Household-level choice-set generation and parameter estimation in activity-based models

Negar Rezvany*[1], Tim Hillel[2], and Michel Bierlaire[3]

[1]Doctoral assistant, School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

[2]Lecturer, Department of Civil, Environmental and Geomatic Engineering (CEGE), University College London (UCL), United Kingdom

[3]Professor, School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Short summary

Traditional Activity-based models (ABMs) treat individuals as isolated entities, limiting behavioural representation. Econometric ABMs assume agents schedule activities to maximise utility, explained through discrete choices. Using discrete choice models implies the need for calibration of maximum likelihood estimators of the parameters of utility functions. However, classical data sources like travel diaries only contain chosen alternatives, not full choice set, making parameter estimation challenging due to unobservable, and combinatorial activity spatio-temporal sequence. To address this, we propose a choice-set generation framework for household activity scheduling, to estimate significant and meaningful parameters. Our methodology adopts a Metropolis-Hastings sampling approach, and extends it to encompass parallel generation for all household agents, household-level choices, and time arrangements. Utilising this approach, we then estimate parameters of household-level scheduling model presented in (Rezvany et al., 2023). This approach aims to generate behaviourally sensible parameter estimates, estimated on ensemble of schedules with consistent alternatives for household members, enhancing model realism in capturing household dynamics.
Keywords: Activity-based modelling, Choice-set generation, Discrete choice modelling, Intra-household interactions.

1 Introduction

Motivation and scope

ABMs represent travel demand driven by spatially and temporally distributed activities, incorporating more behavioural realism compared to traditional trip-based models. This approach has been of interest to modellers and analysts in different domains such as transportation and energy research. Individuals do not plan their day in isolation from other members of the household. Their decision-making involves considering the activities and schedules of other household members and sometimes individuals in their social network. Various interactions, time arrangements, and constraints affect individuals' activity schedules. However, most ABMs do not consider these household dynamics. Hence, models dealing with individual choices need to be revisited to take account of the intra-household interactions.

ABMs research encompasses rule-based computational process models and econometric models. The latter assumes that individuals schedule activities to maximise utility, explained through discrete choices using advanced econometric methods. Nevertheless, these models confront challenges in accurately estimating parameters.

Consistent estimation of parameters requires behavioural data records on hypothetical or unseen situations in addition to the chosen alternative (revealed preference), which are not all necessarily observable and not available in classical data sources such as travel diary surveys or time use data. Moreover, the derivation of choice probabilities and likelihood functions requires the modeller to assume a universal choice set which is finite and enumerable. However, the full choice set of possible activities and their spatio-temporal sequence is combinatorial and cannot be enumerated, while individuals are only aware of a fraction of the full choice set. Therefore, exploring and operationalising appropriate choice set generation techniques is another challenge.

Choice set generation technique using a Metropolis-Hastings based sampling algorithm can be a smart move to strategically sample alternatives, to calibrate econometric activity-based models. This functionality adopts the Metropolis-Hastings based sampling algorithm introduced by Pougala et al. (2021). As intra-household interactions cause additional choice dimensions, time arrangements, constraints, and group decision-making mechanism, the interactions should be considered in the choice set formation to ensure consistency of generated alternatives.

In this paper, we present a choice set generation framework for household activity scheduling, generating an ensemble of schedules with consistent alternatives for all household members. To explore the combinatorial solution space of full set of feasible schedules, we adopt the Metropolis-Hastings based sampling algorithm introduced by Pougala et al. (2021) Necessary considerations in household choice set generation is noted. Utilizing the choice set generation technique, the parameters of a utility-based household scheduling model presented in (Rezvany et al., 2023), the household-level Optimisation-based Activity Scheduling Integrating Simultaneous choice dimensions (OASIS), is estimated. The results and behavioural implications are then discussed.

The remainder of this manuscript is structured as follows. Section 2 discusses household-level choice set generation methodology. Section 3 presents a practical application. Concluding remarks and future research are presented in Section 4.

Relevant literature

ABMs traditionally centered on individual decision-making, often fail to capture the interdependencies between household members. This oversight leads to biased simulations of activity-travel schedules, as household members' schedules are interdependent. Addressing this, earlier we have proposed an operational utility-based scheduling framework, capturing multiple intra-household interactions within a single ABMs, accommodating complex interactions such as allocation of private vehicle to household members, escort duties, joint participation in activities, and sharing rides (Rezvany et al., 2023).

