11th International Conference on Transport Survey Methods

Vehicle cost imputation in travel surveys: Gaining insight into the fundamentals of (auto-) mobility choices

Christine Eisenmann[a], Tobias Kuhnimhof [b]*

[a] *Karlsruhe Institute of Technology (KIT), Institute for Transport Studies, Kaiserstr. 12, 76131 Karlsruhe, Germany*
[b] *Institute of Transport Research, German Aerospace Center, Rutherfordstr. 2, 12489 Berlin*

Abstract

Costs of cars are among the most relevant factors influencing travel behavior. However, there is a lack of data about the true costs of car ownership and how these costs are distributed across different vehicles and across the population. This paper presents a multistage method for imputing car costs by cost item in a German national travel survey data set. Based on vehicle information reported by survey participants, we assign costs to each of the three thousand cars in the data set using the most comprehensive German vehicle cost data base. In addition to combining different data sets, we use model based imputation methods. In order to validate the average costs for private vehicles we also analyze the German income and expenditure survey EVS. The average total cost of ownership for a private car in Germany is about 315 Euros per month. This translates to about 31 Eurocents per auto-km. About one third of the costs are fuel, another third is depreciation, and the rest are other mainly fixed costs (insurance, tax, repair and maintenance). However, the cost distribution is strongly skewed with a long tail to the right. Hence, the majority of motorists pay less than average for their private vehicles while few pay more and evidently some pay a lot more. This imputation approach delivers unprecedented vehicle cost information in particular with regard to the distribution of vehicle costs. Such data is a key for understanding the fundamentals of mobility choices.

Keywords: car ownership, imputation, vehicle cost, cost of car ownership, expenditure, German Mobility Panel, Income and Expenditure Survey

1. Introduction

The cost of holding and using cars is one of the most influential factors that drive long and short term mobility choices. Despite the fact that this truism is widely acknowledged there is surprisingly little knowledge about the costs that drivers incur in reality. This paper presents an approach to impute car costs in travel surveys by assigning costs per item based on detailed vehicle information and a vehicle cost data base. This imputation procedure allows closing a fundamental information gap with regard to understanding mobility choices. The study uses 2016 German Mobility Panel (MOP) data and is a sequel to an earlier vehicle cost imputation study using 2005 MOP data (Kuhnimhof, Ottmann, & Zumkeller, 2008).

In addition to describing the imputation procedure, the paper presents the results of this cost imputation. We also compare the results of this imputation with vehicle expenditure data from an income and expenditure survey. Based on this comparison we discuss advantages and shortcomings of our imputation procedure as well as next steps which we envision to improve the imputation methodology.

2. Data

We used four data sets for compiling the results presented in this paper. Three data sets (i. a fuel consumption survey linked to a national household travel survey; ii. a vehicle cost data base from a German car club; iii. a data set with detailed information on the German vehicle stock) were combined in the actual imputation procedure. An important identifier to combine vehicle information across the three data sets is the HSN-TSN number. Each car configuration (i.e., make, model, version, series) registered in Germany can be identified by an HSN-TSN number (Kraftfahrtbundesamt, 2017a). This number is a combination of a four-digit manufacturer number (HSN, "Hersteller-Schlüssel-Nummer") and a three-digit type code number (TSN, "Typ-Schlüssel-Nummer"). In essence, within each HSN-TSN category vehicles can only differ by year of construction and special features (e.g., trailer hitch, sunroof, color).

In addition to these three data sets, we use a fourth data set (iv. German income and expenditure survey data set) as an independent source for car related expenditures. This section presents all four data sets used for this study. Table 1 gives an

* Corresponding author. Tel.: +49-30-670 55 168.
E-mail address: Tobias.Kuhnimhof@dlr.de

overview of vehicle cost categories included in the data bases ii and iv. The table also shows the cost categories considered in this study, and how we combined these into common denominator categories.

Table 1: Vehicle cost categories in the ADAC vehicle cost data base and the EVS as well as common denominators used in this study

	ADAC data base cost categories	Common denominators used in this study	EVS cost categories
Cost categories considered in this study	fuel oil adblue	fuel & lubricants	fuel & lubricants
	depreciation	depreciation	expenditures for buying and leasing vehicles (minus) Income from selling vehicles
	insurance	insurance	insurance
	repair, parts, maintenance	repair and maintenance	maintenance parts and accessories
	tax	tax	tax
Cost categories not considered in this study	washing		
			garage rental
			other expenses (e.g., parking fees, tickets)

2.1. The German Mobility Panel (MOP)

The MOP is a German national household travel survey that has been conducted every year since 1994 (Weiß, Chlond, Behren, Hilgert, & Vortisch, 2016). The survey is carried out on behalf of and funded by the German Federal Ministry of Transport and Digital Infrastructure. The market research firm KANTAR TNS is responsible for the field work (i.e., recruitment and data collection) and the Institute for Transport Studies of the Karlsruhe Institute of Technology is in charge of the design and scientific supervision of the survey. The MOP consists of two parts: a) a one-week travel diary (everyday mobility survey; MOP-EM) in fall with an annual sample size of about 1.500 households; b) a two-month fuel consumption and odometer reading survey (MOP-FCOR) in spring in which car-owning households of the MOP-EM participate. MOP-FCOR comprises an annual sample size of about 1.500 cars. The sample is weighted by car properties (car age, cylinder capacity) in order to ensure the representativeness of the vehicle sample.

