Modelling mobility tool availability at a household and individual level: A case study of Switzerland

Tim Hillel[1], Janody Pougala[1], Patrick Manser[2], Raphael Luethi[2], Wolfgang Scherr[2], and Michel Bierlaire[1]

[1]*École Polytechnique Fédérale de Lausanne, Switzerland*
[2]*Swiss Federal Railways (SBB), Switzerland*

Abstract

Prediction of the availability of mobility tools, including vehicles, driving licenses, and public transport subscriptions, is a key task for modelling individual travel and activity-scheduling behaviour. In this paper, we present a new framework for prediction of mobility tool availability, based on sequential logit models of individual driving license ownership, household car ownership, and individual public transport subscriptions. The framework is applied to data derived from the Swiss Mobility and Transport Microcensus (MTMC) to estimate a mobility tool availability model for Switzerland. The utility functions in each sequential model are determined using a *machine learning assisted specification* approach.

The resulting model, which is being incorporated into the Swiss Federal Railways' nationwide microscopic Agent Based Model (ABM), *SIMBA MOBi*, is used to simulate mobility tool ownership for a full synthetic population. The predictions are validated against regional and national-level statistics. The results show the new model is able to accurately represent nationally aggregated driving licence ownership by age-group, as well as regionally-aggregated household vehicle ownership and individual *half fare travelcard* ownership, though further calibration is needed for annual public transport subscriptions. Crucially, the new model will enable the modelling of complex household interactions during activity scheduling and mode choice.

Introduction

Individual travel and activity-scheduling behaviour is heavily dependent on the availability of mobility tools, including vehicles, driving licenses, and public transport subscriptions. It is therefore crucial to accurately represent mobility tool availability in transport demand models in order to make realistic behavioural predictions. For Agent Based Models (ABMs) that make use of a synthetic population, mobility tool availability must be predicted for each individual in the population.

Existing approaches to this problem rely on joint models of individual car availability and public transport subscriptions (Danalet and Mathys 2018; Loder and Axhausen 2018). These models have three key limitations. Firstly, as these models make predictions at an individual level, they do not model interactions of household and individual decisions. Secondly, car availability is typically modelled as binary (available or not available), and so do not model different levels of vehicle sharing. Finally, as predictions are made jointly, the relationship between short-term decisions, in our case buying a public transport subscription, and long-term decisions, like getting a driving licence or buying a car.

In this paper, we introduce a new framework for prediction of mobility tool availability, which addresses the above limitations through the use of sequential logit models of individual driving license ownership, household car ownership, and individual public transport subscriptions, as shown in Fig. 1.

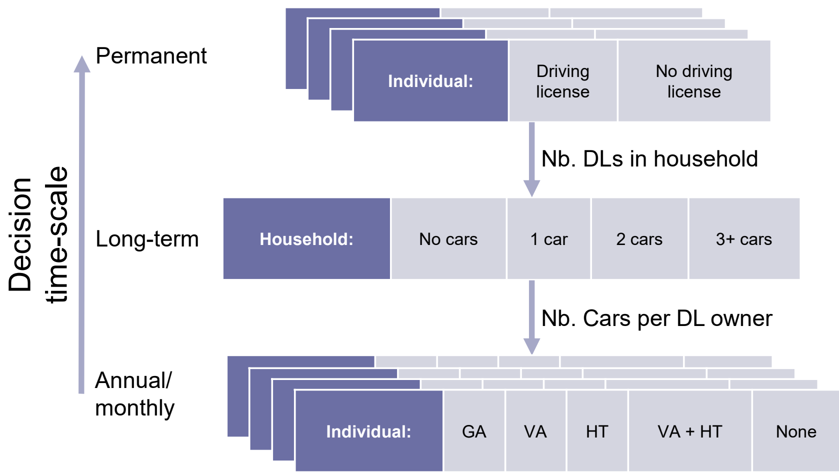

Figure 1: Proposed sequential structure for mobility tools availability model

The framework is applied to data derived from the Swiss Mobility and Transport Microcensus (MTMC) to estimate a mobility tool availability model for Switzerland which will be incorporated into Swiss Federal Railways' nationwide microscopic ABM, *SIMBA MOBi*. As well as the sequential structure shown in Fig. 1, the mobility tool availability model developed in this paper incorporates three further recent contributions from existing research. Firstly, the utility functions in each sequential model are identified using the *machine learning assisted specification* approach introduced by Hillel et al. (2019). Secondly, the socio-economic and household structure information from the MTMC data is augmented with transport service indicators and network statistics calculated for over 8000 zones covering Switzerland (including measures of public transport, car, and combined accessibility, and parking costs), building on the existing work of (Loder and Axhausen 2018). Finally, logit model estimation is performed using the HAMABS algorithm (Lederrey et al. 2019), resulting in drastically reduced fit-times.