Model calibration in utility-based ABMs is challenging due to limited data in traditional surveys like travel diaries, which focus on revealed preferences without illuminating the complete choice set of alternatives. The choice set of alternatives is typically latent or unobservable to the analyst. Defining a choice set representative of activity-travel patterns in household activity pattern problem is thus, necessary for operationalising household random utility models.

Xu et al. (2017) develop a choice set generation technique for Household activity pattern problem (HAPP) (Recker, 1995) using a clustering approach developed by Allahviranloo et al. (2014) to identify representative patterns, optimised for information gain. Shakeel et al. (2022) model potential joint leisure activities within households using a latent class model, underscoring the need for further research in joint activity generation and integration into operational activity-based models.

Applying Metropolis-Hastings algorithm to sample alternatives in an activity-based context has been explored in the literature (Pougala et al., 2021; Danalet & Bierlaire, 2015). Considering their promising results, we explore this approach to expand it to a household-level choice set generation in ABMs.

2 Methodology

We propose a household-level choice set generation technique to estimate the parameters of the utility-based household scheduling model presented in Rezvany et al. (2023). For explanation and formulations of the household-level scheduling framework, we refer the reader to (Rezvany et al., 2023). To explore the combinatorial solution space of full set of feasible schedules, a Metropolis-Hastings algorithm is used. This functionality adopts the Metropolis-Hastings based sampling algorithm introduced by Pougala et al. (2021). In the remainder of this section, we first give a brief synopsis of the base Metropolis-Hastings based sampling strategy and then present the household-level choice set formation framework for model estimation.

Base Metropolis-Hastings based sampling strategy: A brief synopsis

This is a strategy to generate a choice set containing only feasible alternatives that can be used for estimating parameters of a utility-based activity-based model. The alternatives for each individual are full daily schedules. Using a strategic generation with Metropolis-Hastings algorithm, it generates an ensemble of high probability schedules, to estimate significant and meaningful parameters, while still containing low probability alternatives to decrease the model bias. The choice set generation is modelled as a Markov process. The algorithm is initialised with with a random schedule (e.g. the reported schedule in the diary dataset can be used as the initial state). States are defined as daily schedules with choice dimensions such as activity participation, timings, location, and transportation mode. The choice set is generated by exploring the neighbouring schedules of each state using operators with a known probability, and accept or reject the change based on an acceptance probability defined by the modeller. Operators are heuristics that modify specific aspects of the schedule and can be created according to the modeller's needs and specifications. Block, Assign, Swap, and Anchor are example operators, which their description can be found in (Pougala et al., 2021). A set of validity constraints should be checked for the generated states to ensure choice set only contains feasible schedules. The process is carried until the defined Markov chain reaches stationarity.

In order to obtain unbiased parameters using a smaller subset of alternatives, an alternative specific correction term is added in the choice probabilities. C_n is the generated choice set for individual n. Thus, the probability that an individual n chooses alternative $i_n \in C_n$, associated with a deterministic utility V_{in}, is defined as follows (Ben-Akiva & Lerman, 1985):

$$P(i_n|C_n) = \frac{\exp\left[V_{in} + \ln q(C_n|i_n)\right]}{\sum_{j \in C_n} \exp\left[V_{jn} + \ln q(C_n|j)\right]} \tag{1}$$

in which $lnq(C_n|i_n)$ is the alternative specific correction term defined as:

$$q(C_n|i_n) = \frac{1}{q_{in}} \prod_{j \in C_n} \left(\sum_{j \in C_n} q_{jn} \right)^{J+1-\hat{J}} \tag{2}$$

where C_n is choice set of size $J+1$ with $\hat{J}$ unique alternatives for individual n. j represents alternative sampled from the target distribution of the Metropolis-Hastings algorithm with probability q_{jn}. This formulation implies that if all alternatives have equal selection probabilities, the estimation on the subset is the same as the estimation on the full set of alternatives.

A detailed explanation of the Metropolis-Hastings sampling strategy can be found in (Pougala et al., 2021).

Household-level choice set generation and parameter estimation

Intra-household interactions affect how members schedule their day. Causing additional choice dimensions, time arrangements, constraints, and group decision-making mechanism which should be considered in the generated choice set for more behaviourally realistic estimations. The main aspects are summarised as follows.

