MODELING MOTIF CHOICE: COMPARING THE PREDICTION PERFORMANCE OF DIFFERENT DISCRETE CHOICE MODELS TO ANALYZE TEMPORAL STABILITY

SEBASTIAN BÜSCHER

Bielefeld University, Econometrics

DIETMAR BAUER

Bielefeld University, Econometrics

MANUEL BATRAM

Bielefeld University, Econometrics

Abstract. Daily human mobility can be summarized using so-called mobility motifs. These are directed mathematical graphs that depict the locations visited and the trips between these for a full day. The significance of the motif is that out of many possible motifs only a few occur frequently in reality. The frequency with which the motifs are chosen is surprisingly stable across different observation conditions both in terms of measurement technology (cellphone-based as well as traditional mobility surveys) and setting (different cultural context including American and European regions and cities).

In this paper, we use a large German mobility survey panel collected over more than two decades in order to show that the motif choice patterns are also stable over time. Moreover, we model the choice of one of the eleven most frequently chosen motifs as a function of underlying socio-demographic data.

Hereby three different models are used: The Multinomial Logit (MNL) model, Mixed Multinomial Logit (MMNL) models and the Multinomial Probit (MNP) model. We compare the corresponding results both in terms of their estimation accuracy as measured using the log-likelihood value out-of-sample as well as the corresponding numerical effort. For the MMNL model we also investigate the temporal stability of the model.

The main results can be summarized as follows: First, the out-of-sample performance is modest for all three approaches. This hints at the fact that the motif choice is not much influenced by most of the socio-demographic variables. Second, the MMNL and the MNP models improve the out-of-sample performance modestly, hinting at existing correlations between the motif choices. And third, the estimated models do not change much over time for the sample investigated.

All these findings contribute to the perception that motifs choice is extremely stable in many aspects. Therefore motif choice could be a valuable basis on which agent-based mobility simulation models can be based upon.

Keywords: mobility motif; temporal stability; MNP model; MACML

1. Introduction

The modelling of human mobility is of importance, for example for the assessment of the effects of changes in the mobility system on the usage of critical traffic infrastructure and the corresponding environmental impacts. Traditionally for this purpose mobility simulation models have been calibrated using a trip based approach [12, 6].

In the recent past activity-based models, such as the one employed in the software MATSim [10], have become more popular. These activity-based models represent the daily mobility of persons as their main topic.

It has been discovered recently [13], that daily human mobility patterns can be described with human mobility motifs which are directed graphs encoding the visited locations as nodes and the trips between

E-mail addresses: Sebastian.Buescher@uni-bielefeld.de, Dietmar.Bauer@uni-bielefeld.de, Manuel.Batram@uni-bielefeld.de.

the locations as edges. Although there is a vast amount of possible motifs to choose from, it has been observed in several studies for different regions and diverse data sources, that only a small number of motifs suffices to explain large parts of the observed motif choices. This finding has been confirmed using a number of different spatial settings as well as different observation technology [13, 14, 17].

The reason underlying the motif choice, however, has not been analyzed carefully up to now. Secondly, the before mentioned studies only use very recent data material. Thus it is of importance to investigate whether the motif choice frequencies are also stable over time.

This is the research gap this paper tries to fill: We investigate the temporal stability of motif choice as well as the underlying determinants. This is done on a data set from the German Mobility Panel (MOP), a mobility survey spanning over 20 years of daily German mobility data.

To this end, MNL, MMNL, and MNP models are fitted to subsets of the data, and their prediction performance and computational cost are compared. Thus, as a second contribution of this paper, we compare different choice models in terms of their application to a large real-world data set. The data set shows a relatively large number of alternatives (we use the 11 most frequent motifs plus the 'all other' category) from a large number of choices in a panel data set with potentially up to 21 repeated observations for each individual. In this paper, however, the individuals are treated separately for each year reducing the number of observed choices per person to a maximum of 7. Additionally only weekdays are considered and days with reported abnormalities are excluded, leaving up to 5 observations per individual. We compare both the prediction accuracy of the various models as well as the numerical effort for estimation.

Finally, the temporal stability and the dependence of motif choice on various socio-demographic variables is investigated by firstly analyzing the changes in choice frequencies over time in the data and secondly, by partitioning the data into consecutive time periods and fitting the most promising model to the underlying socio-demographic variables for each of the time periods.

The paper is organized as follows: In the next section, the definition of mobility motifs is provided. Preliminary descriptive analysis is used to present the data set. In section 3, we describe the models and estimation algorithms used in this paper. The application of the estimation to the motif data sets is described in section 4. The best-fitting model is then investigated in terms of its temporal stability in section 5. Finally, section 6 concludes the paper.

2. Motifs: Definition and descriptive Analysis

Human mobility motifs were developed by [13] to analyze daily mobility patterns in passively collected data obtained from sources like mobile phone billing data or GPS data.

[13] define a mobility 'motif' as a mathematical graph containing nodes (representing visited locations) and directed edges, representing the trips between them. A major difference to classical definitions of trip-chains is that the locations in motifs are unlabeled, which is due to the anonymous nature of the data source for which the motifs were developed.

Beside GPS tracks motifs can also be computed based on traditional mobility diary data. Figure 1 sketches this process[1] Plot (a) sketches an exemplary mobility diary of one day, (b) abstracts the information from (a) into a movement graph whereby, the different locations are symbolized as nodes of the graph and the directed edges are given by the trips from one location to another. Plot (c) leads to the motif by aggregating multiple edges between the same nodes, adding artificial vertices for round tours[2] Since only the purpose of the trip is given, only the endpoint of each trip can be directly inferred. For some purposes (such as home or work-based trips) identical nodes can be identified while purposes like 'shopping' result in new vertices for each shopping trip.

For the first observed trip for each person, the origin is unknown in the mobility survey. Here we assume that the trip is started from home.

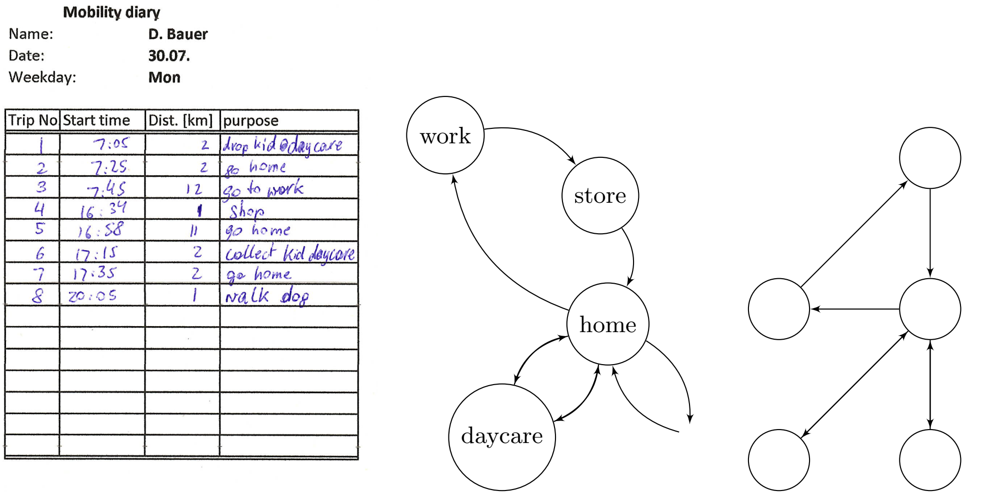
A related issue concerns the start and the end of trips: Most of the observed days end at the same location where they started. In the remaining cases, two options are possible: to introduce motifs (referred to as open motifs below) that are not related to closed graphs or to omit these observations. For this paper, we choose the first option. But since the frequency of open motifs is small compared to the most frequently chosen motifs, all motifs relating to an open graph ended up in the category for 'others'.