Methodology

Modelling framework

The modelling framework proposed in this paper consists of three logit models applied in sequence:

1. **Individual driving license ownership**: Binary Logit (BL) model predicting the driving license (DL) ownership (DL/no DL) for each adult in the household
2. **Household car ownership**: Multinomial Logit (MNL) model predicting the number of cars owned by a household, given the number of driving licenses in the household (from model 1).

3. **Individual Public Transport (PT) subscription**: MNL model predicting the PT subscriptions owned by each adult in the household, given the household car ownership (from model 2) and individual driving license ownership (from model 1).

This structure is shown visually in Fig. 1.

The proposed framework has three advantages over the joint estimation approach. Firstly, the sequence models intra-household interactions by modelling household car ownership dependent on total individual driving license ownership, as well as individual PT subscription ownership dependent on household car ownership and structure. This allows the model to capture household behaviours, e.g. high proportional car ownership households which are less likely to use (and therefore own subscriptions for) public transport services. Secondly, by determining the number of cars in the household alongside the number of driving licenses, this structure allows for a complex modelling of car-sharing arrangements (i.e. the number of cars per driver). This will allow for more realistic behavioural modelling for activity planning and mode choice in ABMs. Finally, the sequence provides a realistic representation of the causal effects of decisions at different time-frames. Annual/monthly PT subscription decisions are made dependent on long-term car-ownership decisions, which in turn are dependent on the lifelong choice of acquiring a driving license.

Case study

SBB has developed a state-of-the-art microscopic, activity-based transport model called *SIMBA MOBi*, which is currently used to support medium and long-term transport planning projects (Scherr, Manser, and Buetzberger 2019). SIMBA MOBi is composed of three major modules: a synthetic population based on Switzerland's real 2017 population (Bodenmann et al. 2019), an activity-based demand model (Scherr, Joshi, et al. 2019) and a network flow simulation using the MATSim open-source framework (Horni, Nagel, and Axhausen 2016).

SIMBA MOBi currently incorporates a MNL model individual car availability and PT subscription trained on an indivdual's socio-economical features and some geographical attributes (Danalet and Mathys 2018). The model jointly predicts binary car availability (car/no car) and five-classes of PT subscription: (i) the *GA travelcard* (GA), which provides unlimited nationwide PT travel; (ii) a *regional travelcard* (VA), which provides unlimited PT travel in a defined region of Switzerland; (iii) the *half fare travelcard* (HT), which provides half-price travel on PT services nationwide; (iv) a regional travelcard in combination with a half fare travelcard (VA+HT); and (v) no PT subscription. This results in 10 joint classes (car+GA, car+VA, car+HT, car+VA+HT, car only, GA only, VA only, HT only, HT+VA with no car, no mobility tools).

For the second iteration of SIMBA MOBi, SBB intends to incorporate more realistic household interactions in the activity planning and mode-choice predictions. Furthermore, SBB wishes to incorporate relationships between transport service indicators/network metrics and mobility tool ownership into the model. As such, the mobility tools availability model developed in this paper, which addresses limitations of joint models of car availability and PT subscription ownership, is intended to replace the existing model.

Data

The base dataset used to estimate the mobility tools availability model is the 2015 Swiss MTMC data, a nationwide survey conducted every 5 years that aims to obtain a global understanding of Swiss residents' mobility patterns and behaviours. The 2015 MTMC data is used, which contains data for 163 843 individuals across 57 090 households.

Each sequential model in the mobility tools availability model is estimated on a different dataset composed of socio-economic and household structure variables from the MTMC data, augmented with transport service indicators and network metrics. These include network accessibility measures, parking costs, and population density.