For this study we use the MOP-FCOR data set. MOP-FCOR participants are asked to report dates, odometer mileage, and the amount of fuel purchased for each refueling event during eight weeks in spring. This information allows for calculating average fuel consumption and monthly mileage for every car in the sample. Information about the socio-demographics of the participants, the availability of cars and bicycles in the household and vehicle details are also collected. The vehicle details include parameters such as make, model, fuel type, engine size, and year of construction. In the MOP-FCOR, data are collected through a paper-and-pencil questionnaire (PAPI). We imputed car cost information for cars of the 2015 and 2016 MOP-FCOR survey. The total sample size of cars for which costs were imputed was 2,977.

2.2. ADAC vehicle cost data base

Vehicle cost data come from a German car club (Allgemeiner Deutscher Automobil-Club ADAC). This car club maintains a large car cost data base. The main purpose of this data base is to provide individual information to ADAC customers (ADAC, 2017): when buying or selling a car, one can look up and compare car prices and running costs on the ADAC website based on detailed vehicle specifications (make, model, type of fuel, year of construction etc.). Moreover, subsets of the data set can be purchased from ADAC with prices of the data depending on the use.

The underlying identifiers in the ADAC vehicle cost data base are HSN-TSN numbers (see above), meaning the data base contains vehicle costs for each HSN-TSN/year-of-construction combination. Vehicle costs are differentiated by cost items as follows: new car price in the year of construction, used car price in 2016, repair and maintenance in 2016, fuel in 2016, oil in

2016, AdBlue (diesel exhaust fluid) in 2016, car wash in 2016, tax in 2016, and insurance in 2016 (fully comprehensive cover, partly comprehensive cover, liability).

Generally, the ADAC data base is the most comprehensive source for car costs in Germany. However, the data base also has several shortcomings:

1) Car cost and residual value information of cars older than 12 years is not included in the ADAC data base.
2) Repair and maintenance costs base on rates of authorized car garages (e.g., Mercedes car garage); in reality, however, many motorists prefer independent car garages with less expensive rates.
3) On the other hand, larger unforeseeable repairs, such as damages to the bodywork, are not included in the ADAC repair and maintenance costs.
4) On average, car insurance costs in the ADAC data base are overestimated. The reason is that insurance costs in the data base do not take account of the bonus-malus scheme ("Schadensfreiheitsklasse") in the German car insurance system: In this bonus-malus system the individual insurance premium depends on how long the policy holder has driven without insurance claim. As a consequence, individual insurance premiums range between 30% and 135% of the initial insurance premium (Autobild, 2016). In the dataset, bonus-malus is set as 100%, although many car users in Germany pay considerably less for their car insurance. Even the ADAC experts themselves confirmed in personal communication that in reality insurance holders on average probably pay only about 30% of what the vehicle cost data base suggests.
5) The residual value is based on assumed average odometer reading values only and not on real odometer mileages per vehicle.
6) The ADAC assumes washing costs of 21€ per car and month. We believe this figure is much exaggerated as even flat rates at car washes in Germany are available for 20€ per month (see for example (STAYCLEAN, 2017)). Due to these unrealistic assumptions and due to the fact that this cost category is not available in the EVS we have not included washing costs in our study.

In the context of the vehicle cost imputation presented here, we purchased vehicle costs broken down by cost item from ADAC for 14,999 HSN-TSN/ year-of-construction combinations. However, we did not use all of these observations as will be explained later on.

2.3. German vehicle stock data base

The German Federal Motor Transport Authority (KBA, "Kraftfahrtbundesamt") keeps a data base containing the vehicles in use (i.e., with a valid registration/number plate) in Germany, the central vehicle register (ZFZR, "Zentrales Fahrzeugregister") (Kraftfahrtbundesamt, 2017b). As of January 1, 2016 this data base contained about 63 million individual entries, i.e., all in-use vehicles including two-wheelers and trailers in Germany with vehicle and owner details. This data base as such is only available to KBA and not for research or other purposes. However, very detailed aggregate statistics based on this data base can be obtained from KBA.

For this research project we were able to use a data base that provided a complete overview of the German vehicle stock as of January 1, 2016 broken down by HSN-TSN/ year-of-construction-combination. The observations in this data base are the individual HSN-TSN/ year-of-construction-combinations; the variables in this data base are additional vehicle details such as the vehicle trading name (e.g., "Volkswagen Golf"), engine size/displacement, horsepower, etc. as well as the number of vehicles in the respective category registered in Germany.

This data set was used for two purposes in the context of our study: a) to associate eligible HSN-TSN-numbers with vehicles from the MOP-FCOR based on reported vehicle details; b) to identify common HSN-TSN/ year-of-construction-combinations on German roads as will be explained later.

2.4. Income and expenditure survey (EVS)

In addition, we analyzed car related expenditures as reported in the German income and expenditure survey 2013 (EVS, "Einkommens- und Verbrauchsstichprobe") (Destatis, 2017). Following a common income and expenditure survey format, the EVS asks respondent households (sample size 42,792 households) to report all incomes and expenditures during a three month reporting period broken down by very detailed categories. This includes various categories relating to transportation and vehicles.

