Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 32 (2018) 421–435

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Steering Committee for Transport Survey Conferences

Some pay much but many don't: Vehicle TCO imputation in travel surveys

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 specifically on 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 analyze the German income and expenditure survey EVS. The average total cost of ownership for a private car in Germany is about 310 Euros per month. This translates to about 30 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 key for understanding the fundamentals of mobility choices.

Peer-review under responsibility of the International Steering Committee for Transport Survey Conferences (ISCTSC).

Keywords: car ownership; imputation; vehicle costs; cost of car ownership; German Mobility Panel; Income and Expenditure Survey

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

2352-1465
Peer-review under responsibility of the International Steering Committee for Transport Survey Conferences (ISCTSC).
10.1016/j.trpro.2018.10.056

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. Research on total costs of ownership (TCO) of cars often focusses on new vehicles, for example in the context of alternative drive trains (Hagman et al. 2016; Wu et al. 2015; Letmathe and Suares 2017). Moreover, there is information – often by car clubs – on the costs of specific vehicles or vehicle types (AAA 2017; ADAC 2017). However, to the knowledge of the authors there is no data set available that provides insight into how real costs are distributed across the population of vehicles and drivers. Nevertheless, such information appears to be paramount for a broad variety of mobility market analyses such as establishing the market potential of mobility-as-a-service concepts.

This paper presents an imputation of car costs in a German National Travel Survey by assigning costs per item based on detailed vehicle information and a vehicle cost data base. The study uses the vehicle data set from the 2015 and 2016 German Mobility Panel (MOP) data and is a sequel to an earlier vehicle cost imputation study using 2005 MOP data (Kuhnimhof et al. 2008). Imputation in the context of travel surveys usually aims at minimizing the number of missing values among items in the questionnaire (Meister 2016; US Department of Transportation 2011; Lepanjuuri et al. 2016). Our vehicle cost imputation differs from this in that we add variables to the vehicle data which did not exist in the data set before and were not part of the questionnaire. We do so by fusing information from different data sets.

As a result of this imputation procedure we obtain a data set which contains monthly costs per item (depreciation, repair and maintenance, tax, insurance, fuel) for about 3,000 cars. Monthly averages from this data set can be compared to results from income an expenditure surveys. We use such comparison in order to validate our results. Based on this comparison we discuss advantages and shortcomings of our imputation procedure. What is unique about our imputation data set is its ability to provide distributions of car costs across the population of vehicles, including different vehicle types, ages and usage patterns. The paper also presents such distributions allowing for important conclusions about the true costs of car ownership in Germany. These results imply that for many driving is much less costly than average values or costs of new cars suggest.

2. Data

This section presents all four data sets that were used in this study. 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. Table 1 gives an 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 et al. 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 that the sample represents the population of vehicles in private German households best possible. Selectivity studies (Kuhnimhof et al. 2006) and thorough annual data quality checks documented by transparent reporting (Weiß et al. 2016) indicate a high validity of the MOP survey.

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 VKT (vehicle kilometers traveled) for every car in the sample. Monthly VKT as measured in the MOP-FCOR survey align well with annual average VKT figures of the German mileage survey in 2014 (Bäumer et al. 2017): on average, private cars in the MOP-FCOR survey reported 1,052 km monthly VKT in 2014 which translate into 12,624 (=12*1,052) annual VKT. This matches well with 12,333 km annual VKT per private car as measured in the German mileage survey. This indicates that annual VKT derived from monthly MOP-FCOR results represent average annual VKT in Germany reasonably well and seasonality is not a big concern.

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 which we list in the following along with key words on how these costs were generated (a more detailed explanation is available from ADAC (ADAC Fahrzeugtechnik 2018)):

- new car price in the year of construction: manufacturer information;
- used car price in 2016: drawn from Germany's most extensive data base on used car sales run by Deutsche Automobil Treuhand (DAT Group 2018);
- repair and maintenance in 2016: based on spare part costs and average hourly workshop rates for typical maintenance and wear repairs;
- fuel in 2016: based on EU-driving cycle fuel consumption and average fuel costs (we did not use ADAC fuel costs in our cost imputation);
- oil in 2016: average costs;
- AdBlue (diesel exhaust fluid) in 2016: average costs;
- car wash in 2016: lump sum of annually 250€ (we did not use car wash costs in our cost imputation);
- tax in 2016: computed on the basis of the official vehicle taxing scheme;
- insurance in 2016 (fully comprehensive cover, partly comprehensive cover, liability): based on the ADAC's own vehicle insurance rates;

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.