Even when only considering days with no more than six visited locations, more than one million possible mobility motifs arise. However, [13] found that the 17 most commonly chosen motifs made up over 90% of the observed daily mobility motifs in their data set. [14] and [17] were able to confirm these

[1]The figure also appears in the forthcoming chapter [5].

[2]This is a somewhat arbitrary design choice. Alternative options would be to omit round trips or to introduce edges from a node into itself.

FIGURE 1. Stylized example of the motif extraction process. Source: [5].



Trip No	Start time	Dist. [km]	purpose
1	7:05	2	drop kid @ daycare
2	7:25	2	go home
3	7:45	12	go to work
4	16:34	1	shop
5	16:58	11	go home
6	17:15	2	collect kid daycare
7	17:35	2	go home
8	20:05	1	walk dog

(A) Example page from a stylized mobility diary

(B) Resulting mobility graph

(C) Resulting motif

findings when observing similar choice frequencies among the most commonly chosen motifs in other data sets, obtained from different regions and varying data sources.

The data set used in this paper is from the MOP (for details on the data set see [19] and `https://mobilitaetspanel.ifv.kit.edu`). The MOP is a mobility survey with a rotating sample of households which are kept in the survey for up to three years. We use data from 1994 up to 2013 spanning two decades. In total, information on 8,722 households, 15,864 persons and 230,769 individual daily mobility patterns from all over Germany is available in the data set used. The data set contains information from week-long trip diaries from the household members as well as a wealth of socio-demographic data, which include information on a personal-level (age, sex, employment status, education, ...) and on household-level (number of household members, number of cars, number of young children, location ...), as well as trip specific effects (weekend, weather, special days ...). The trip diary data is used to assign the according human mobility motif to each recorded person and day. A stylized example of the motif extraction process can be seen in the figure 1.

The motif choice of each person for each recorded day depending on several socio-demographic variables is then used as the discrete choice to be modelled. It was previously shown that even though there is a vast amount of possible motifs to choose from, a small fraction of motifs covers large parts of the observed motifs. Consequently, we model the choice between the 11 most commonly chosen motifs whilst the remaining motifs are aggregated in one group as "others".

The same stability of motif distribution is also present in the MOP data set. Moreover, due to the long record of data over two decades, it is also possible to observe a temporal stability of the frequency of the chosen motifs. This stability is shown in figure 2[3]. For each of the 11 most common motifs in the MOP data set a time series is plotted showing the relative frequency of the motif for each year. It can be seen that the ordering of the seven most frequent motifs is identical over most of the 20 year time span. Moreover their frequency is almost constant.

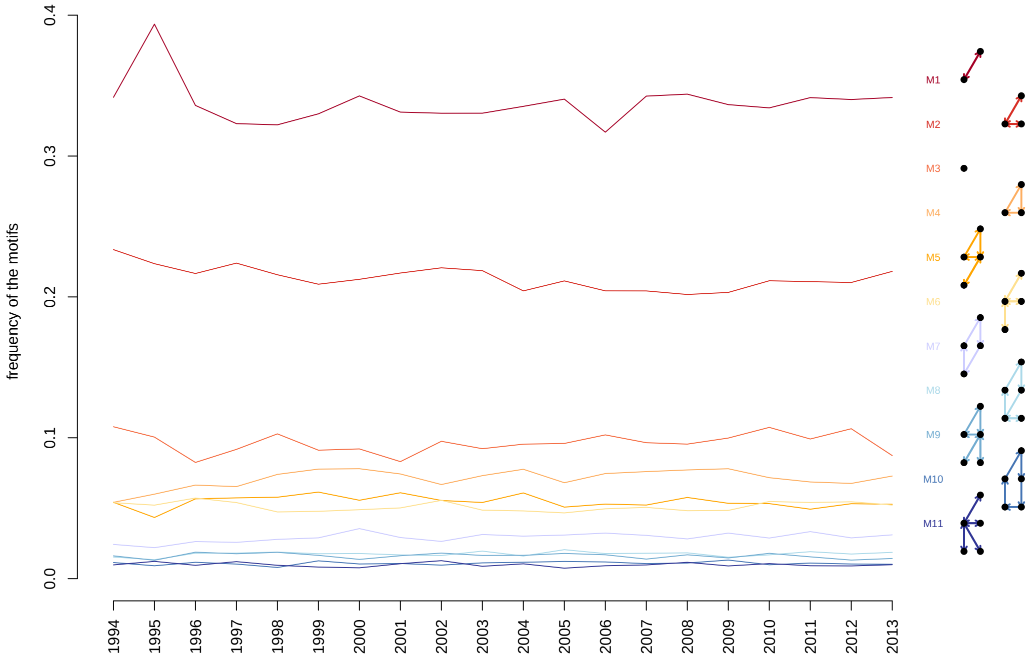
In addition to the data of the mobility diaries, the MOP data set contains a wealth of explanatory variables on the personal and household level. This includes sex, age, employment status, the number of cars available to the household, type of household, access to public transportation as well as various other variables (for more details on the data set see [19] and `https://mobilitaetspanel.ifv.kit.edu`).

3. CHOICE MODELS USED AND ESTIMATION ALGORITHMS

In this paper, Multinomial Logit (MNL), Mixed Multinomial Logit (MMNL) and Multinomial Probit (MNP) models are considered to model the motif choices. Whilst MNL models are easiest to compute due to the closed form of the choice probabilities, the models suffer from the independence of irrelevant alternatives (IIA) assumption. In MMNL models this assumption is relaxed and the MNP model is completely unaffected by the IIA assumption.

[3]This plot is taken from [5].

FIGURE 2. Frequency of the 11 most common motifs for all 20 years (1994-2013). Source: [5].



For the MNL models, following [15], the probability that decider n chooses alternative i out of the choice set of $j = 1, ..., J$ alternatives in her t-th decision has the form

$$P_{nti}(\beta) = L_{nti}(\beta) = \frac{\exp(X'_{nti}\beta)}{\sum_{j=1}^{J} \exp(X'_{ntj}\beta)}.$$

Here X_{ntj} describes the characteristics of the j-th alternative influencing the preferences for alternative j. This model has been given the interpretation as a Random Utility Model (RUM).