Due to the three-month reporting period survey design, such expenditure surveys are unable to deliver sensible car expenditure distributions. This is because very few households have extremely high car related expenditures, namely those that purchased a car during the reporting period. However, most households have no or relatively low car related expenditures, e.g., due to regular vehicle fueling during the reporting period. Regular annual (e.g. tax and insurance) or irregular and unforeseeable expenses (e.g. repairs) which fall into the reporting period at random add to that problem. However, expenditure surveys are a useful source to compute average values for car related expenditures.

Hence, to assess the validity of the results of the car cost imputation we compare average vehicle costs as obtained by the imputation with average vehicle costs as surveyed in the EVS 2013. For the purpose of this study, we did not rely on published EVS reports but analyzed the EVS microdata set in order to ensure a best possible match of the car cost categories with the cost items in our imputation procedure. We computed the total amount of expenditure by car cost item for private households and divided this by the number of cars in the households. Income generated from selling cars was subtracted from expenditure spent for buying and leasing cars to make this cost category comparable to the depreciation as obtained from the imputation procedure. In this analysis, both expenditures and cars only relate to private vehicles. We did not include company cars for which users usually do not incur any visible costs (as reported in the EVS) but a deduction from their net-income. Table 3 shows the result from this descriptive EVS-analysis.

3. Cost imputation methodology

Our cost imputation procedure fell into two parts: Firstly, there was an initial cost imputation which was purely based on combining data from different data sources, most importantly from our car use survey (MOP-FCOR) and the ADAC vehicle cost data base. This, however, left many cases with missing cost item data, mainly concerning the new car price and the residual value of the car for which the ADAC data base contained no entries in many cases. Moreover, after this initial cost imputation the data set did not contain information on annual or monthly depreciation and many residual vehicle values were not correct as will be explained below. For this reason, we secondly estimated and applied a multivariate model to close these existing data gaps. This section presents these different stages of our vehicle cost imputation in greater detail.

3.1. Initial cost imputation

The initial cost imputation comprised three steps involving the MOP-FCOR data set, the ADAC vehicle cost data base and the German vehicle stock data base:

- First, we identified suitable HSN-TSN-aliases for each car in the 2015/2016 MOP-FCOR data set: In order to limit the respondent burden, MOP participants do not report HSN-TSN numbers of their vehicles but vehicle details which they usually know out of the top of their head such as make, model trading name, type of fuel, engine size, year of construction and horse power. Based on these variables and using the German vehicle stock data base, we associated all HSN-TSN-numbers that were suitable for each vehicle in our MOP-FCOR data set. On average, we found five such numbers (“aliases”) for each MOP-FCOR car, i.e. our MOP-FCAR car data set increased from about 2,977 cars to about 14,999 HSN-TSN/ year-of-construction-combinations.
- Second, car cost information from the ADAC vehicle cost data base was added to each of the 14,999 HSN-TSN/ year-of-construction-combinations. This step was performed by ADAC through combining our data set with their vehicle cost data set using the appropriate identifiers. (However, for a relatively large number of HSN-TSN/year-of-construction-combinations there was no match in the vehicle cost data base resulting in a large number of missing values).
- Third, we reduced our data set back to 2,977 observations by identifying one HSN-TSN-alias for each MOP-FCOR car. Therefore, we only considered aliases with available car cost information. If that information was available for more than one HSN-TSN-alias, we selected the alias with the highest number of vehicles on the road among all suitable aliases. Therefore, we again used the German vehicle stock data base.

However, after this initial cost imputation there were still a large number of cars in the sample with missing car cost information. New and used car price values were missing for 28% of the sample and tax/insurance/maintenance information for 2%-4% of the sample. Specifically, the missing vehicle price value gave rise to the second step in our imputation procedure as described in the next two sections.

3.2. Modelling new car prices, residual values and depreciation

As a next step, we estimated and applied two linear regression models in order to predict the new car price (in €) and the residual value (percentage of the new car price) of the vehicles in our data set. There were three reasons behind these regressions:

- first, closing the existing cost data gaps with regard to vehicle value by imputing missing values;
- second, correcting the residual vehicle values resulting from the initial cost imputation, which are based on average odometer reading values only;
- third, translating residual vehicle values into annual or monthly depreciation costs as a function of increasing age and mileage of the vehicles.

Table 2 shows the results of both multivariate regression models. In both cases, we used simple linear regressions with the new car price and the residual value being the explained variables. Explanatory variables were car drive, motor power, car brand (premium/non-premium, country of manufacture), cylinder capacity, car segment as well as total car mileage and car age (the latter two account for the residual value model only). Only significant variables were incorporated in the model. We broke the explanatory variables cylinder capacity and motor power down into categories and implemented them as dummy variables in the model. We also wanted our model to consider the fact that new cars depreciate faster than old cars. Therefore, we employed multiple linear functions to approximate the regressive relationship between the residual car value and car age as well as car mileage (see also (Hughes, Liu, & Castro, 2015)). From the 2,977 observations in our data set after the initial cost imputation, we used only observations with complete new car price and residual value information, which reduced the data set to 2,000 cars.