Despite these shortcomings, the ADAC data base is the most reliable source for car costs in Germany. It provides costs by item for all cars for which costs can be reliably established. From the judgement of ADAC these are all vehicles up to the age of 12 years. 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.

Income and expenditure surveys represent a useful source for average spending by expenditure items, e.g., spending on fuel or purchase of vehicles. However, 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. Hence, distributions on car related expenditures from income and expenditure surveys suggest a variance which is much higher than in reality. (The same is true for other household expenditure categories). 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. Due to the time lag between the surveys (EVS is from 2013, MOP-FCOR data sets from 2015 and 2016) we cannot expect perfect consistency of the results. At the same time, substantial changes in car related expenditures between 2013 and 2016 are not to be expected either, as this was a period of relative stability with regard to the economic situation and consumer prices in Germany (Statistisches Bundesamt 2018). Fuel prices which decreased substantially from 2013 to 2015/2016 are an exception which we discuss later as we present the results. In light of these considerations, comparing with EVS2013 is the best option to compare imputation results with secondary sources on transportation expenditures.

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.

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 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 regression models to predict the new car price (in €) and the residual value (percentage of the new car price) of the cars in our data set. There were three reasons for 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 odometer mileage of the vehicles.

Table 2 shows the results of both multivariate regression models. In both cases, we used linear regressions with the new car price and the residual value being the explained variables. Lagrange multiplier tests were applied to test for heteroscedasticity; test results (p-value >0.3 for all lag windows) indicate that the assumption of homoscedastic error terms is valid. Explanatory variables were car drive, motor power, car brand (premium/non-premium, country of manufacture), cylinder capacity, car segment as well as total odometer mileage and car age (the latter two account for the residual value model only). The car segment definition of the German Federal Motor Transport Authority KBA is applied to our dataset; cars are classified by visual, technical and market-oriented properties (Kraftfahrtbundesamt 2018). 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 odometer mileage (see also (Hughes et al. 2015)). Interaction variables (e.g., odometer mileage and drive, odometer mileage and brand) were tested, but turned out not to be significant. 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.

We are aware that linear regression models have theoretical shortcomings in this case. Firstly, both – new car prices and residual values – are not normally distributed which suggests linearizing the explained variable. Hence, we also tested a log-linear model. 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 (i.e., re-transforming the log of the error term would lead to error term results, which average in 1 instead of in 0 due to the properties of the log-function). However, for our imputation obtaining values which – on average – do not overestimate or underestimate vehicle values was key. Secondly, 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. However, irrespective of the type of model used, there is a more fundamental restriction to predicting residual values of old cars (21% of the MOP-FCOR sample is older than 12 years and 3% of the MOP-FCOR is even older than 20 years): depending on the cars' state of maintenance, there is the potential that residual values for vintage cars rise in Germany with increasing age (VDA 2017). This phenomenon cannot be reflected by a residual value model as the ADAC vehicle cost data does contain residual values of cars older than 12 years. ADAC's reasoning for not providing these residual values are the great uncertainties associated with old vehicles' state of maintenance. Hence, even a more sophisticated model, e.g. a truncated regression model (i.e., applied when the available data to estimate the dependent variable are truncated) or a censored regression model (i.e., applied when the dependent variable is censored), would not be able to capture residual values beyond a certain age correctly.

In light of these considerations and because of the very good fit of our regression models (R-Squares of 0.84 and 0.96) which do not leave much room for improvement we opted for using linear regression despite their theoretical disadvantages. Finally, random elements were not included in the value imputation. This is consistent with the other cost items drawn from the ADAC vehicle cost data base (average values per HSN-TSN and model year combination) which do not incorporate possible variation of car costs within each HSN-TSN category either.