Firstly, the choice set of all individuals in a household should be generated in parallel, as they are inter-related. This is a key matter in household-level choice-set formation differing it from the individual-level approach. The relation between individuals and their household is lost in individual-level choice-set formations, leading to separate choice set formation procedures with no feedback between them. In our framework, as an output, a sample containing clusters of schedules for all individuals in a household is generated, satisfying intra-household validity constraints. This approach leads to more complexities in the generation model. An individual is selected in each household as index. At each step of the random walk, their combinatorial solution space is explored using the Metropolis-Hastings algorithm. Their state is then used as the benchmark for ensuring schedule synchronisations with other agents in the household. This ensures compatibility between the generated schedules for individuals in the household at each step.

Secondly, in terms of capturing household interactions, we move from individual utility function to household utility function where the individual sub-utilities are aggregated. The household utility is defined as the weighted sum of the utility that each agent n in the household of size N_m gains from her/his schedule over the considered time period. The weights w_n, capture the relative "power" of each individual in the household-oriented decisions.

$$HUF = \sum_{n=1}^{n=N_m} w_n \, U_n \tag{3}$$

Moreover, we ensure that the possible interaction aspects are captured in the utility function such as terms to capture motivations for joint engagements and (dis)utility of doing an escorting task.

Third, the operators must generate a state that meets the household-level constraints, as well as individual-level schedule validity constraints. As the within-household interactions lead to additional and more complex constraints, these interplays must be also accounted for in the generated choice set. Resource constraints, sharing household maintenance responsibilities, joint activity participation, joint travels, and escorting are examples of intra-household interactions.

Furthermore, new operators are introduced to modify choice dimension aspects related to household scheduling, such as activity participation mode. We define *partic_ mode* operator $\omega_{\text{partic\_mode}}$ to change the participation mode p_{a_n} of a randomly selected activity a_n for individual n, with a given probability $P_{\text{partic\_mode}}$. The modes are chosen from a set of possible activity participation modes which is considered known (e.g., solo or joint). The selection of a participation mode is done according to distribution $P_\pi(a_n)$ which is conditional on the activity and assumed to be exogenous to the choice-set generation. This operator can only be applied to activities that can have multiple participation modes defined by the modeller.

In case of change in participation mode, the schedule synchronisation among agents in the household is checked and the corresponding activity is planned in the schedule of accompanying member(s) with the same timings and participation mode. To respect validity requirements, the resulting schedule must always start and end at home and the participation mode of home cannot be changed. The transition probability associated to this change is defined as the product of probability of this operator to be chosen multiplied by the probability of choosing a valid block and the probability of selecting one of the possible participation modes. This can be written as follows:

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{partic\_mode}} P_\pi \frac{T-2\delta}{T\delta}, & \text{if } b_i \notin \{b_o, b_T\} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where X_t and X_{t-1} are the state at time t and $t-1$, respectively. Each state corresponds to a schedule within the considered time budget T (e.g., 24 hours) and is discretised into blocks of length δ. b_p is a block at position p, $p = (0, \delta, ..., T-\delta, T)$.

3 Empirical investigation

The data from the 2018 – 2019 UK National Travel Survey (NTS) (Department for Transport, 2022) is used for generating choice sets for households and estimating parameters of the ABMs model (Rezvany et al., 2023). NTS is a household survey containing information on daily trips and socio-economic characteristics of individuals in a household within the UK. Using the Metropolis-Hastings algorithm to generate the choice sets, we estimate the parameters of a sample of schedules for 2-membered households of 2 adults. The 2018 – 2019 version of the data contains $4'802$ individuals, belonging to $2'401$ households of size 2 with 2 adults, and $22'698$ trip diaries.