In our data set – as is common for mobility choice data – the regressors X_{ntj} include alternative specific constants (ASCs) influencing the preferences for an alternative which are not explained by underlying characteristics. Moreover, the ASCs vary depending on the socio-demographics of the deciders. This limits the interpretation as a Random Utility Model (RUM) since the utility inferred from a choice is no longer explained by characteristics of the alternatives but determined by some given constants biasing the preferences. For motif choice, in this paper, all regressors are of this type.

For identifiability, the utility of one alternative needs to be restricted to zero, as only differences in systematic utilities $(X_{ntj} - X_{nti})'\beta$ matter. We will always restrict the utility of the alternative corresponding to the 'stay at home' motif $M3$ to zero.

The MNL model can be easily estimated by minimizing the likelihood. The panel data setting is no more complicated than the cross-section as the errors are considered to be independent over choices. The likelihood, as well as first and second derivatives, can be readily calculated analytically. For this paper, we used the `R`-package `apollo` [8] to estimate all MNL models.

However, the MNL suffers from the IIA assumption. One way to alleviate this assumption is to use nested logit models [15]. In the case of motif choice, no obvious nesting structures exist. It has been noted, however (see [13]), that individuals tend to use certain motifs jointly such that classes of motifs can be defined within which switching is more frequent than out of the class. This is weak evidence for the violation of the IIA assumption.

MMNL models [15] allow incorporating correlations between error terms. They are obtained from the MNL model by allowing for randomly distributed coefficients β. The choice probabilities are changed to

$$P_{nti}(\beta) = \int L_{nti}(\tilde{\beta}) f_\beta(\tilde{\beta}) d\tilde{\beta} = \int \frac{\exp(X'_{nti}\tilde{\beta})}{\sum_{j=1}^{J} \exp(X'_{ntj}\tilde{\beta})} f_\beta(\tilde{\beta}) d\tilde{\beta}. \tag{1}$$

Different distributions $f_\beta(\tilde{\beta})$ have been proposed in the literature. Most commonly the various components are modelled as being independently distributed with marginal distributions being chosen from three densities: normal, log-normal or tent-distribution (supported on an interval). Furthermore, it is common to allow some components to be random while others are kept deterministic. While correlations between components of $\tilde{\beta}$ can be included in the specification most often this is not done.

Estimation of MMNL models is more complicated than estimating MNL models. The integral contained in equation (1) can typically not be evaluated analytically. Thus the standard approach for estimation is the Maximum Simulated Likelihood (MSL) method. [15] provides a detailed discussion of the statistical properties of the corresponding estimates. The number of random draws for the estimation is typically guided by the numerical affordability. Instead of independent, identically distributed (iid) draws correlated draws such as Halton draws [15] are often used to reduce the computation load.

For the MMNL models in this paper we chose to treat the ASCs as mixed between individuals with normal distribution and no correlation between them. The other parameters were treated as being deterministic. This way we can accord for random taste variations while keeping the straight forward interpretation of the other parameters. All MMNL models were estimated with MSL using 100 Halton draws for each mixed parameter. Also here the `R`-package `apollo` was used to perform the estimations.

A different approach is chosen for MNP models [15]. Here the error $e_{nt:} = (e_{nti})_{j=1,...,J}$ in the random utility formulation

$$U_{nti} = X'_{nti}\beta + e_{nti}$$

is assumed to be multivariate normally distributed. The alternative with the maximal random utility is chosen. Additionally assuming β to be random allows to model unobserved heterogeneity in preferences.

For mixed MNP models hence the choice probabilities are given as

$$(2) \qquad P_{nti}(\beta) = \int M_{nti}(\tilde{\beta}) f_\beta(\tilde{\beta}) d\tilde{\beta} = \int \Phi_{J-1}\left(\tilde{e}_{ntj|i} \leq (X_{nti} - X_{ntj})'\tilde{\beta}, j = 1, ..., J, j \neq i; \Sigma_i\right) f_\beta(\tilde{\beta}) d\tilde{\beta},$$

where $\Phi_{J-1}(., \Sigma_i)$ denotes the multivariate normal cumulative distribution function (MVNCDF) corresponding to a multivariate normally distributed random variable of dimension $J - 1$ with expectation zero and variance Σ_i denoting the variance of $(\tilde{e}_{ntj|i})_{j\neq i} = (e_{ntj} - e_{nti})_{j\neq i}$.

The error terms of the utility functions are assumed to be independent across individuals and time but correlated across alternatives for a given individual and time point. So, Σ_i is a general covariance matrix. As we will discuss later, a suitable normalization strategy needs to be applied to make the model estimable because not all of those parameters are identified. In correspondence to the MMNL model we only allow for independent random effects of the ASCs. These random effects are modelled as multivariate normal with diagonal covariance matrix Ω.

Again, such a model cannot be evaluated analytically. Estimation using MSL is one option.

For the panel data case, [2] suggested a different estimation approach, Maximum Approximate Composite Marginal Likelihood (MACML). This approach combines two ideas, the approximation of the MVNCDF and usage of the Composite Marginal Likelihood (CML) instead of the full likelihood. For the CML, MACML relies on a pairwise marginal likelihood specification and the analytic approximation of the MVNCDF is the Solow-Joe approximation (see [11]). Simulations in [4] show that this approach potentially reduces the numerical load in comparison to MSL methods for a given accuracy (in terms of closeness to the maximum likelihood estimate), in particular for data sets with a relatively large choice set and a large number of repeated choices per person in the panel. Both conditions are met for our data set. The asymptotic properties of the MACML method are discussed in [1]. There it is shown that the method can be interpreted as using a distorted mapping $\tilde{P}_{nti}(\beta)$ depending on the approximation concept chosen. In this sense standard asymptotics hold.

We use a variant of the MACML approach but in order to improve the computational performance, we deviate from the original MACML approach by only using pairs of adjacent observations instead of all possible pairs in the CML part. The idea of only considering adjacent pairs was first considered in [16] who also showed by simulation that this method works well for time-dependent categorical data. In a more general sense, it is just a weighting scheme. Weights are commonly used in the CML literature to adjust the relative impact of marginal likelihoods in order to improve the efficiency of the resulting estimator (see [9]). In this regard, the results in [18] suggest that if the dependency structure in the panel data is not too complex then the efficiency loss of only considering adjacent pairs is minor.

Using this estimation algorithm, our aim is to estimate the linear coefficients (β) as well as the covariance parameters of the random effect (Ω) and the error terms (Σ). The covariance parameters are estimated using the corresponding Cholesky decomposition. To ensure identification, we take utility differences with respect to the "stay-at-home" motif. This motif is chosen to be the reference in order to aid interpretation.[4] In the same vein, the dimensions of Ω and Σ are reduced, which do then represent

[4]For this normalization, the interpretation is somewhat similar to the ordered probit model.

the correlation across the various differences with respect to the stay at home motif. For Σ we fix the first diagonal entry to ensure identification.[5]

For MACML estimation there is no standard package and all computations are done in the R statistical computing language with performance-critical parts written in C++. The CML is optimized using the quasi-Newton BGFS routine of `nlm()` function with analytic gradients but other than that default settings. The optimizer is initialized using draws from the standard normal distribution for all parameters.