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 odometer mileages (see Table 2). However, the ADAC data base only assumes average annual VKT. This might differ substantially from the individual odometer 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.

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 odometer mileage [10.000 km]	.	.	-0.030	<.0001
Total odometer mileage over 50.000 km (i.e., max (0; tot. mileage – 50.000 km) [10.000 km]	.	.	0.022	<.0001
Total odometer 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*	

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 odometer mileage in 2017. This was done by adding 12 times the car's monthly VKT as reported in the MOP-FCOR survey to the 2016 total odometer mileage (see section 2.1). 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.

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 VKT, ADAC test cycle fuel consumption and assumed average prices for fuel. However, from the MOP-FCOR, which collects fuel consumption along with odometer mileage, we have 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 VKT (monthly VKT 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 private cars, commercially registered cars (i.e., company cars) and all 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 most results is absolutely satisfactory. Nevertheless, there are differences between the EVS and imputation data results which point to potential deficiencies of the imputation process. In light of the findings from both approaches Table 3 also lists estimates for real figures for average costs per cost item for private cars. In the following we discuss selected cost issues and the associated imputation problems as well as our reasoning for the average real cost estimates. 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 2015/2016				**EVS 2013**	**Estimated 2015/2016**	
	Mean	**StdDev**	**Min**	**Max**	**Mean**	***Mean***	***Comment***
Private cars							
Fuel & lubricants [€/month]	92.0	58.6	0.0	553.0	101.8	*92*	*Imputed value*
Depreciation [€/month]	105.6	109.9	0.0	1,573.0	100.4	*103*	*Rounded mean*
Insurance [€/month]	102.7	40.9	28.0	360.0	35.8	*36*	*EVS value*
Repair and maintenance [€/month]	80.0	18.4	46.0	224.0	54.5	*67*	*Rounded mean*
Tax [€/month]	12.9	8.5	2.0	61.0	11.6	*12*	*Rounded mean*
Total costs [€/month]	393.1	163.2	140.0	2,268.0	304.2	*310*	
Sample size	*2,546 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*						
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*						

4.1. Vehicle depreciation

According to EVS, German households on average spend 100.4 Euros per car per month for buying and leasing vehicles after accounting for income generated through selling cars. By and large, these expenditures should reflect the average vehicle depreciation per car and month. For private cars, our imputation procedure arrives at a very similar result (105.6 Euros per car 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 do at least decrease only marginally 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, see 3.2) 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 and we estimate the average monthly cost of depreciation per private car to be 103 Euro.

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 plausible explanations for the differences between the EVS and the cost imputation values. 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. An estimate of an average of about 67 Euros appears to be plausible.

4.3. Fuel and lubricants

According to EVS, Germans spent around 102 Euros per car per month on fuel in 2013. According to the imputed data this figure was around ten Euros less in 2015/2016. In this case, fuel price decreases between 2013 (Super 95 E5 Gasoline: 160.3 Eurocent per liter; Diesel: 143.4 Eurocent per liter) and 2015 (Super 95 E5 Gasoline: 139.2 Eurocent per liter; Diesel: 117.2 Eurocent per liter) and 2016 (Gasoline: 130.1 Eurocent per liter; Diesel: 108.7 Eurocent per liter) are a likely explanation. Given the reference period (2015/2016) of our imputed data, we believe the result of the imputation is closer to reality.

4.4. Tax and insurance

With regard to tax, the absolute difference between the EVS result (11.6 Euros per car per month) and the imputation result (12.9 Euros per car per month) is small. This conforms to expectation as the fixed taxing scheme allows for almost no uncertainty as regards the tax given the vehicles properties.

As for consistency of EVS and imputation results, insurance costs are the big exception with costs per month according to the imputed data being about 65€ higher than according to the EVS. In section 2.2 we explain the reasons for this divergence that mainly origin in the disregard of any discounts that are very common in reality. It appears very likely that average insurance expenditures according to EVS are closer to reality than those resulting from the cost imputation procedure.

