
N_d^i	Frequency of travel PAS activities done by individual i having RCS alternative d
$\hat{t}_{m_k}^{i,b}$	Simulated travel time in [min] for unobserved journey k by transport mode n of individual i in region b
$d_{m_j}^{i,b}$	Travel distance in [km] reported by individual i for transport m on journey j in region b
$a_{1,n}^b$	Linear regression parameter for simulated transport mode n and region b
$a_{0,n}^b$	Constant term of the linear regression model for simulated transport mode n and region b

List of Acronyms

CNL	Cross-nested Logit
EMPU	Expected Maximum Perceived Utility
LRM	Linear Regression Model
MAD	Median Absolute Value Deviation
MNL	Multinomial Logit
MOP	German Mobility Panel (household travel survey)
MXL	Mixed Logit
NGEV	Nested Generalized Extreme Value Logit
NL	Nested Logit
PAS	Perceived Activity Set
PT	Public Transport
PTf	Public Transport Season Ticket nest (free access)
PTp	Public Transport Season Single ticket (paid access)
RCS	Restricted Choice Set
RLRM	Robust Linear Regression Model
RMAE	Relative Mean Absolute Error
RMSE	Root Mean Squared Error
SD	Standard Deviation

Appendix A

Table A1. NL with the PT nest and the non-robust simulated dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.042	0.005	0.000
	Travel activity frequency RCS (car)	0 ^f		
	Travel activity frequency RCS (car, bicycle)	−0.113	0.011	0.000
	Travel activity frequency RCS (car, PT)	−0.046	0.023	0.050
	Travel activity frequency RCS (car, bicycle, PT)	−0.146	0.019	0.000
	Importance of ‘escort’ PAS activities	0 ^f		
	Importance of ‘leisure’ PAS activities	158.000	15.000	0.000
	Importance of ‘other’ PAS activities	154.000	21.600	0.000
	Importance of ‘shopping’ PAS activities	114.000	10.800	0.000
	Importance of ‘work’ PAS activities	109.000	20.100	0.000
Model Parameters	Nest ‘PT free’	1 ^f		
	Nest ‘PT paid’	1.000	1.80e + 308	1 *
Model Statistics	Adjusted ρ^2	0.696		
	Final log-likelihood	−560.350		
	Number of observations	1352.000		

* Non Significant at 95%; f . . . fixed value.

Table A2. NL with the bicycle nest and the non-robust simulated dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.009	0.003	0.000
	Travel activity frequency RCS (car)	0 ^f		
	Travel activity frequency RCS (car, bicycle)	−0.021	0.007	0.000
	Travel activity frequency RCS (car, PT)	−0.033	0.013	0.010
	Travel activity frequency RCS (car, bicycle, PT)	−0.095	0.012	0.000
	Importance of ‘escort’ PAS activities	0 ^f		
	Importance of ‘leisure’ PAS activities	51.100	5.760	0.000
	Importance of ‘other’ PAS activities	33.600	6.490	0.000

Table A2. Cont.

		Value	Std Error	p-Value
Model Paramters	Importance of 'shopping' PAS activities	41.500	5.130	0.000
	Importance of 'work' PAS activities	25.700	6.480	0.000
	Nest 'bicycle'	1.000		
	Nest 'no bicycle'	2.820	0.328	0.000
Model Staitistics	Adjusted ρ^2	0.609		
	Final log-likelihood	−564.428		
	Number of observations	1352.000		

f . . . fixed value.