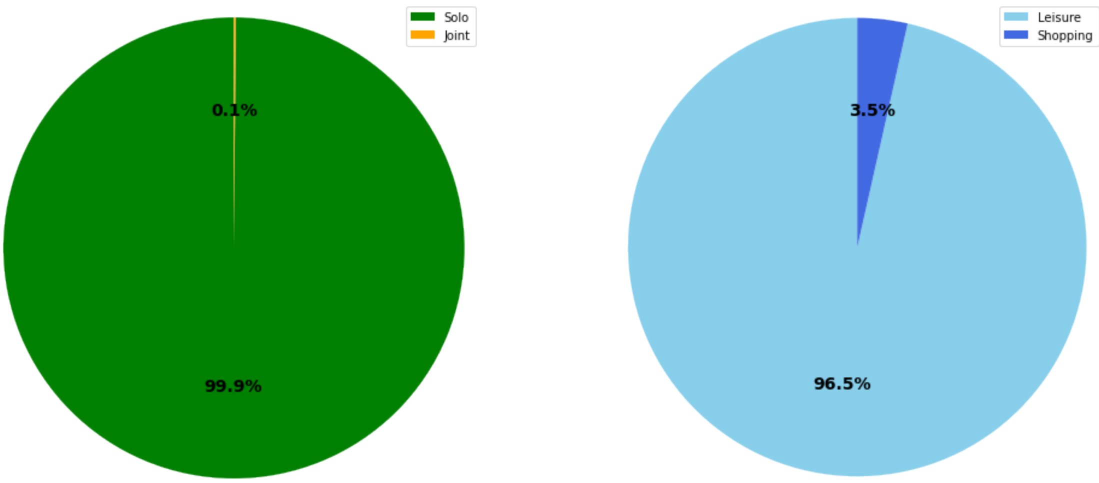
First we process the data to convert the trip diaries to daily activity schedules. Data points with missing information are excluded. For this case study, a sample of schedules for 500 households is used. We group the activities into 6 categories: Home, Work, Education, Leisure, Shopping, and Personal business (eg. eat/drink, using services like medical appointments).

The mean of start times and durations for each activity from the distribution across the households of 2 adults are assumed as indicators for desired start and duration times in the model (Table 1). The scheduling preferences are assumed to be homogeneous across the individuals. For identification purposes, 'Home' is used as reference. Home is interpreted as absence of activity in this study due to absence of information on in-home activities in the dataset, which can be relaxed with richer data containing in-home activities such as time use surveys. For the sake of simplification, travel parameters are not estimated to focus solely on activity scheduling parameters. We generate choice sets of size $N = 100$ for each household using the Metropolis-Hastings algorithm. We consider Block, Assign, Swap, Anchor, Partic_mode, and Combination Meta-operators for the random walk. The operators have equal probability of being selected.

Table 1: Scheduling preferences

Activity	Desired start time [hh:mm]	Desired duration[hh:mm]
Work	09:15	06:55
Education	10:30	5:10
Leisure	12:48	02:50
Shopping	12:35	01:05
Personal business	12:20	01:10

As we study interactions within household members, activity participation modes (solo/joint) are extracted from the data, using a set of rules inspired by Ho & Mulley (2013) for identifying joint participation within household. Analyzing diaries in NTS, we observe that only 0.1% of activities in diaries are performed jointly (Figure 1a). Among which Leisure activities make a substantial portion (97%) of joint activities (Figure 1b). Thus, we only consider Leisure activities to have the possibility to be done jointly in our choice set generation.



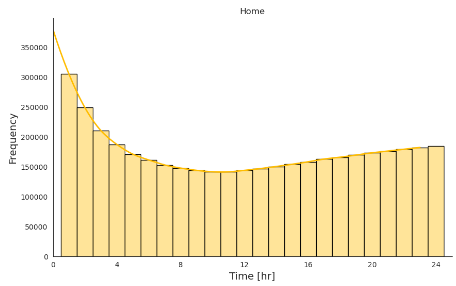
(a) Share of activity participation modes

(b) Share of activity types with joint participation

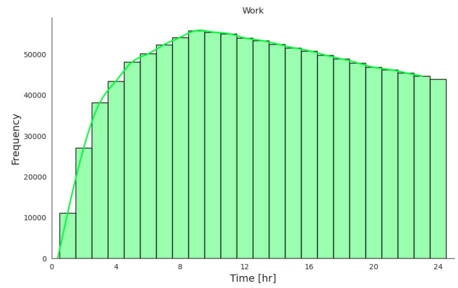
Figure 1: Analysis of activity participation modes in NTS data

Generated choice set: analysis and discussions

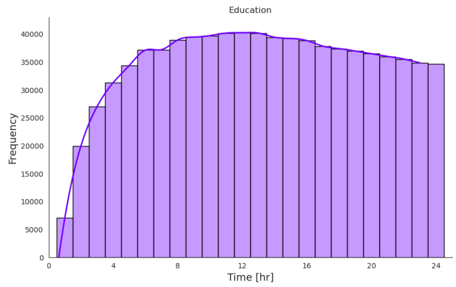
We run 1000 iterations of the algorithm for a sample of 500 households of 2 adults, generating choice sets of sizes $N = 100$ alternatives for each household. The accepted schedules are sampled after a warm-up period. Figure 2 depicts the distribution of activity participation across different hours of the day for each activity type in the generated sample. The distributions are sensible according to