4.5. Key findings concerning private vehicle TCO

On average, holding a private car in Germany costs about 310 Euros per month which translates to about 30 Eurocents per km given an annual mileage of 12,333 km (Bäumer et al. 2017). 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, the real advantage of the MOP-FCOR data set with imputed costs is its ability to provide car cost distributions which income and expenditure surveys cannot provide. Figure 1 shows weighted empirical cumulative distributions for costs per car per month by cost item based on the imputed data. This data has been scaled with a correction factor applied to all observations such that the means per cost item align with our estimates for real figures from Table 3.

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, indicating that the majority of vehicles cost substantially less than the average. Hence, most motorists pay less than average for their cars.

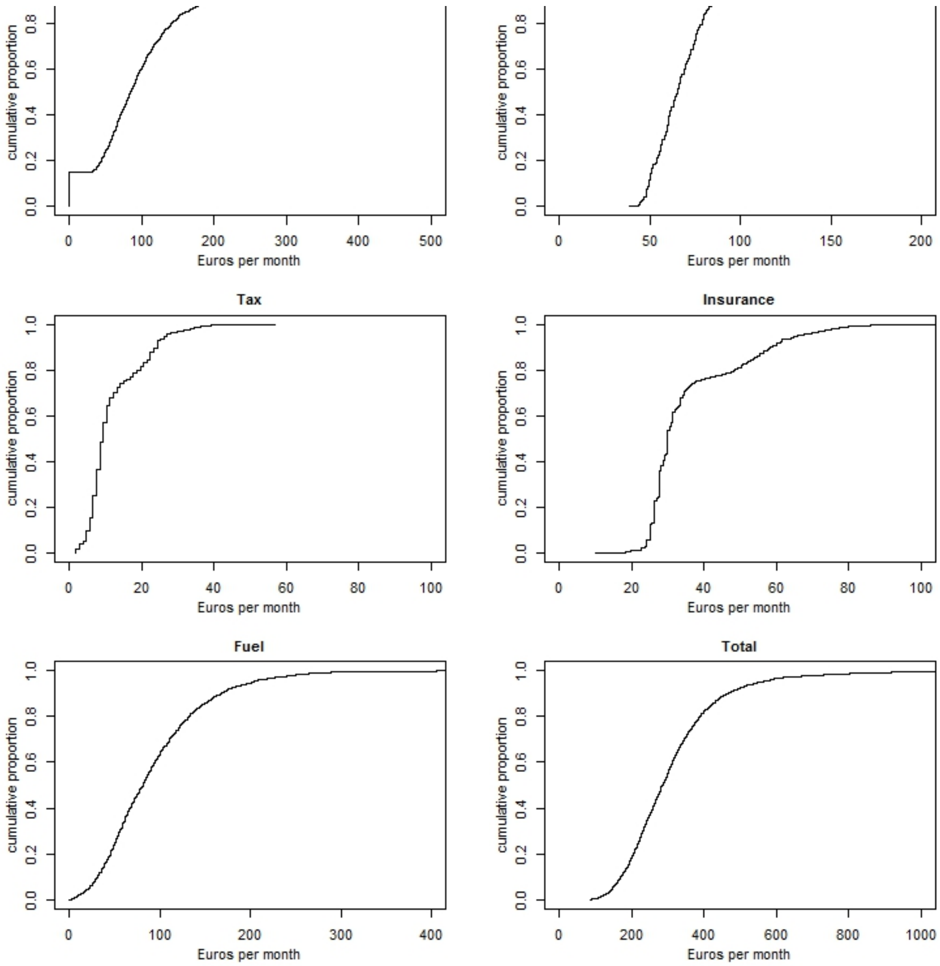

Figure 1: Empirical cumulative distributions of car costs per item based on the MOP-FCOR data set with imputed costs (private cars)

4.6. 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 rather small. Second, a 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 VKT of commercially registered cars of about 2,040 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 and company car taxation (Finanztip 2017).