The transport service indicators and network metrics are calculated for 8000 zones covering all of Switzerland. The accessibility measures are generated for each zone using a utility based approach. To calculate the accessibility for a given zone, the *attraction* of each other zone (a compound measure of the population and jobs in that zone) is multiplied by the utility of travelling to that zone (estimated from SBB's microscopic traffic simulation model), and summed over all other zones. This approach is used to calculate accessibility measures for each zone for the car, public transport, and combined car and public transport modes. The parking costs are calculated as the average cost of parking per hour in each zone. Finally, the density is calculated as the combined total number of inhabitants and jobs per square kilometre.

An overview of the variables in the finished dataset is given in Table 1. The individual models use all variables, whilst the household car ownership model uses the household and zonal variables.

Table 1: Description of individual, household, and zonal variables in dataset.

Level	Variable	Type	Categories
Individual	Nationality	Categorical	Swiss, other
	Age	Integer	
	Employment rate	Categorical	Full-time, part-time, unemployed
Household	N. inhabitants	Integer	
	N. adults	Integer	
	N. children	Integer	
	Type	Categorical	Couple with children, couple without children, single, other
	Language	Categorical	German, French, Italian, Other
Zonal	Degree of Urbanisation	Categorical	City, town, rural
	Density	Continuous	
	Parking cost	Continuous	
	PT accessibility	Continuous	
	Car accessibility	Continuous	
	Combined PT+Car accessibility	Continuous	

Model specification and estimation

According to the proposed framework, three logit ownership models are estimated for the Swiss mobility tool availability model: (i) a binary individual driving license ownership; (ii) a four-class household car ownership (no cars, one car, two cars, three-or-more cars); and (iii) a five-class individual PT subscription (none, GA, VA, HT, VA+HT).

The utility specifications in each model are identified using the machine learning assisted specification approach proposed by Hillel et al. (2019). Following this approach, a Gradient Boosting Decision Trees (GBDT) model is first estimated for each sequential model. The structure of each trained GBDT models is then used to select the most relevant features. For continuous features, the distributions of the split-points in the trees in the ensemble are analysed to identify non-linear transforms of input features, including *log, Box-Cox,* and *piecewise-linear* transforms. Due to word-limit constraints, a detailed overview of the assisted specification process is not given here.

The logit models are estimated in Biogeme using the *HAMABs* algorithm (Bierlaire 2018; Lederrey et al. 2019). This results in greatly reduced fit-times, allowing for several different forms of the utility functions to be tested during model specification.

The utility functions, parameter values, and summary statistics for each finalised logit model are given in Tables 2, 3, and 4 in the appendices.

Model validation

The nature of the MTMC data means it is not possible to directly validate the complete modelling sequence directly, as only one respondent per household provides full socio-economic information required to predict individual driving license and PT subscription ownership (and so the full chain cannot be estimated). Instead, the model is used to simulate mobility tools ownership for a synthetic population of Switzerland, and validated against aggregate regional control values. SIMBA MOBi's synthetic population is based on Switzerland's real population in 2017, and contains 8.6 million persons (7 million adults) across 3.8 million households.

The sequential model estimated on the MTMC data is applied directly to the synthetic population, without further calibration. Discrete decisions for each individual/household are drawn from the output choice probabilities of each model, according to the Monte Carlo method. The simulation of driving license ownership is used to define the inputs to the household car-ownership model (e.g. number of driving licenses, number of driving licenses per adult, etc.). The simulations of driving license ownership and car ownership are then used to define the inputs to the PT subscription model.

The model estimates on the synthetic population are validated externally against aggregate control totals measured at a regional level (106 regions). Note that, as the ground-truth data is for the real population and the mobility tools availability model is used to simulate ownership for the synthetic population, this process does not directly validate the ownership model. As such, any discrepancies may be the result of the ownership model (including any bias in the MTMC data) *or* the synthetic population. However, the validation approach effectively validates the combination of the mobility tools availability model in combination with the synthetic population.

Results

Driving license ownership

Driving license information is not available for Switzerland at a regional level. Instead, the mobility tools model is validated against nationwide driving license statistics. Figure 2 shows the number of driving licenses simulated for all adults in the synthetic population compared to the control totals aggregated by age-group (ASTRA 2018).