While both models show a good fit with R-Squares of 0.84 and 0.96 respectively, we are aware that these models have shortcomings specifically relating to using a linear regression. First, linear regressions do not prevent predicted values to go below zero, which does not make sense and is a specific concern in the case of the residual value model. Moreover, the residual value model is not well suited to predict residual values of vintage cars and other rather old cars, since the ADAC vehicle cost data do not incorporate residual values of cars older than 12 years. Therefore, the fact that residual values for vintage cars tend to rise in Germany with increasing car age – depending on the cars' state of maintenance, of course – is not properly reflected in the residual value model (VDA, 2017). Second, both – new car prices and residual values – are not normally distributed and linearizing the explained variable would have made sense. However, we also tried a log-linear model. While the model fit improved, the problems of these models arise when re-transforming the predicted values to real values, i.e. prices and percentages. The assumptions about the error term in the linear regression lead to over-estimation of the car prices when re-transforming correctly. However, deriving plausible absolute figures for the new car prices and the residual values of the vehicles was paramount when applying the model in our context. For this reason, we opted for the methodologically deficient but robust linear regression models. However, we acknowledge that there is room for improvement regarding these models.

As indicated above, the application of these linear models firstly served to impute missing vehicle price and value information for about a third of our data set, which does not need further explanation. Beyond that, we applied the residual value model to all MOP-FCOR cars in order to correct the residual vehicle value after the initial cost imputation: Residual values of used cars depend on individual car mileages (see Table 2). However, the ADAC data base only assumes average annual mileages. This might differ substantially from the individual vehicle mileage of the cars in MOP-FCOR data set. Hence, residual values of all vehicles in our MOP-FCOR data set after the initial cost imputation needed to be corrected based on the actual individual vehicle mileages. The application of the linear regression model on used car prices allowed for this correction of the residual value.

Finally, we applied the residual value model in order to derive annual or monthly vehicle depreciation. In order to do so, the residual values of a car at two points in time are needed and depreciation can be calculated as the difference between the two. Therefore, we compared the residual value of each vehicle in 2016 (which is available in the dataset) with a predicted value in 2017 (one year foresight). In order to generate the predicted residual value in 2017 we applied the linear model by advancing the vehicle age by one and predicting the vehicle mileage in 2017. The 2017 odometer reading value was predicted by adding 12 times the car's monthly mileage as reported in the MOP-FCOR survey to the 2016 vehicle mileage. In order to correct for the shortcomings of the residual car model for vintage cars, we set the depreciation to be zero for cars with a negative residual value.

Table 2: Estimation results (and the corresponding levels of significance) of linear regression models in new car prices and residual car values.

Variable	**New car price [€]**		**Residual car value [% of new car price]**	
	Parameter Estimate	**Pr > |t|**	**Parameter Estimate**	**Pr > |t|**
Intercept	11,171	<.0001	0.655	<.0001
Car drive: diesel	2,547	<.0001	0.007	0.0013
Car drive: hybrid, electric, gas	.	.	0.013	0.0061
Motor power: 75-99 PS	1,507	0.0002	0.011	<.0001
Motor power: 100-124 PS	4,176	<.0001	0.012	<.0001
Motor power: 125-149 PS	6,456	<.0001	0.019	<.0001
Motor power: 150-199 PS	8,715	<.0001	0.025	<.0001
Motor power: 200 PS and more	19,753	<.0001	0.036	<.0001
Car brand: premium car manufacturer	2,318	<.0001	0.009	<.0001
Car brand: German	1,855	<.0001	0.025	<.0001
Car brand: French	.	.	-0.011	<.0001
Car brand: Japanese	.	.	0.015	<.0001
Cylinder capacity: 1.400-1.599 ccm	556	0.0733	-0.008	0.0001
Cylinder capacity: 1.600-1.999 ccm	1,321	0.0002	-0.006	0.0104
Cylinder capacity: 2.000 ccm and more	3,668	<.0001	-0.007	0.0261
Segment: small	1,506	0.0009	0.020	<.0001
Segment: compact	3,431	<.0001	0.029	<.0001
Segment: middle class	7,213	<.0001	.	.
Segment: upper middle class	13,131	<.0001	.	.
Segment: upper class	25,646	<.0001	-0.022	0.0593
Segment: cross-country	10,968	<.0001	0.055	<.0001
Segment: sport	12,690	<.0001	0.047	<.0001
Segment: mini van	3,431	<.0001	0.017	<.0001
Segment: large van	7,113	<.0001	0.010	0.0004
Segment: utility	6,988	<.0001	0.014	0.001
Segment: motorhome	20,281	<.0001	0.077	<.0001
Segment: SUV	6,098	<.0001	0.044	<.0001
Total car mileage [10.000 km]	.	.	-0.030	<.0001
Total car mileage over 50.000 km (i.e., max (0; tot. mileage – 50.000 km) [10.000 km]	.	.	0.022	<.0001
Total car mileage over 150.000 km (i.e., max (0; tot. mileage – 150.000 km) [10.000 km]	.	.	0.006	<.0001
Car age [years]	.	.	-0.031	<.0001
Car age, after 10 years (i.e., max(0, car age-10) [years]		.	0.019	<.0001
Sample size	*2,000*		*2,000*	
R-Square	*0.8363*		*0.9642*	
Adjusted R-Square	*0.8344*		*0.9636*	

3.3. Fuel and insurance costs

As a final step in our cost imputation, we corrected fuel expenditures per vehicle and narrowed down insurance costs. The fuel expenditures as resulting from the initial cost imputation (i.e., according to the ADAC vehicle cost data base) are based on assumed average annual mileages, ADAC test cycle fuel consumption and assumed average prices for fuel. However, from the MOP-FCOR which collects fuel consumption along with vehicle mileage we had more accurate information available per vehicle for these data items. Hence, in the final data set with imputed expenditures we did not use the ADAC fuel cost information. Instead, our vehicle cost data set contains fuel costs based on each car's average fuel consumption (liter per 100 km), the annual mileage (monthly mileage during the reporting period times 12 for every month of the year) as reported in MOP-FCOR and average fuel prices in 2016, differentiated by fuel type (ARAL, 2017).