5. Conclusions and outlook

This paper presented a multistage method for imputing car costs by cost item in the fuel consumption and odometer reading survey (MOP-FCOR) of the MOP. Based on vehicle information reported by MOP survey participants, we assigned suitable car model specifications to each car 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. To close remaining data gaps and to compute vehicle depreciation over time we estimated regression models predicting vehicle values. Through this imputation procedure we generated a vehicle data base with 3000 vehicles including vehicle costs per cost item. Based on this data base we computed average monthly costs per vehicle and cost distributions. In order to validate the average cost figures for private vehicles we 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 plausible and provided additional insights.

On average, the total cost of ownership for a private car in Germany is about 310 Euros per month. This translates to about 30 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.

Despite caveats of the imputed cost data which we discuss in the paper, 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 presented in this paper 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

AAA (2017) AAA Newsroom. Your Driving Costs. https://newsroom.aaa.com/auto/your-driving-costs/. Accessed 27.3. 2018

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

ADAC Fahrzeugtechnik (2018) ADAC Autokosten, Berechnungs-Grundlage für die standardisierte Kostenberechnung. ADAC. https://www.adac.de/_mmm/pdf/autokosten_grundlagen_47084.pdf.

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

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

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

DAT Group (2018) Information centre for the European automobile industry. https://www.dat.de/en-int/home.html. Accessed 26.3. 2018

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

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

Hagman J, Ritzén S, Stier JJ, Susilo Y (2016) Total cost of ownership and its potential implications for battery electric vehicle diffusion. Research in Transportation Business & Management 18:11-17. doi:https://doi.org/10.1016/j.rtbm.2016.01.003

Hughes T, Liu Z, Castro P (2015) Residual Car Values Forecasting Using AutoCycle. 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). https://www.kba.de/DE/ZentraleRegister/ZFZR/zfzr_node.html. Accessed September 5 2017

Kraftfahrtbundesamt (2018) Bestand am 1. Januar 2017 nach Segmenten. https://www.kba.de/DE/Statistik/Fahrzeuge/Bestand/Segmente/segmente_node.html. Accessed 27.3. 2018

Kuhnimhof T, Chlond B, Zumkeller D (2006) Nonresponse, selectivity, and data quality in travel surveys - Experiences from analyzing recruitment for the German mobility Panel. Travel Survey Methods, Information Technology, and Geospatial Data (1972):29-37

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 8th Conference on Transport Survey Methods, Annecy, France, 28.5.2008

Lepanjuuri K, Cornick P, Byron C, Templeton I, Hurn J (2016) National Travel Survey 2015, Technical Report. NatCen,

Letmathe P, Suares M (2017) A consumer-oriented total cost of ownership model for different vehicle types in Germany. Transportation Research Part D: Transport and Environment 57:314-335. doi:https://doi.org/10.1016/j.trd.2017.09.007

Meister R (2016) Evaluation verschiedener Imputationsverfahren zur Aufbereitung großer Datenbestände am Beispiel der SrV-Studie von 2013. Technische Universität Dresden, Dresden

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

Statistisches Bundesamt (2018) Preise, Erzeugerpreise gewerblicher Produkte (Inlandsabsatz), Preise für leichtes Heizöl, schweres Heizöl, Motorenbenzin und Dieselkraftstoff. https://www.destatis.de/DE/Publikationen/Thematisch/Preise/Erzeugerpreise/ErzeugerpreisePreisreiheHeizoelPDF_5612402.pdf?__blob=publicationFile. Accessed 10.12. 2017

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

US Department of Transportation FHA (2011) 2009 National Household Travel Survey, User's Guide. U.S. Department of Transportation, Federal Highway Administration,

VDA (2017) Deutscher Oldtimer Index legt 2016 um 4,4 Prozent zu.

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

Weiß C, Chlond B, Behren Sv, Hilgert T, Vortisch P (2016) Deutsches Mobilitätspanel (MOP) - Wissenschaftliche Begleitung und Auswertungen Bericht 2015/2016: Alltagsmobilität und Fahrleistung. Institut für Verkehrswesen, Karlsruhe

Wu G, Inderbitzin A, Bening C (2015) Total cost of ownership of electric vehicles compared to conventional vehicles: A probabilistic analysis and projection across market segments. Energy Policy 80:196-214. doi:https://doi.org/10.1016/j.enpol.2015.02.004