Overall, Fig. 2 shows a good fit of the model results. The model slightly underestimates the number of driving licenses of people between 25 and 74 years old, whilst slightly overestimating the number of driving licenses owned by adults bellow the age of 24 or above the age of 75.

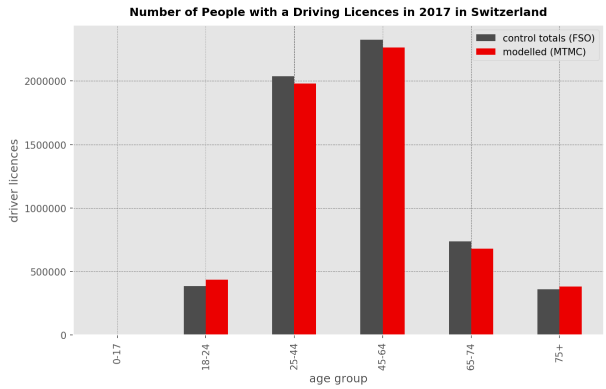

Figure 2: Validation of the driving license ownership predictions by age-group, aggregated at nationwide-level

Car ownership

The mobility tools availability model predicts the number of cars owned by each individual household. Figure 3 shows the results aggregated by the 106 MS regions (Federal Statistical Office 2016) compared to the control totals provided by ASTRA based on the MOFIS-database (Federal Statistical Office 2020). The scatter-plot on the left compares the absolute number of cars while the one on the right compares the number of cars per 1000 inhabitants. The results show that the the model generalizes considerably well on the synthetic population.

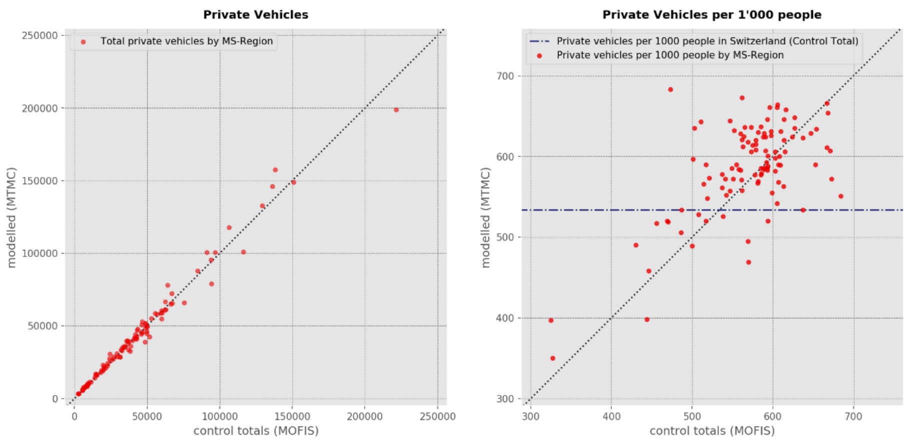

Figure 3: Validation of the household car ownership predictions, aggregated at regional level

PT subscription ownership

Predicting PT subscription ownership is inherently more challenging than either driving license or car ownership. This is because PT subscriptions depend highly local fares and personal affinity to public transport. There are strong regional variations which cannot necessarily be explained by the features included in the model (there is no calibration by canton or commune during model estimation). For instance, whilst they are both well connected cities with high relative density and accessibility measures, there is a much lower rate of public transport subscriptions in Geneva compared to Zurich. Furthermore, PT subscription ownership is very influenced by both home and work locations. This explains why certain *commuter towns*, e.g. Brig, have much higher rates of PT subscription.

Figure 4 shows the total number of predicted PT subscriptions of each type grouped by MS-region and compared to the control totals from the SBB customer database (Swiss Federal Railways 2017). While the number of nationwide travelcards (GA) is marginally overestimated (especially in areas with high public transport accessibility), the model applied to the synthetic population substantially underestimates the number of regional travelcards (VA). This may be due to respective over and under-representation in the MTMC data. The estimation for owning a half-fare travelcard (HT) shows the best fit compared to the empirical data.