Table A3. NL with the PT nest and the robust simulated dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.052	0.007	0.000
	Travel activity frequency RCS (car)	0 ^f		
	Travel activity frequency RCS (car, bicycle)	−0.156	0.015	0.000
	Travel activity frequency RCS (car, PT)	−0.045	0.028	0.11 *
Model Paramters	Travel activity frequency RCS (car, bicyle, PT)	−0.208	0.027	0.000
	Importance of 'escort' PAS activities	0 ^f		
	Importance of 'leisure' PAS activities	120.000	15.300	0.000
	Importance of 'other' PAS activities	448.000	65.600	0.000
	Importance of 'shopping' PAS activities	116.000	14.500	0.000
	Importance of 'work' PAS activities	223.000	34.500	0.000
	Nest 'PT free'	1 ^f		
	Nest 'PT paid'	1.000	1.80e + 308	1 *
	Adjusted ρ^2	0.794		
	Final log-likelihood	−376.523		
Model Staitistics	Number of observations	1352.000		

* Non Significant at 95%; f . . . fixed value.

Table A4. NGEV using the non-robust simulated dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.006	0.001	0.000
	Travel activity frequency RCS (car)	0 ^f		
	Travel activity frequency RCS (car, bicycle)	−0.071	0.010	0.000
	Travel activity frequency RCS (car, PT)	−0.014	−0.087	0.13 *
Model Paramters	Travel activity frequency RCS (car, bicyle, PT)	−0.087	0.011	0.000
	Importance of 'escort' PAS activities	0 ^f		
	Importance of 'leisure' PAS activities	73.000	10.500	0.000
	Importance of 'other' PAS activities	220.000	46.800	0.000
	Importance of 'shopping' PAS activities	70.800	9.790	0.000
	Importance of 'work' PAS activities	87.400	16.600	0.000
	Nest 'bicycle'	1 ^f		
	Nest 'PT'	10 **	0.000	0.000
	Nest 'PT free'	1 ^f		
	Nest 'PT paid'	1.000	1.80e + 308	1.00 *
	Node RCS (car, bicycle) membership to nest 'PT paid'	1.000	1.80e + 308	1.00 *
	Node RCS (car, bicycle) membership to nest 'Bicycle'	0.000	1.80e + 308	1.00 *

Table A4. Cont.

		Value	Std Error	p-Value
Model Statistics	Node RCS (car, bicycle, PT) membership to nest 'PT free'	1.000	1.80e + 308	1.00 *
	Node RCS (car, bicycle, PT) membership to nest 'Bicycle'	0.000	1.80e + 308	1.00 *
	Adjusted ρ^2	0.820		
	Final log-likelihood	−322.148		
	Number of observations	1352.000		

* Non Significant at 95%; ** Outside or close to the boundaries.

Table A5. NGEV using the robust simulated dataset.

		Value	Std Error	p-Value
Utility Parameters	Fixed cost [EUR/week]	−0.012	0.003	0.000
	Travel activity frequency RCS (car)	0 ^f		
Model Parameters	Travel activity frequency RCS (car, bicycle)	−0.099	0.012	0.000
	Travel activity frequency RCS (car, PT)	−0.010	0.018	0.59 *
	Travel activity frequency RCS (car, bicycle, PT)	−0.106	0.015	0.000
	Importance of 'escort' PAS activities	0 ^f		
	Importance of 'leisure' PAS activities	69.400	10.800	0.000
	Importance of 'other' PAS activities	159.000	39.500	0.000
	Importance of 'shopping' PAS activities	81.900	11.100	0.000
	Importance of 'work' PAS activities	91.400	17.200	0.000
	Nest 'bicycle'	1 ^f		
	Nest 'PT'	10 **	0.000	0.000
	Nest 'PT free'	1 ^f		
	Nest 'PT paid'	1.000	0.000	1 *
	Node RCS (car, bicycle) membership to nest 'PT paid'	1.000	0.000	0.000
	Node RCS (car, bicycle) membership to nest 'Bicycle'	0.0001 **	0.000	0.000
	Node RCS (car, bicycle, PT) membership to nest 'PT free'	1.000	1.80e + 308	1 *
	Node RCS (car, bicycle, PT) membership to nest 'Bicycle'	0.000	1.80e + 308	1 *
Model Statistics	Adjusted ρ^2	0.870		
	Final log-likelihood	−230.729		
	Number of observations	1352.000		

* Non Significant at 95%; ** Outside or close to the boundaries; f . . . fixed value.