As for vehicle insurance, the ADAC vehicle cost data base provided three different data items per vehicle: costs for i. liability insurance, ii. partially comprehensive insurance, iii. fully comprehensive insurance. Obviously, to an individual vehicle only one of these insurance schemes apply at a time and we needed to identify a likely insurance scheme per vehicle. Car owners in Germany are obliged to take out liability insurance for their car. They can also take out additional fully comprehensive or partly comprehensive covers if they wish for greater insurance protection, but the latter two are not obliged by law (Verbraucherzentrale, 2016). However, a higher insurance cover is advisable for cars with high values, e.g., new cars. Old cars with low residual values often only have liability insurance. 25% of registered vehicles in Germany only have liability insurance, 30% have an additional partly comprehensive cover and 45% have a fully comprehensive cover (Statista, 2015). Therefore, we assumed that all cars aged 4 years and younger have fully comprehensive cover, cars aged 5 to 8 years have partly comprehensive cover and cars older than 8 years have liability insurance only. Based on these assumptions we selected the most likely insurance costs per vehicle.

4. Results and Discussion

Table 3 shows weighted averages, standard deviations and extreme values for various car cost items for all cars, and separating privately registered cars and commercially registered cars (i.e., company cars) as resulting from our cost imputation. In addition, the table lists corresponding average expenditures per car as measured by the EVS. As private households usually do not incur costs (aside from net-income reductions) for company cars, the corresponding EVS values only relate to private vehicles. As explained above, only average values can be compared across the different data sets.

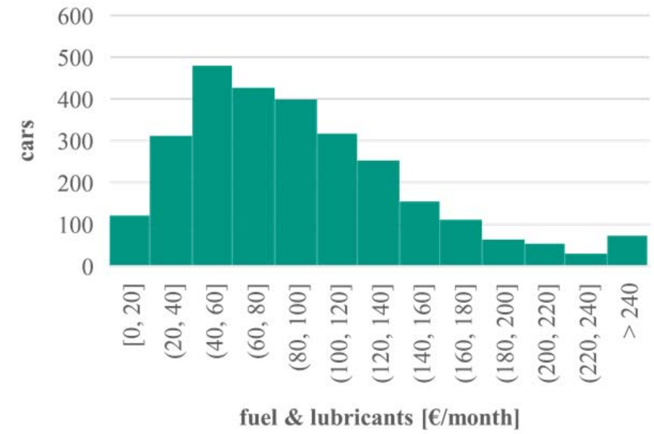
Given the substantial differences in the two approaches (EVS vs. MOP-FCOR microdata with imputed cost) and the time lag between the data sources (2013 vs. 2015/2016) we believe that the consistency of the most results is absolutely satisfactory. Specifically with regard to expenditures for fuel, depreciation and tax the results are surprisingly consistent. Insurance costs are the big exception with costs per month according to the imputed data being about 65€ higher than according to the EVS. The reasons for this are stated above. It appears very likely that average insurance expenditures according to EVS are closer to reality than those resulting from the cost imputation procedure.

However, the real advantage of the MOP-FCOR data set with imputed cost is its ability to provide car cost distributions, which the EVS cannot provide. Figure 1 shows such distributions. It is evident and makes sense that these distributions are generally skewed with a long tail to the right. This means that the median is lower than the average, meaning that the majority of vehicles cost substantially less than the average. This conforms to expectation; however, it is important to keep this in mind when interpreting average values in this context.

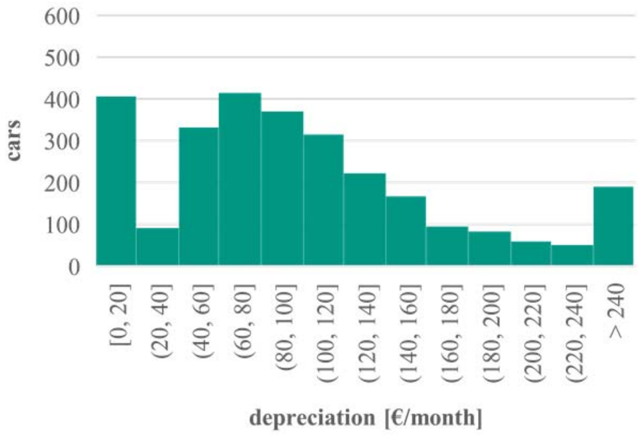
In the following we will discuss selected cost issues and the associated data and imputation problems by comparing the EVS and imputation data findings and drawing insights from the presented distributions. This discussion mainly focusses on private vehicles because of the small sample size of the commercial vehicles and the comparability with EVS results. Moreover, some of the drawbacks of the cost imputation procedure that mainly affect old cars are not a concern for company cars which are almost exclusively relatively new cars.