3.1. **Model Selection.** In this paper, the three methods MNL, MMNL and MNP, are compared for modeling the motif choices in the MOP data set, both in terms of prediction performance as well as numerical effort.

The process of model selection in the current context comprises two different issues: First, the impact of socio-demographic variables on prediction accuracy needs to be decided upon. And second, we need to distinguish the three models (MNL, MMNL, MNP) with respect to their appropriateness.

Concerning the socio-demographic variables we choose the following ones from the large number of variables available in the MOP data set:

- Sex: The data set contains males and females over the age of 10.
- Age: The age in years at the time of data collection is given in the data set.
- Household type: With the variables available in the MOP we classified the households into single-person households without children, couples without children, single with one child, other with children, other without children.
- Region: classified into large cities, medium-sized cities, and rural areas.
- number of cars available in the household: zero, one, and more than one car.
- Occupational status: full-time, part-time employees, unemployed, students and apprentices, retirees, and home keepers.

Except for age, all other variables are categorical in nature with relatively few categories. Age has an impact on the motif choice even if the other variables are corrected for. This will be visible below. In terms of the functional dependence, preliminary investigations lead us to use cubic dependency on age, with linear and quadratic specifications as alternatives. Note that the introduction of the cubic term, additional to the squared term, adds another eleven parameters to the MNL model.

For the remaining variables, dummy coding of these variables leads to category-specific ASCs for each category. Two different methods of fitting models can be distinguished: Either one model with many parameters is fitted or separate models containing only ASCs and continuous variables for all observations for each category are fitted. The first approach amounts to very large models with many variables and a large sample size. The latter approach involves fitting many models with small sample size and relatively few parameters involved in each model. Numerical reasoning leads to using the latter method for our setting.

In a first step, the role of the various variables is investigated using MNL models with ASCs and cubic dependence on age. A large number of different models is fitted and validated for several categories using K-fold cross-validation with $K = 5$. Each category is uniquely defined by the five previously mentioned socio-demographic variables.

For the cross-validation, the corresponding population $\mathbf{S}_{\text{pop}}$ is randomly partitioned into 5 equally sized groups $\mathbf{S}_{\text{pop}_1}, \ldots, \mathbf{S}_{\text{pop}_5}$. The model is then fitted to the sets of observations $\mathbf{S}_{\text{pop}_{-k}} := \bigcup_{i \neq k} \mathbf{S}_{\text{pop}_i}$, resulting in K different parameter vectors $\hat{\beta}_{\text{pop}_{-k}},\ k = 1 \ldots K$. For each of these models, the log-likelihood per observation

$$ll_{\text{obs}_k}(\hat{\beta}_{\text{pop}_{-k}}) = ll(\hat{\beta}_{\text{pop}_{-k}}, \mathbf{S}_{\text{pop}_k})/|\mathbf{S}_{\text{pop}_k}|,$$

of the validation group $\mathbf{S}_{\text{pop}_k}$ is calculated, where $|\mathbf{S}_{\text{pop}_k}|$ denotes the number of observations in group $\mathbf{S}_{\text{pop}_k}$. These are the category-specific out-of-sample log-likelihoods per observation.

To compare the prediction performance of the category-specific model fit to a more general model fit, one model is fitted once to the whole data set, not selecting for specific categories. The fitted parameter vector for the reference model is $\hat{\beta}_{\text{ref}}$. The reference log-likelihood per observation is then calculated as

$$ll_{\text{obs}_k}(\hat{\beta}_{\text{ref}}) = ll(\hat{\beta}_{\text{ref}}, \mathbf{S}_{\text{pop}_k})/|\mathbf{S}_{\text{pop}_k}|$$

[5]Furthermore, it is necessary to refill the column and row, which are 'lost' due to the differencing, with zeros. The reason is that all other utilities are differenced with respect to an alternative in order to compute its probability (see [3] for a more detailed discussion of those standard computations in MNP model estimation).

and compared to the category-specific out-of-sample log-likelihoods per observation by calculating the weighted mean difference

$$\overline{\Delta ll_{\text{obs}}} = \sum_{k=1}^{K} \frac{\left(ll_{\text{obs}_k}(\hat{\beta}_{\text{pop}_{-k}}) - ll_{\text{obs}_k}(\hat{\beta}_{\text{ref}})\right) \; |\mathbf{S}_{\text{pop}_k}|}{\sum_{r=1}^{K} |\mathbf{S}_{\text{pop}_r}|}.$$

A positive value suggests that the category-specific model has a better out-of-sample performance than the reference model, which in turn indicates that the inclusion of a dummy variable for this category might improve the general model. Note, that these log-likelihoods are computed for the same observations, but the underlying models are fitted on different observations.

To make sure that the increase in out-of-sample prediction performance is statistical significant a Diebold-Mariano [7] inspired t-test is performed with

$$t_{\text{DMS}} = \frac{\overline{\Delta ll_{\text{obs}}}}{s_{\Delta ll_{\text{obs}}}/\sqrt{K}},$$

where $s_{\Delta ll_{\text{obs}}}$ is the weighted corrected sample standard deviation

$$s_{\Delta ll_{\text{obs}}} = \sqrt{\frac{K}{K-1} \sum_{k=1}^{K} \frac{\left(ll_{\text{obs}_k}(\hat{\beta}_{\text{pop}_{-k}}) - ll_{\text{obs}_k}(\hat{\beta}_{\text{ref}}) - \overline{\Delta ll_{\text{obs}}}\right)^2 \; |\mathbf{S}_{\text{pop}_k}|}{\sum_{r=1}^{K} |\mathbf{S}_{\text{pop}_r}|}}.$$

With the assumption that the test statistic t_{DMS} has a t-distribution with 4 degrees of freedom, we can do a one-sided test at the 95% confidence level with the test rejecting for values larger than 2.13.

For the selection of the variables we estimate models splitting the sample according to each variable in turn. Thus for example all 145 308 observations are split into those of persons without access to a car, those living in households with one car and those living in households with more than one car. Only in the category with no car we obtain a positive and statistically significant value according to the t_{DMS} test. Full results can be seen in table 1.

TABLE 1. log-likelihood difference statistics for categories with one fixed variable.

Category	$\lvert\mathbf{S}_{\text{pop}}\rvert$	$\overline{\Delta ll_{\text{obs}}}$	$s_{\Delta ll_{\text{obs}}}$	t_{DMS}
no car	15040	0.0113725	0.00418856	6.07
one car	74429	-0.0003288	0.00019308	-3.81
more than one car	55839	0.0000732	0.00065086	0.25
single without children	24266	0.0062789	0.00154739	9.07
couple without children	52373	0.0019977	0.00034802	12.84
single with one child	350	-0.9366935	1.79593373	-1.17
other with children	45541	0.0036872	0.00071807	11.48
other without children	22778	0.0031211	0.00166672	4.19
full-time employed	48612	0.0062745	0.00163108	8.60
part-time employed	20824	0.0123107	0.00672594	4.09
unemployed	4930	0.0085079	0.01078326	1.76
student or apprentice	20701	0.0022936	0.00253132	2.03
retired	39305	0.0007736	0.00077879	2.22
home keeper	10936	0.0227095	0.00904428	5.61
large city	79592	-0.0001245	0.00038159	-0.73
medium sized city	49348	-0.0000351	0.00065107	-0.12
rural area	16368	-0.0003050	0.00236645	-0.29
male	69639	0.0036981	0.00060844	13.59
female	75669	0.0031526	0.00136931	5.15

In accordance with these results and to facilitate interpretability, the following categories were included in the subsequent models as dummy variables for the ASCs: no car, single without children, couple without children, other without children, full-time employed, part-time employed, student or apprentice, retired, home keeper, and female. This leaves the group of unemployed males living in a household with children and at least one car as the reference category.