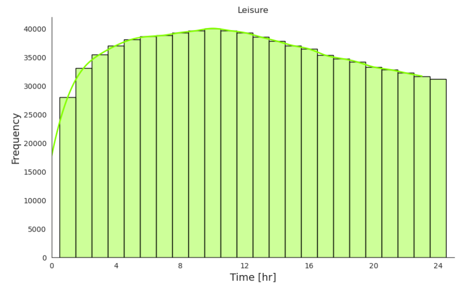
(a)



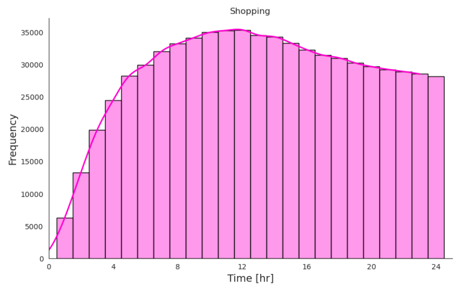
(b)



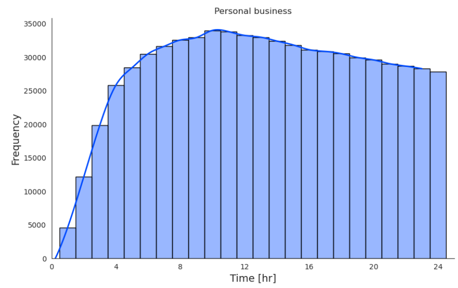
(c)



(d)



(e)



(f)

Figure 2: Distribution of activity participation across different hours of day in generated choice sets

expectations. Home activity has a peak at midnight which aligns with the common resting period. It declines sharply as people typically begin their day and participate in out-of-home activities, with a gradual increase towards the evening suggesting return to home after the daily activities. Figure 2b indicates distinct peak activity times for work with concentrated density during typical office hours. Leisure have a more spread-out pattern, reflecting more scheduling flexibility and less constrained feasible activity hours throughout the day.

We take a closer look at the impact of each operator on acceptance of a generated schedule. Figure 3 illustrates the proportion of operators among accepted moves. In this case study, Block and Assign are the most frequently used operators, followed by Swap and Partic_mode. The Block operator only modifies the time discretisation, which thus does not affect the schedule validity, leading to feasible new states. Moreover, frequent application of Assign operator is also expected. Adding activities is favourable in terms of utility gain as the constant utility gain for participating in activities is usually larger in scale than the penalties for schedule deviations. The lower proportion of use for partic_mode operator makes sense. When changing the activity participation mode, household schedule synchronisation validity constraints can be obviated. Thus, partic_mode operator has a lower proportion in accepted schedules. Figure 4 shows the typology of combination of Meta-operators in accepted schedules, illustrating the prevalence of each operator in the accepted Meta-operator combinations (Figure 4b) and the distribution of their lengths (Figure 4a).

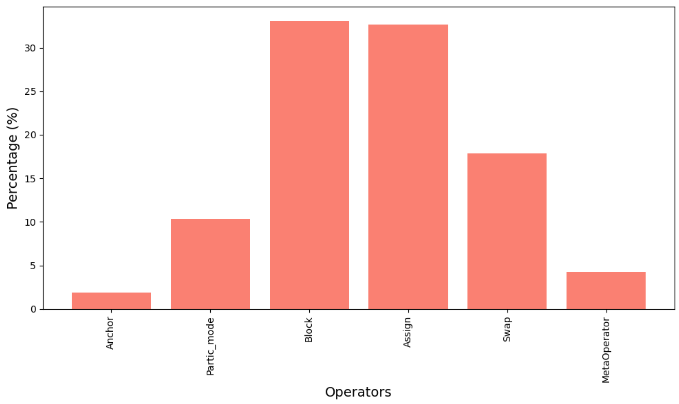
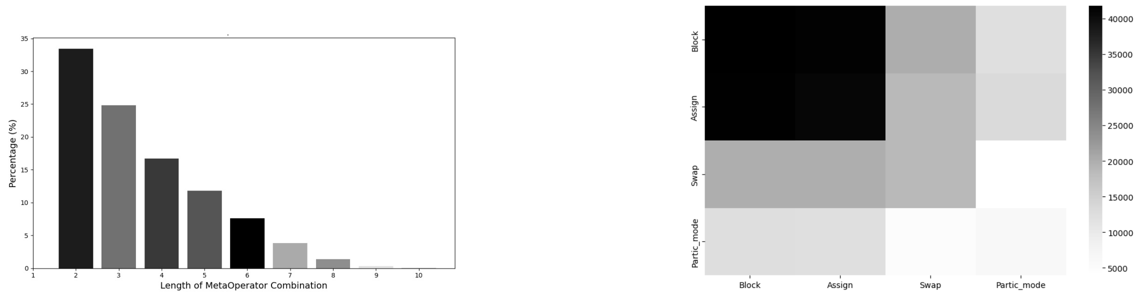