Conclusions and further work

In this study, we present a new framework for prediction of mobility tool availability. The framework is used to estimate a new mobility tool availability model for Switzerland, to be incorporated into SBB's nationwide ABM *SIMBA MOBi*. The finalised model is validated against regional and national ownership statistics. It is found that the model applied to the synthetic population predicts well aggregate driving license ownership at a national level and car and half-fare travelcard ownership at a regional level. The model needs to be further improved to accurately predict *unlimited travel* PT subscriptions. In particular, we find strong regional variations of PT subscription which are not adequately represented by the features currently included in the model.

As well as improving the models predictive ability, a major direction for further work will be incorporating the added flexibility of the proposed mobility tools availability framework into a functional ABM. More specifically, by predicting the number of cars and driving adults in each household, further realistic behavioural assumptions (including complex household interactions) can be incorporated into the activity planning and mode-choice models. Furthermore, the modularity offered by separating the mobility availability model into three sequential steps, each with a different time horizons, opens new possibilities for interaction with other long-term decisions in the ABM. For example, the work or school location choice could be modelled before the car-ownership/PT-subscription decision, therefore allowing the work location to be used as an input feature.

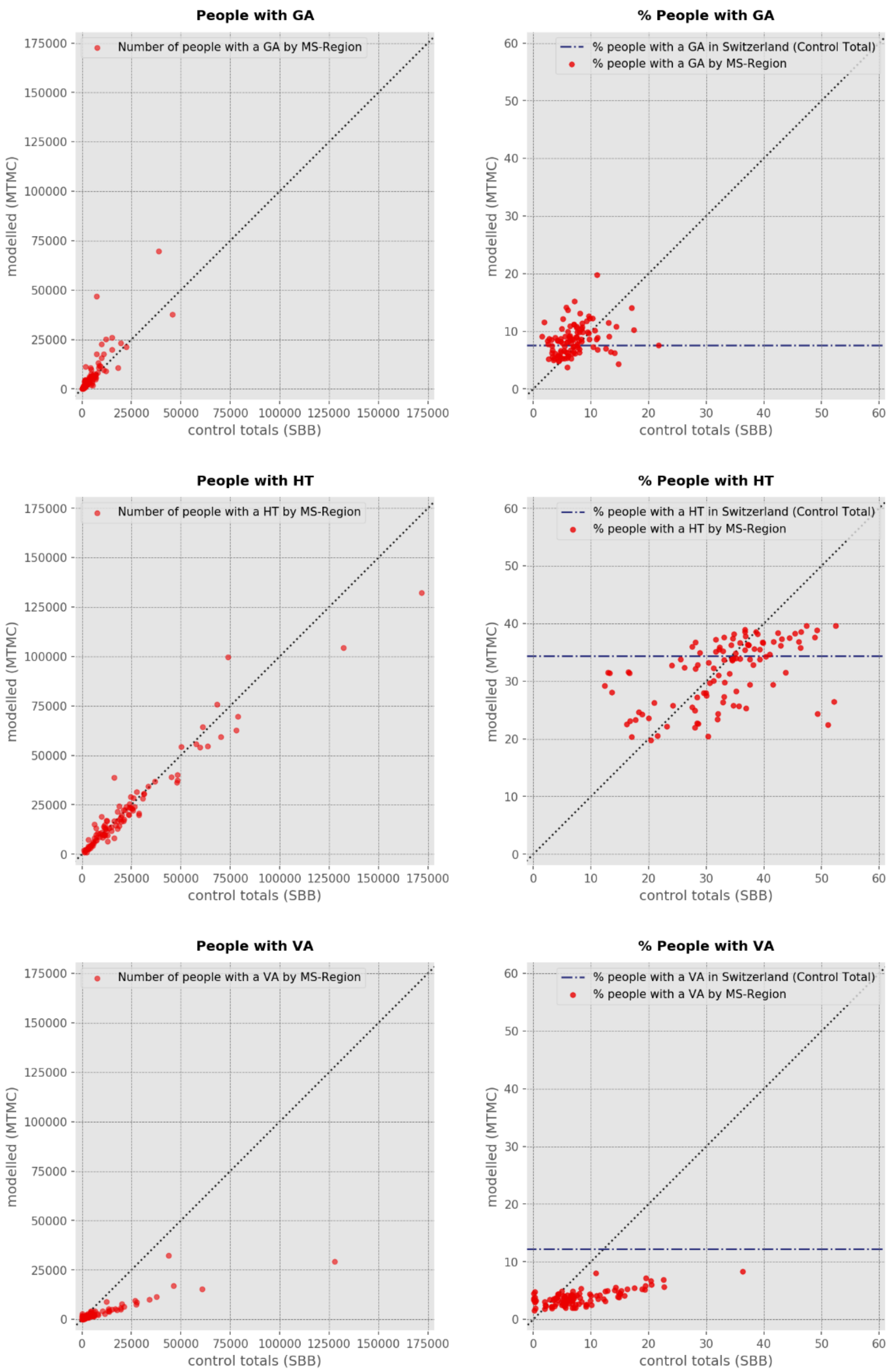

Figure 4: Validation of the individual PT subscription predictions, aggregated at regional level

Appendices

Table 2: Parameter values and summary statistics for individual driving license model. All parameters are significant at 2.5% significance level.

Variable	Parameter values	
	No DL	DL
ASC		
ASC No DL	-	-
ASC DL	-	-11.3
Individual		
Age 18-22 (piecewise)	-	0.514