The region seems to play no role in the motif choice and is hence omitted in the following.

4. Comparison of Models

To compare the performance of the MNL, MMNL and MNP models, a K-fold out-of-sample prediction performance comparison is used, similarly to section 3 but with $K = 3$.

For reasons of computational feasibility, a subset of the data set with data from the year 2013 was used for this part. Weekends and days with reported abnormalities were again excluded. This subset contains information on $2,285$ decision makers with a total of $10,515$ decisions.

The models were fitted three times on two-thirds of the data set ($7,010$ observations) using a maximum likelihood approach and the average log-likelihood per observation of the remaining third ($3,505$ observations) used as a measure of the out-of-sample performance.

The out of sample prediction accuracy and the corresponding required computation time are then used to compare the different models.

In order to facilitate a comparison between the logit and the MNP models we rely on the comparison of relative gains as in Mc Fadden's Pseudo-R^2. The reason for using a different criterion than the likelihood is that MNP models are estimated using a different criterion and hence the corresponding values are not directly comparable. Let Q denote the objective function being maximized, then

$$R^2_{RG} = \frac{Q_{fit} - Q_0}{Q_{max} - Q_0} = 1 - \frac{Q_{max} - Q_{fit}}{Q_{max} - Q_0},$$

where Q_{fit} is the function value for the fitted model, Q_{max} is the best possible function value and Q_0 is a baseline model. The objective function for the MNL and MMNL model are the negative log-likelihoods and the objective function of the MNP is the negative logarithm of the CML. Those objective functions share the property that Q_{max} is equal to zero, which corresponds to the chosen alternatives all having a predicted probability of one. As a suitable baseline model we pick the 'coin flip' model where all linear coefficients of the utility functions are set to zero and hence utility differences are non-existent. Correspondingly, the variances are fixed at one and for the MNP model the off-diagonal entries of Σ are zero. Therefore, for each out-of-sample fold we consider

$$R^2_{RG} = 1 - \frac{Q_{fit}}{Q_0},$$

and additionally compare the out-of-sample likelihood values for the logit models.

The results of this comparison can be seen in table 2.

Table 2. Results of out-of-sample prediction and computation time performance comparison for MNL, MMNL, and MNP with MACML.

	MNL	MMNL	MNP with MACML
required time (h)	2.37	5.09	14.92
$R^2_{\text{obs}_1}$	0.2080	0.2194	0.4904
$R^2_{\text{obs}_2}$	0.2174	0.2348	0.2697*
$R^2_{\text{obs}_3}$	0.2105	0.2299	0.2868*
$\overline{R^2_{\text{obs}}}$	0.2120	0.2280	–

The best model corresponding to this cross-validation procedure is then selected for the subsequent analysis. The results for MNP show that the higher model complexity leads to longer computation times, the estimation takes three times longer than for MMNL when choosing 100 Halton draws. Furthermore, two of the three folds for MNP (marked with the star) show signs of non-convergence because the BFGS-algorithm, which was used to optimize the objective function, terminated because it failed to locate a point lower than the current estimate.[6] Usually the termination criterion should be that the relative gradient is close to zero. This also explains the difference between the first relative gain, which is two times larger than for MMNL, and fold two and three which show a performance which is only on par with MMNL. Re-estimating these two folds using the estimated parameters of the successfully converged model as initial values gives Pseudo R^2 values of 0.4753 and 0.4910 respectively. However, the required processing time to estimate these two models was 23.68 hours. This highlights how important good starting values are for MACML models.

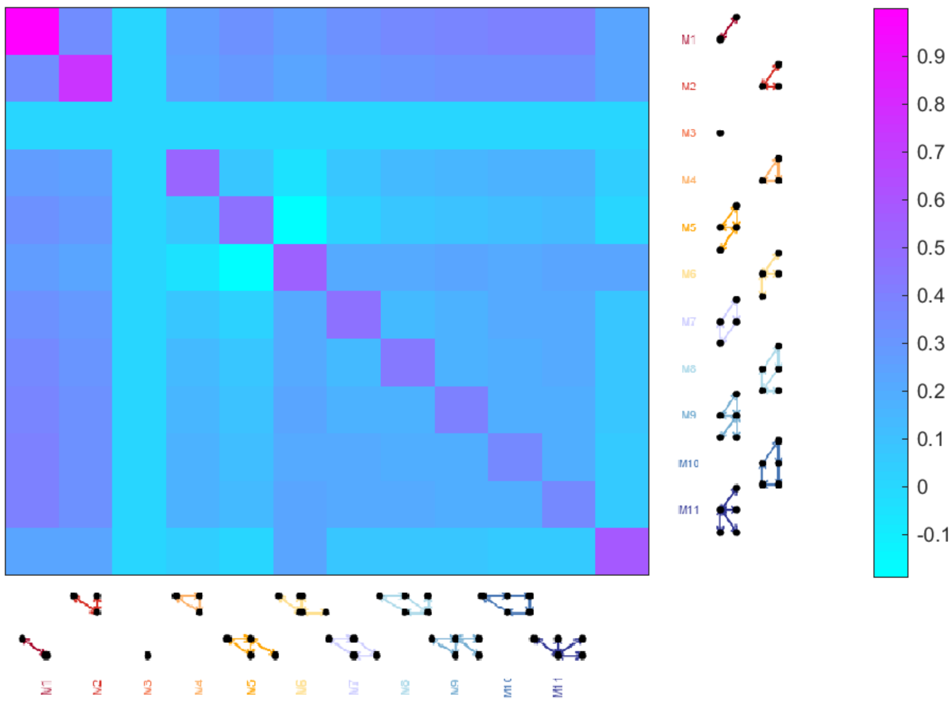
To provide some information on the reasons for the better performance of the MNP model consider figure 3 which provides a plot of the estimated covariance matrix Σ for the first fold. Note that here

[6]This corresponds to termination code 3 of the nlm package.

the variance of the first motif is set to equal 1, the third motif is the reference and hence the covariance is zero. The last motif corresponds to all other. It can be seen that the variance of the error for the various motifs varies between 0.35 (for motif $M11$) and 0.75 (for motif $M2$). Almost all covariances are positive, only the covariance between $M5$ and $M6$ is substantially negative. The positive covariances lead to substantial correlations of approximately 0.5.

These results show that the better performance of the MNP model is due to the ability to capture correlations within motifs as detected in [13].