References

1. Le Vine, S.; Lee-Gosselin, M.; Sivakumar, A.; Polak, J. A New Concept of Accessibility to Personal Activities: Development of Theory and Application to an Empirical Study of Mobility Resource Holdings. *J. Transp. Geogr.* **2013**, *31*, 1–10. [CrossRef]
2. De Jong, G.; Fox, J.; Daly, A.; Pieters, M.; Smit, R. Comparison of Car Ownership Models. *Transp. Rev.* **2004**, *24*, 379–408. [CrossRef]
3. Dieleman, F.M.; Dijst, M.; Burghouwt, G. Urban Form and Travel Behaviour: Micro-Level Household Attributes and Residential Context. *Urban Stud.* **2002**, *39*, 507–527. [CrossRef]
4. Fagnant, D.J.; Kockelman, K.M.; Bansal, P. Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2536*, 98–106. [CrossRef]
5. Yang, J.; Liu, A.A.; Qin, P.; Linn, J. The Effect of Owning a Car on Travel Behaviour: Evidence from the Beijing License Plate Lottery. 2016. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2789824 (accessed on 21 March 2018).

6. Mulalic, I.; Pilegaard, N.; Rouwendal, J. Does Improving Public Transport Decrease Car Ownership? Evidence from the Copenhagen Metropolitan Area. 2016. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2710547 (accessed on 21 March 2018).
7. Simma, A.; Axhausen, K.W. Structures of Commitment in Mode Use: A Comparison of Switzerland, Germany and Great Britain. *Transp. Policy* **2001**, *8*, 279–288. [[CrossRef](#)]
8. Nicola, C.; Anne, N. Car Ownership and Mode of Transport to Work in Ireland. *Econ. Soc. Rev.* **2010**, *41*, 43–75.
9. Paola, A.; Deniz, A.; Willem, H.; Chris, T.; Martina, V. Quantifying the Explanatory Power of Mobility-Related Attributes in Explaining Vehicle Ownership Decisions. *Res. Transp. Econ.* **2016**. [[CrossRef](#)]
10. Le Vine, S. Strategies for Personal Mobility: A Study of Consumer Acceptance of Subscription Drive-It-Yourself Car Services. Ph.D. Thesis, Imperial College London, London, UK, 2011.
11. Andrew, D.; Michel, B. A General and Operational Representation of Generalised Extreme Value Models. *Transp. Res. Part B Methodol.* **2006**, *40*, 285–305. [[CrossRef](#)]
12. Kelpin, R. Mobility Panel Germany. Available online: <http://daten.clearingstelle-verkehr.de/192/> (accessed on 19 March 2018).
13. Washington, S.; Ravulaparthi, S.; Rose, J.M.; Hensher, D.; Pendyala, R. Bayesian Imputation of Non-Chosen Attribute Values in Revealed Preference Surveys. *J. Adv. Transp.* **2014**, *48*, 48–65. [[CrossRef](#)]
14. Verboven, S.; Hubert, M. LIBRA: A MATLAB Library for Robust Analysis. *Chemom. Intell. Lab. Syst.* **2005**, *75*, 127–136. [[CrossRef](#)]
15. Maronna Ricardo, A.; Martin, R.D.; Yohai, V. *Robust Statistics*; John Wiley & Sons, Ltd.: Chichester, UK, 2006. [[CrossRef](#)]
16. Moshe, B.-A.; Lerman, S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand*; The MIT Press: Cambridge, MA, USA, 1985.
17. Sochor, J.L.; Strömberg, H.; Karlsson, M.A. An Innovative Mobility Service to Facilitate Changes in Travel Behaviour and Mode Choice. In Proceedings of the 22nd World Congress on Intelligent Transportation Systems, Bordeaux, France, 5–9 October 2015.



© 2018 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/>).