Table 3: car costs for the imputed MOP-FCOR sample (mean, StdDev, minimum, maximum) and the EVS 2013

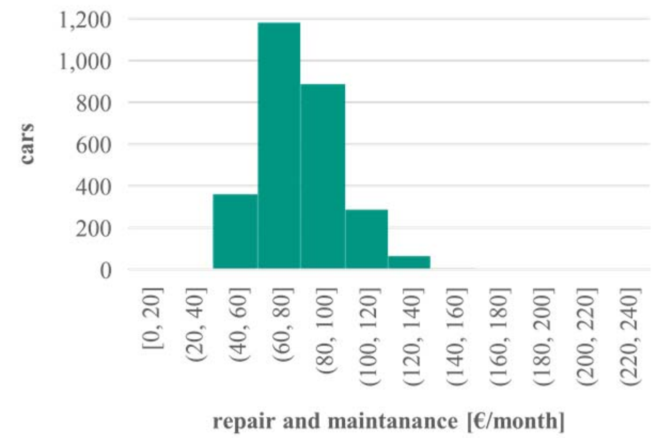
	Imputed data				**EVS 2013**
	Mean	**StdDev**	**Minimum**	**Maximum**	**Mean**
All cars					
Fuel & lubricants[€/month]	97.2	66.1	0.0	655.0	-
Depreciation [€/month]	113.9	128.6	0.0	1,573.0	-
Insurance [€/month]	105.6	43.5	28.0	534.0	-
Repair and maintenance [€/month]	80.4	18.5	46.0	241.0	-
Tax [€/month]	13.2	8.5	2.0	61.0	-
Total costs [€/month]	410.2	191.5	140.0	2,268.0	-
Sample size	*2,795 cars*				
Private cars					
Fuel & lubricants [€/month]	92.0	58.6	0.0	553.0	101.8
Depreciation [€/month]	105.6	109.9	0.0	1,573.0	100.4
Insurance [€/month]	102.7	40.9	28.0	360.0	35.8
Repair and maintenance [€/month]	80.0	18.4	46.0	224.0	54.5
Tax [€/month]	12.9	8.5	2.0	61.0	11.6
Total costs [€/month]	393.1	163.2	140.0	2,268.0	304.2
Sample size	*2,546 cars*				*35,673 households with cars, 49,578 cars*
Company and business cars					
Fuel & lubricants [€/month]	182.4	103.4	19.0	655.0	-
Depreciation [€/month]	242.1	235.4	0.0	1,482.0	-
Insurance [€/month]	151.9	52.8	56.0	534.0	-
Repair and maintenance [€/month]	88.2	18.9	51.0	241.0	-
Tax [€/month]	18.0	8.0	2.0	55.0	-
Total costs [€/month]	682.6	311.1	230.0	2,110.0	-
Sample size	*202 cars*				



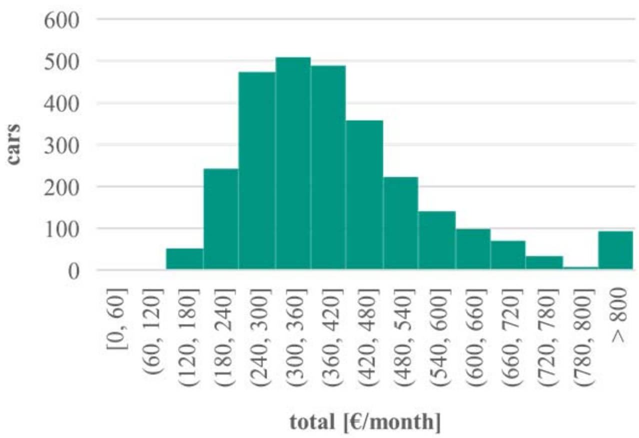
a) fuel & lubricants



b) Depreciation



c) Repair and maintenance



d) Total costs per month

Figure 1: Distributions of car costs per item based on the MOP-FCOR data set with imputed costs (private cars)

4.1. Vehicle depreciation

According to EVS, German households on average spend 100.4 Euros per month for buying and leasing vehicles per vehicle after accounting for income generated through selling cars. By and large, these expenditures should reflect the average vehicle depreciation per car and month. We estimated similar average depreciation figures: according to our model, the depreciation of private cars is 105.6 Euros per month.

One of the problems of the ADAC vehicle cost data base is that it did not contain residual value figures for vehicles older than 12 years. The fact that residual values of vintage cars might increase or does at least decrease marginally only is not appropriately taken into account in the linear regression model. Our correction (i.e., set depreciation to be zero for cars with a negative residual value) is addressing this issue only partly. However, in light of this issue the consistency of the EVS findings and the cost imputation finding are satisfactory.

4.2. Repair and Maintenance

Because the ADAC vehicle cost data base does not contain repair costs for irregular and unforeseeable damages to the car, we expected that the cost imputation would underestimate expenditures for repair and maintenance. We expected that this would predominantly affect old cars. Old cars usually don’t have comprehensive insurance that takes care of specific types of damage to the car (e.g., damage caused by hailstorm or animal bites) and also the failure rate of costly single vehicle components (e.g., lambda sensor, alternator) increases with vehicle age.

However, repair and maintenance costs per car per month from EVS (55 Euros) and the cost imputation data (80 Euros, private cars) were contrary to our expectation. We believe the reason for this is that in reality motorists find numerous ways to get away with lower expenditures per repair or vehicle service than what ADAC assumes. While ADAC assumes cost rates from authorized garages, motorists frequently prefer unauthorized garages, take care of the damage themselves or even go without repair. This appears to over-compensate for the bias that our cost imputation has on the side of larger repairs which we assume to be relevant mainly for old cars.