FIGURE 3. Estimated error covariance matrix Σ from the first estimated MNP model.



5. Temporal Evolution of the Best Fitting Model

Finally, the temporal stability of the motif choice is investigated. For this, firstly, the choice frequency of the 11 most common motifs is determined for each of the 20 years and compared to show the stability of the choice frequencies. The overall stability of the choice frequencies can be seen in figure 2.

Secondly, the most promising model is used to analyze the temporal stability of the choice determinants. To do this, the model is fitted to each of the 20 years separately. Comparing the estimated parameters β for the different model fits allows the analysis of the temporal stability of the determinants of the motif choice.

From the predictive performance alone, according to the results shown in table 2, we should use the MNP model since it outperforms the MNL and MMNL by far. However, due to the greatly increased computation time and due to the issues with non-convergence, the MMNL model was used instead. Thus an MMNL model including ASCs (treated as random with independent normal distribution for each ASC), dummies for the various categories identified above as well as linear, quadratic and cubic age as a regressor is estimated for each year separately. The choice probabilities are thus calculated according to equation 1 with

$$X'_{nti}\tilde{\beta} = \tilde{\beta}_{ASC,ni} + x_{nti,\text{no car}}\beta_{\text{no car},i} + x_{nti,\text{single}}\beta_{\text{single},i} + \cdots + x_{nti,\text{female}}\beta_{\text{female},i},$$

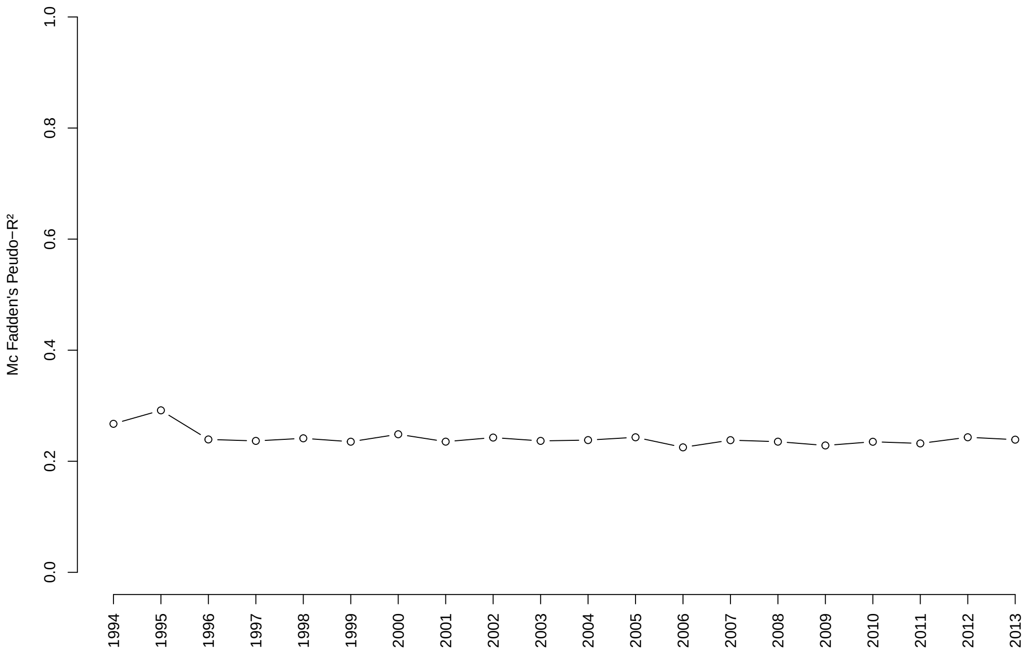
where $\tilde{\beta}_{ASC,ni} \overset{iid}{\sim} \mathcal{N}\left(\beta_{ASC,i,\text{mean}}, \beta_{ASC,i,\text{sd}}\right)$.

It should be noted, that for the year 1995 there was no observation in the data set with household type 2, couples without children. This means that the corresponding parameter can not be identified and the model is not identifiable. Results for this year are therefore to be considered carefully.

To highlight the stability of the model, figure 4 shows the Mc Fadden's Pseudo-R^2 values for each of the years 1994 to 2013. Two things should be noted here. First, the Pseudo-R^2 values are very stable over the years varying between 0.2250 and 0.2486 with only the values for 1994 and 1995 being slightly higher with 0.2674 and 0.2916 respectively.

Second, the explanatory power of the models is rather modest, indicating that motif choice is very stable and only in a small part driven by socio-demographic factors.

FIGURE 4. Mc Fadden's Pseudo-R^2 values for MMNL from 1994 to 2013.



The estimated mean of the ASCs for the MMNL models can be seen in figure 5.

The means of the ASCs are mostly stable over time. There are some fluctuations for the less frequent alternatives, including a major decline for the alternative $M11$ in the year 2002. However, this is mitigated by an increase in the occupation related parameter estimates for $M11$. As seen in the example considering full-time workers shown in figure 6. When adding up the means of the ASCs and the full-time occupation variable, one can see that the dents almost all cancel out, so the resulting utility for the different occupation groups stay stable (see figure 7). These changes in the parameters can be explained by variations in the reference group of unemployed males in a household with children and at least one car, which is rather small.

From figure 6 one can not only infer the stability of the motif choice but also how full-time employed people move during the day, compared to unemployed people. Since the motif $M3$, the 'stay-at-home' motif, is the reference motif all positive parameter values referring to another motif indicate a preference for this motif compared to the 'stay-at-home' motif.

As can be seen in figure 6 most estimated parameter values are greater than zero, indicating a higher utility when traveling.

Motif $M11$ has negative estimated parameter values in 14 of the 16 years for the group of full-time employed people, indicating that full-time employed people are less likely to choose this complex hub-motif compared to unemployed people. It could be reasoned that people with full-employment do not have the time for many extra trips during the day, going back and forth to one place.

As mentioned previously, even though controlling for occupation status and household type, one can see in figures 8(A) to 8(C) that the age still has a significant effect on motif choice, up to cubic dependency. The estimated parameters are mostly also stable over time. Note that the parameter for squared dependency has usually the opposite sign of the linear and the cubic parameter.

The effect of the age on the utility of the different motifs can be seen in figure 8(D). It can be observed that the utility of the simple motifs decreases with age and later rises slightly again. The utility

FIGURE 5. Coefficient estimates for the ASCs in the years 1994 to 2013. A solid square indicates that the coefficient is significant at the 0.05 level.