Age 23-26 (piecewise)	-	0.0879
Age 27-69 (piecewise)	-	0.0102
Age 70-89 (piecewise)	-	-0.125
Age 90+ (piecewise)		-0.199
Nationality	-	0.631
Employment rate: full-time	-	1.43
Employment rate: part-time	-	0.749
Household		
Type of household: couple with children	-	0.578
Type of household: couple without children	-	0.488
Type of household: other	-	0.161
Language: French	-	0.183
Network		
PT accesibility (Box-Cox)	-	-0.647
Combined Car+PT accessibility (Box-Cox)	-	0.413
Parking cost (linear)	-	-0.751
Parking cost (log)	-	0.344
Free parking	-	-0.554

Summary statistics	
Number of parameters	18
Sample size	$48\,629$
Initial log-likelihood	$-33\,707.05$
Final log-likelihood	$-17\,212.77$
Cross entropy loss	-0.354
$\bar{\rho}^2$	0.489
Estimation time (s)	5.41

Table 3: Parameter values and summary statistics for household car ownership model. All parameters are significant at 2.5% significance level.

Variables	**Parameter values**			
	0	**1**	**2**	**3+**
ASC				
ASC 0	-			
ASC 1	-	2.81		
ASC 2	-		2.41	
ASC 3+	-			2.55
Household				
Type of household: couple with children	-	0.588	1.75	1.32
Type of household: couple without children	-	0.992	1.89	1.22
Type of household: other	-	-0.576	0.329	-
Language: German	-	-0.389	-0.747	-0.78
Number of driving licenses	-	-0.314	-	1.1
Number of driving licenses per adult	-	4.59	7.26	6.47
Number of children	-	0.175	0.227	0.086
Number of adults	-	1.1	1.77	1.68
Network				
Combined PT+Car accessibility	-	-1.32	-2.2	-2.49
PT accessibility (log)	-	-0.339	-0.587	-0.723
Car accessibility (Box-Cox)	-	0.484	0.967	1.09
Parking cost (linear)	-	-0.697	-1.62	-2.35
Parking cost (log)	-	0.317	0.693	1.07
Free parking	-	-0.555	-1.23	-1.86
Degree of Urbanisation: city	-	-0.177	-0.307	-0.292

Summary statistics	
Number of parameters	46
Sample size	57 065
Initial log-likelihood	$-79\,108.89$
Final log-likelihood	$-49\,757.06$
Cross entropy loss	-0.872
$\bar{\rho}^2$	0.371
Estimation time (s)	44.1

Table 4: Parameter values and summary statistics for individual PT subscription model. All parameters significant at 2.5% significance level.