These are likely explanations for the differences between the EVS and the cost imputation values. In addition, the distribution of the repair and maintenance cost (Figure 1, private cars) does not raise any suspicion. Again, we believe that the order of magnitude for average expenditures for repair and maintenance per private car per month is about right in both data sets and ranges from about 50 Euros to 80 Euros.

4.3. Key findings concerning private vehicle TCO

Table 4 compiles estimated likely figures for expenditures for private vehicles broken down by cost item. These estimated figures are based on the considerations above. For insurance we chose the EVS value because we believe it is closer to reality. In the other cases we chose the mean between the EVS and imputed data value. Table 4 also shows the resulting average cost per km based on 12,333 km annually for private cars as measure by the most recent German mileage survey in 2014 (Bäumer et al., 2017).

On average, holding a private car in Germany costs about 315 Euros per month resulting in about 31 Eurocents per km. About one third of the cost of private cars is fuel, one third is depreciation and one third is made up by other – mostly fixed – costs. However, given the skew of the distribution (see Figure 1) most motorists actually pay less for their cars.

Table 4: Estimated likely figures for vehicle expenditures per cost item for private vehicles

	Imputed data	EVS	Estimated Figure	Comment
Fuel & lubricants[€/month]	92.0	101.8	97	Rounded mean
Depreciation [€/month]	105.6	100.4	103	Rounded mean
Insurance [€/month]	(102.6)	35.8	36	EVS value
Repair and maintenance [€/month]	80.0	54.5	67	Rounded mean
Tax [€/month]	12.9	11.6	12	Rounded mean
Total costs per month [€]			315	
Total costs per km, mileage weighted [€]			0.31	

4.4. Key findings concerning company car TCO

The results from our cost imputation concerning company cars, i.e., commercially registered vehicles in use by private households, must be interpreted with great care (see Table 3). Firstly, the sample size (202 cars) is very small. Secondly, comparison with EVS figures is not possible; hence, there is no external data source to check the validity of these results.

Nevertheless, the findings on company cars in Table 3 appear consistent with expectation and other data. Fuel expenditure per company car is about twice the fuel expenditure per private car, conforming to the average monthly mileage of commercially registered cars of about 2040 km (twice that of private cars) (Bäumer et al., 2017). Depreciation is 2.3 times as high as for private cars. This is logical as company cars are often premium cars and almost exclusively new cars that are subject to high depreciation. Again, we assume that insurance costs of company cars are overestimated; however, it is unclear to which degree. The higher repair and maintenance cost of company cars of about 90 Euros appear reasonable. This is because company cars are often expensive cars with higher repair and maintenance rates. In addition, service and repair of company cars is usually through authorized garages and dealerships. Average tax rates for company cars are higher as these are usually cars with larger engines and higher CO_2-emissions, on which tax rates are based.

In light of the uncertainties about company car insurance premiums, the monthly cost per company car in Germany appears to range from about 600 to 700 Euros. Hence, in total the average company car is about twice as expensive as the average private car. These costs, however, are only partly borne by private households through net-income reductions. The most common accounting scheme for company cars in Germany is the 1%-scheme (Finanztip, 2017): under this scheme, the car is treated as a non-monetary benefit; company car users have to pay income tax on 1% of the new car price for the car that has been given to them by their employer; this rate increases by 0.03% for each km of single commuting distance. We illustrate this using an example: The new car price of a given company car is 40.000 Euro; the commuting distance of the employee is 15 Km (German average); and the assumed income tax rate is 20%. In this example the monthly net-income reduction is 116 Euros resulting from an assumed 580 Euro (40.000*[0.01+15*0.0003] = 580) non-monetary benefit for which 20% tax has to be paid.

While this example is inspired by realistic figures, it does not claim to be representative. Nevertheless, it illustrates that it is very likely that most company car beneficiaries only pay a fraction of the costs that the vehicle really causes.

5. Conclusions and outlook

This paper presented a multistage method for imputing car costs by cost item in a national travel survey data set, the fuel consumption and odometer reading survey 2015 and 2016 (MOP-FCOR) of the German Mobility Panel (MOP). Based on vehicle information reported by survey participants, we assigned suitable car model specifications to each of the three thousand cars in the data set. Using these model specifications, car costs per item were assigned to each vehicle using the most comprehensive German vehicle cost data base maintained by the largest German car club ADAC. After this initial cost imputation, there were still numerous missing values in our data set. To close these data gaps and to compute vehicle depreciation over time we estimated linear regression models predicting vehicle values. Through this imputation procedure we generated a vehicle data base with three thousand vehicles including vehicle costs per cost item. Based on this data base we computed average costs per vehicle per month and cost distributions. In order to validate the average cost figures for private vehicles we also analyzed the German income and expenditure survey EVS. The comparison with the EVS figures pointed to some differences between imputed cost information and average EVS expenditures. According to our assessment, there were logical explanations for these differences, which were by and large not implausible but provided additional insights.

On average, the total cost of ownership for a private car in Germany is about 315 Euros per month. This translates to about 31 Eurocents per auto-km. About one third of the costs are fuel, another third is depreciation, and the rest are other mainly fixed costs (insurance, tax, repair and maintenance). However, the cost distribution is strongly skewed with a long tail to the right. This means that the majority of motorists pay less than average for their private vehicles while few pay more and evidently some pay a lot more. On average, company cars produce more than twice the costs of private cars, mostly because they are more expensive, newer and are used more intensively. However, private households only bear a fraction of the costs of their company cars.