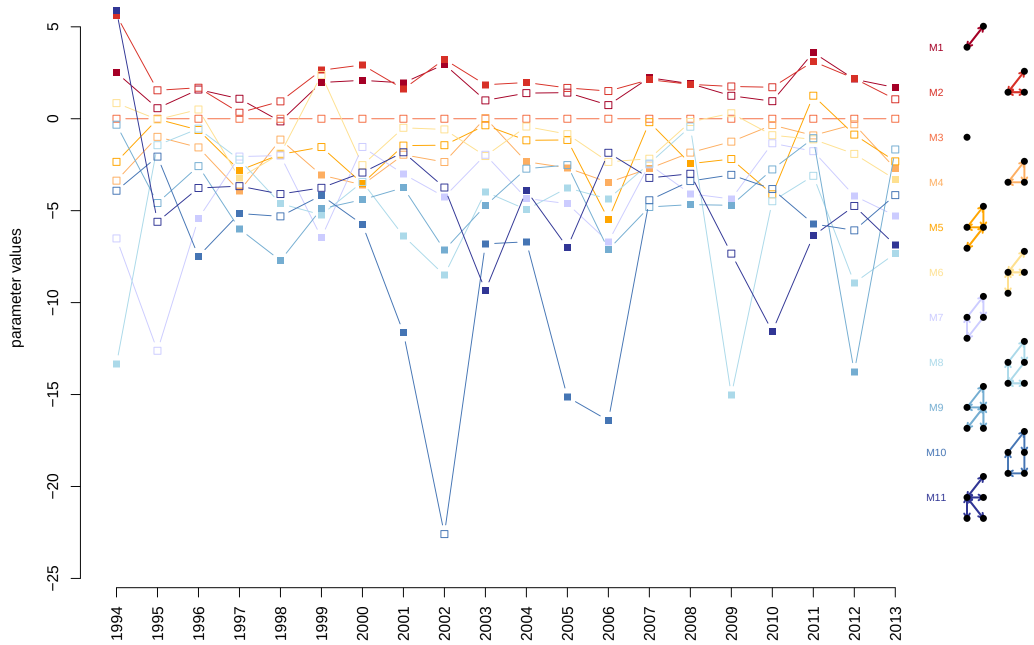
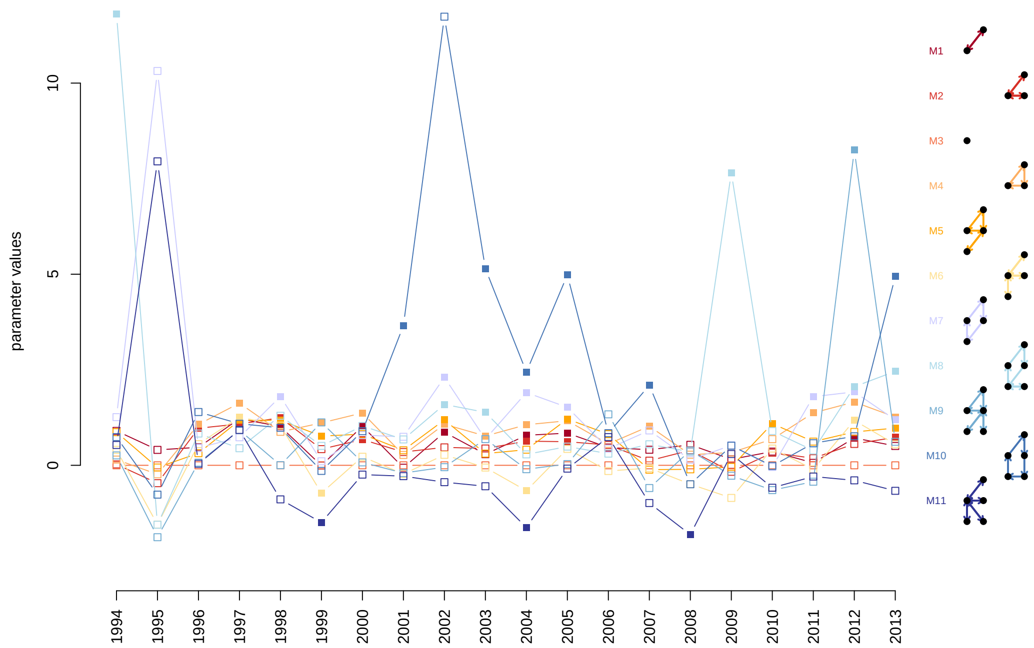


FIGURE 6. Coefficient estimates for the full-time occupied group in the years 1994 to 2013. A solid square indicates that the coefficient is significant at the 0.05 level.



of the more complex motifs increases with age till about the age of 40 and decreases afterwards again. This indicates that people become more active and travel more until they reach a peak mobility age of just under 40 years. The motif $M8$ is a bit of an outlier in this plot, since the utility increases again after the age of 70.

Overall, the results demonstrate the stability of the estimated parameters over the 20 years.

Figure 7. Coefficient estimates for the full-time occupied group plus ASCs in the years 1994 to 2013. A solid square indicates that the occupation coefficient is significant at the 0.05 level.

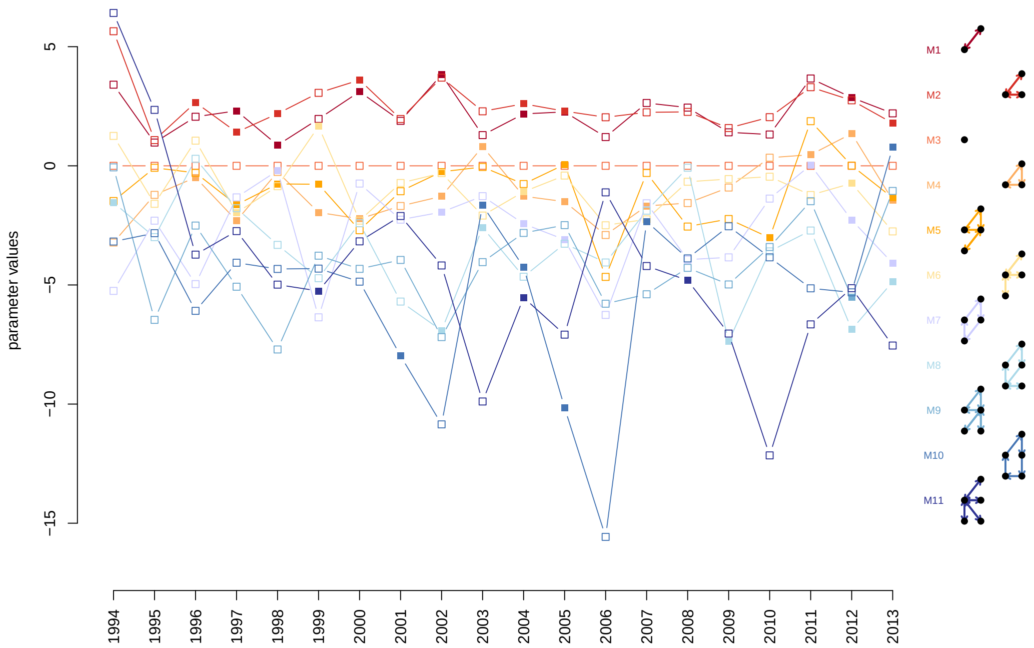
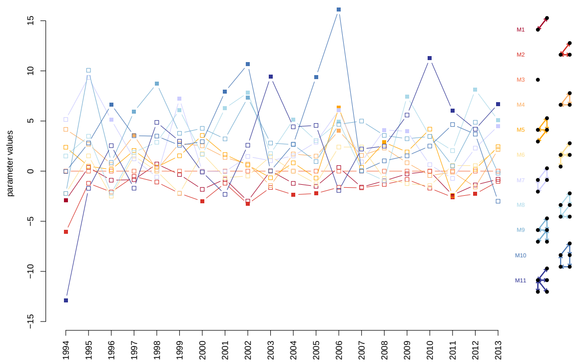
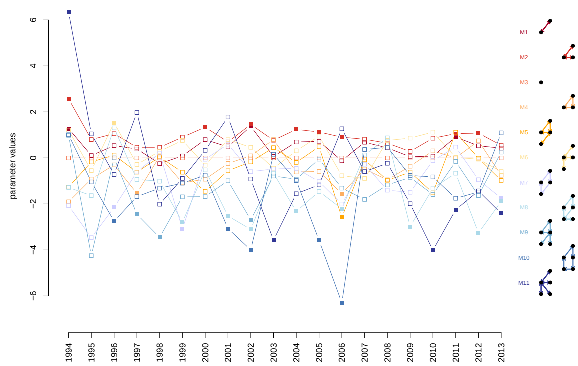