Variables	Parameter values				
	None	GA	VA	HT	HT+VA
ASC					
ASC None	-				
ASC GA	-	2.48			
ASC VA	-		6.33		
ASC HT	-			4.24	
ASC VA+HT	-				4
Individual					
Age 18-22 (piecewise)	-	-0.156	-0.371	-0.29	-0.341
Age 23-26 (piecewise)	-	-0.227	-0.168	0.0821	-0.14
Age 27-69 (piecewise)	-	-0.004	-0.013	0.012	-
Age 70-89 (piecewise)	-	-0.028	-	-0.0376	-0.021
Age 90+ (piecewise)	-	-0.209	-0.336	-0.432	-0.197
Nationality	-	1.11	0.111	0.927	0.848
Employment rate: full-time	-	0.16	-	-	0.236
Employment rate: part-time	-	0.151	0.268	0.305	
Owns driving license	0.187	-			
Household					
Language: German	-	0.803	-0.203	0.791	0.801
Numbers of cars	-	-	-0.125	-0.295	-
Number of cars per adult	-	-2.31	-1.51	-	-0.701
Network					
PT accessibility PT (Box-Cox)	-	-	-	-0.192	-
Car accessibility (Box-Cox)	1.48	-	-	-	-
Combined PT+Car accessibility (Box-Cox)	-1.85	-	-	-	-
Degree of Urbanisation: city	-	-	0.755	-	-

Summary statistics	
Number of parameters	45
Sample size	48 636
Initial log-likelihood	$-78\,276.62$
Final log-likelihood	$-55\,715.11$
Cross entropy loss	-1.15
$\bar{\rho}^2$	0.288
Estimation time (s)	25.2

References

ASTRA (2018). *Führerausweisstatistik*. URL: https://www.newsd.admin.ch/newsd/message/attachments/52768.pdf (visited on 02/13/2020).

Bierlaire, Michel (2018). *PandasBiogeme: A Short Introduction*. TRANSP-OR 181219, p. 22.

Bodenmann, Balz, Pascal Burkei, Camilla Philipp, Nadja Bernhard, and Krill Muller (2019). *Synthetische Population 2017 – Modellierung mit dem Flächennutzungsmodell FaLC*. Federal Office for Spatial Development (ARE) and Swiss Federal Railways (SBB), p. 95.

Danalet, Antonin and Nicole Mathys (2018). "Mobility Resources in Switzerland in 2015". In: *18th Swiss Transport Research Conference (STRC2018)*, p. 28.

Federal Statistical Office (2016). *MS-Regionen*. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/raum-umwelt/nomenklaturen/msreg.assetdetail.415729.html (visited on 02/13/2020).

— (2020). *Road Vehicles - Stock, Level of Motorisation*. URL: https://www.bfs.admin.ch/bfs/en/home/statistics/mobility-transport/transport-infrastructure-vehicles/vehicles/road-vehicles-stock-level-motorisation.html (visited on 02/13/2020).

Hillel, Tim, Michel Bierlaire, Mohammed Elshafie, and Ying Jin (2019). "Weak Teachers: Assisted Specification of Discrete Choice Models Using Ensemble Learning". In: *hEART 2019, 8th Symposium of the European Association for Research in Transportation*, p. 12.

Horni, Andreas, Kai Nagel, and Kay W Axhausen (Aug. 10, 2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press.

Lederrey, Gael, Virginie Lurkin, Tim Hillel, and Michel Bierlaire (2019). *HAMABS: Estimation of Discrete Choice Models with Hybrid Stochastic Adaptive Batch Size Algorithms*. TRANSP-OR 191213. Switzerland: Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, p. 43.

Loder, Allister and Kay Werner Axhausen (Apr. 12, 2018). "Mobility Tools and Use: Accessibility's Role in Switzerland". In: *Journal of Transport and Land Use* 11.1.

Scherr, Wolfgang, Chetan Joshi, Patrick Manser, Nathalie Frischknecht, and Denis Métrailler (2019). "An Activity-Based Travel Demand Model of Switzerland Based on Choices and Constraints". In: *hEART 2019, 8th Symposium of the European Association for Research in Transportation*. Budapest, Hungary, p. 13.

Scherr, Wolfgang, Patrick Manser, and Patrick Buetzberger (2019). "Microscopic Mobility Simulation for Corporate Planning". In: 47th European Transport Conference (ETC2019). Dublin, Ireland.

Swiss Federal Railways (2017). *Generalabo / Halbtax*. URL: https://data.sbb.ch/explore/dataset/generalabo-halbtax/information/ (visited on 02/13/2020).