We are aware of caveats in our current imputation procedure. There are two related main sources for bias and error in the imputed cost data: Firstly, missing information on residual values and repair costs for old vehicles; there is little we can do about this data gap because the ADAC vehicle cost data base does not contain this information. Secondly, there is room for improvement concerning the multivariate models that we used to impute missing values and depreciation. As a next step in researching vehicle costs we are planning to move from the current robust linear model to a more advanced model which also does a better job at modelling depreciation of older vehicles.

Despite these caveats in the current version of the imputed cost data, we believe the imputation approach delivers unprecedented vehicle cost information in particular with regard to the distribution of vehicle costs. Vehicle cost distribution information is paramount for understanding car ownership and car usage choices. For example, the majority of cars are much less expensive than average figures suggest. If this is true, the potential for replacing private vehicles by car sharing may be strongly overrated. We believe that in an environment of a new and increasing mobility service economy – possibly additionally stimulated by vehicle automation in the future – it will be paramount to understand the fundamentals of mobility choices. Data on the details and the distribution of vehicle costs as we obtain them through our imputation procedure provide important insights in this context.

Acknowledgements

This paper presents excerpts of work in the projects VEU TP1500 (Tobias Kuhnimhof) funded by the German Aerospace Center DLR and „Substitutionspotenziale der Pkw-Nutzung” funded by Ministry of Transport Baden-Württemberg (Christine Eisenmann).

References

ADAC. (2017). Info, Test & Rat - ADAC Autokosten. Retrieved September 5, 2017, from https://www.adac.de/infotestrat/autodatenbank/autokosten/default.aspx?ComponentId=282164&SourcePageId=287163

ARAL. (2017). Kraftstoffpreis-Archiv. Retrieved September 11, 2017, from http://www.aral.de/de/retail/kraftstoffe-und-preise/kraftstoffpreise/kraftstoffpreis-archiv.html

Autobild. (2016). Ohne Unfall fährt's sich billiger Retrieved September 5, 2017, from http://www.autobild.de/artikel/schadenfreiheitsklasse-und-schadenfreiheitsrabatt-36479.html

Bäumer, M., Hautzinger, H., Pfeiffer, M., Stock, W., Lenz, B., Kuhnimhof, T., & Köhler, K. (2017). Fahrleistungserhebung 2014 – Inländerfahrleistung. In BASt (Ed.), *Berichte der Bundesanstalt für Straßenwesen*. Bergisch Gladbach.

Destatis. (2017). Einkommens- und Verbrauchsstichprobe (EVS). Retrieved 25.4.2017, 2017, from https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/Methoden/Einkommens_Verbrauchsstichprobe.html

Finanztip. (2017). Dienstwagenbesteuerung, Mit Fahrtenbuch oder über die Ein-Prozent-Regel versteuern. Retrieved September 5, 2017, from http://www.finanztip.de/dienstwagen-besteuerung/

Hughes, T., Liu, Z., & Castro, P. (2015). Residual Car Values Forecasting Using AutoCycle. In M. Analytics (Ed.), *ECONOMIC & CONSUMER CREDIT ANALYTICS*.

Kraftfahrtbundesamt. (2017a). Verzeichnis der Hersteller und Typen der für die Personenbeförderung ausgelegten und gebauten Kraftfahrzeuge mit mindestens vier Rädern (Klasse M).

Kraftfahrtbundesamt. (2017b). Zentrales Fahrzeugregister (ZFZR). Retrieved September 5, 2017, from https://www.kba.de/DE/ZentraleRegister/ZFZR/zfzr_node.html

Kuhnimhof, T., Ottmann, P., & Zumkeller, D. (2008). *Adding Value to Your Data: Analysis of Travel Expenses Based on Trip Diary and Enriched Odometer Reading Data*. Paper presented at the ISCTSC Conference 2008, Annecy.

Statista. (2015). Bestand an Verträgen in der Kfz-Versicherung in Deutschland im Jahr 2015 nach Sparten (in Millionen). Retrieved September 5, 2017, from https://de.statista.com/statistik/daten/studie/247196/umfrage/bestand-an-vertraegen-in-der-kfz-versicherung-nach-sparten/

STAYCLEAN. (2017). STAYCLEAN Textile Autowäsche, Flatrate Retrieved September 11, 2017, from https://staycleancarwash.com/flatrate

VDA. (2017). Deutscher Oldtimer Index legt 2016 um 4,4 Prozent zu [Press release]. Retrieved from https://www.vda.de/de/presse/Pressemeldungen/20170203-Deutscher-Oldtimer-Index-legt-2016-um-4-4-Prozent-zu.html

Verbraucherzentrale. (2016). Kfz-Versicherung: Pflicht für alle Halter von Kraftfahrzeugen. Retrieved September 5, 2017, from https://www.verbraucherzentrale.de/Kfz-Versicherung-Pflicht-fuer-alle-Halter-von-Kraftfahrzeugen

Weiß, C., Chlond, B., Behren, S. v., Hilgert, T., & Vortisch, P. (2016). Deutsches Mobilitätspanel (MOP) - Wissenschaftliche Begleitung und Auswertungen Bericht 2015/2016: Alltagsmobilität und Fahrleistung.