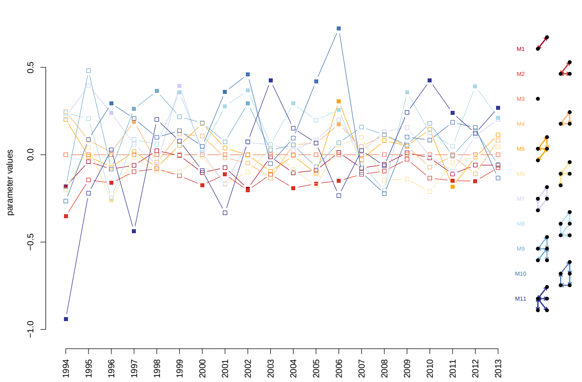
Figure 8. Age related parameter estimates from 1994 to 2013.



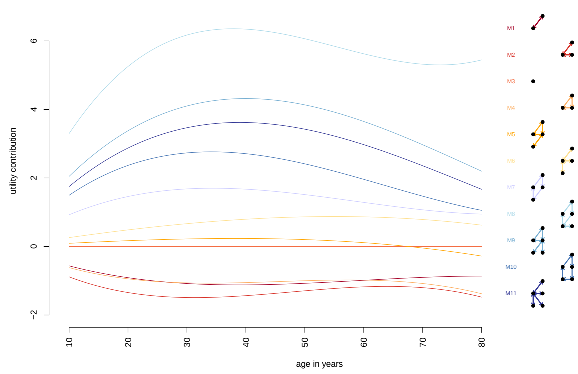
(a) Linear dependency on age



(b) Quadratic dependency on age



(c) Cubic dependency on age



(d) Age dependent utility for the year 2012

6. Conclusions

The results in this paper show several things.

Firstly, the results in section 5 highlight the stability of motif choice. This is in line with the findings of [13, 14, 17] who established spatial stability and independence from different data sources. In addition, we showed temporal stability of the motif choice frequency and the choice determinants.

In section 4 we were able to show that, as can be expected for a more flexible model, MMNL models provide a better prediction accuracy than MNL models. We were further able to show, that mixed MNP models are outperforming MMNL models in terms of prediction accuracy at least for the chosen formulation of independent random ASCs while the MNP model allows for correlations that appear to be present in the data. However, with the MACML formulation used, the model estimation is still too time consuming for evaluating all years and has convergence issues when the starting values are poorly chosen. This indicates great potential for the mixed MNP models for motif choice prediction. It however requires some more work on the computation time and starting values.

Once the MACML procedure can compete with MMNL in terms of computation time for this application, one should also compare the predictive performance of MNP with MMNL models where the parameters have a joint distribution with fully general covariance matrix. To make sure that the more complex models are still estimable, the data set used should be increased such that it includes enough observations for every parameter. In this respect merging some of the groups can be an option, i.e. if the groups do not differ in terms of choice determinants significantly.

With all estimated models of this paper the predictive performance of motif choice is, however, still quite low, indicating that motif choice is only modestly affected by socio-demographics. This indicates that models of motif choice can be transferred from one population to another one without losing much in terms of predictive accuracy.

Acknowledgements

This work has been financed partly by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 356500581 which is gratefully acknowledged. The data set has been made available by the German 'Clearingstelle für Verkehr' run by the DLR and has been collected by the Karlsruhe Institute of Technology which is gratefully acknowledged.

References

[1] M Batram and D Bauer. On consistency of the macml approach to discrete choice modelling. Journal of Choice Modelling, 30:1–16, 2019.

[2] C. R. Bhat. The maximum approximate composite marginal likelihood (macml) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B, 45(7):923–939, 2011.

[3] C. R. Bhat. The composite marginal likelihood (cml) inference approach with applications to discrete and mixed dependent variable models. Foundations and Trends in Econometrics, 7(1), 2014.

[4] C. R. Bhat and R. Sidhartan. A simulation evaluation of the maximum approximate composite marginal likelihood (macml) estimator for mixed multinomial probit models. Transportation Research Part B, 45(7):940–953, 2011.

[5] S. Büscher, M. Batram, and D. Bauer. Using motifs for population synthesis in multi-agent mobility simulation models. In to appear in Proceedings of the workshop SMSA2019: Stochastic Models, Statistics and Their Applications, Springer Nature, 2019.

[6] E Cascetta. Transportation Systems Analysis. Springer Verlag, New York, 2 edition, 2009.

[7] F. X Diebold and R. S. Mariano. Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1):134–144, 2002.

[8] S. Hess and D. Palma. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. Journal of Choice Modelling, 32:100170, 2019.

[9] C. Heyde. Quasi-Likelihood And Its Application. Springer Verlag, 1997.

[10] A. Horni, K. Nagel, and K. W. Axhausen. The multi-agent transport simulation MATSim. Ubiquity Press London., 2016.

[11] H Joe. Approximations to multivariate normal rectangle probabilities based on conditional expectations. Journal of the American Statistical Association, 90(431):957–964, 1995.

[12] J. Ortúzar and L. G. Willumsen. Modelling transport. John Wiley and Sons, 2001.

[13] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González. Unravelling daily human mobility motifs. Journal of The Royal Society Interface, 10(84):1–8, 2013.

[14] C. M. Schneider, C. Rudloff, D. Bauer, and M. C. González. Daily travel behavior: Lessons from a week-long survey for the extraction of human mobility motifs related information. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. ACM, 2013.

[15] K. E. Train. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, 2 edition, 2009.

[16] C. Varin and P. Vidoni. Pairwise likelihood inference for ordinal categorical time series. Computational Statistics & Data Analysis, 51:2365–2373, 2006.

[17] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. C. González. Discovering urban activity patterns in cell phone data. Transportation, 42(4):597–623, Jul 2015.

[18] Y Zhao and H. Joe. Composite likelihood estimation in multivariate data analysis. The Canadian Journal of Statistics, 33(3):335–356, 2005.

[19] D. Zumkeller and B. Chlond. Dynamics of change: Fifteen-year german mobility panel. TRB 88th Annual Meeting Compendium of Papers, 2009.