Figure 3: Frequency of accepted operator changes



(a) Lengths of combinations for MetaOperators

(b) Frequency of pairs in Meta-operators

Figure 4: Typology of accepted combinations of Meta-operators

Estimation results and discussions

Using the generated choice set, the scheduling model has been estimated. The attributes used in the model are related to the activity-specific constants and parameters, as well as scheduling deviation penalties. Table 2 summarises the estimation results.

The estimated parameters are behaviourally sensible. The activity-specific constants are all positive, indicating a baseline preference for doing an out-of-home activity rather than staying at home, all else being equal. Shopping activities bring the most utility per time unit followed by Personal business, Work, and Leisure activities. Most of the parameter estimates are statistically significant ($p - value < 0.05$). The estimates with 0 p-value are indicative of parameters that are highly statistically significant predictors in the model. However, there are example parameters that are not statistically significant such as parameters associated with duration of leisure. This can indicate that leisure is not a particularly time constraining activity, in the sense that it is less likely to trigger trade-offs in the scheduling process compared to other activities.

The estimated joint participation parameter for leisure is significant and positive. This indicates that doing leisure activities with other household member(s) is strongly preferred, highlighting the social aspect of leisure time. Joint participation in activities can be motivated by considerations such as (i) efficiency; which can be gained from time and/or money savings, (ii) altruism, which is a selfless regard in which an individual gains utility by benefiting someone other than oneself, and (iii) companionship.

The penalty parameters have a negative sign, indicating a decline in utility when deviating from their preference. For example the significant negative coefficient for shopping later than preferred suggests individuals find less utility in shopping activities that occur later than their preferred timing, possibly due to increased crowds, reduced availability of items, or personal schedule constraints. Shorter durations than expected are penalised about 3 times more than longer for work activity. The negative and significant estimate for shorter work activities than preferred may reflect

Table 2: Estimation results

Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
Leisure: ASC	2.26	0.0874	8.71	0
Leisure: joint_partic	0.259	8.71	-1.84	0
Leisure: early	-0.778	0.0874	-8.9	0
Leisure: late	-0.737	0.0857	-8.6	0
Leisure: long	0.0095	0.0227	-0.416	0.677*
Leisure: short	-0.14	0.216	0.648	0.517*
Personal business: ASC	4.8	0.682	7.03	2.01e-12
Personal business: early	-0.96	0.113	-8.51	0
Personal business: late	-0.775	0.0977	-7.93	2.22e-15
Personal business: long	-0.547	0.165	-3.31	0.000944
Personal business: short	-1.5	0.507	-2.95	0.00316
Shopping: ASC	7.45	0.944	7.89	2.89e-15
Shopping: early	-1.23	0.166	-7.43	1.09e-13
Shopping: late	-0.697	0.0927	-7.52	5.28e-14
Shopping: long	-0.803	0.165	-4.88	1.08e-06
Shopping: short	-3.43	0.789	-4.35	1.36e-05
Education: ASC	1.38	1.07	1.29	8.15e-04
Education: early	-2.36	0.58	-4.06	3.02e-02
Education: late	-0.399	0.174	-2.29	4.24e-02
Education: long	-2.44	0.989	-2.47	1.44e-03
Education: short	-1.52	0.257	-5.88	1.36e-05
Work: ASC	4.28	0.476	8.99	0
Work: early	-0.828	0.108	-7.68	1.58e-14
Work: late	-0.45	0.0975	-4.62	3.92e-06
Work: long	-0.272	0.0438	-6.22	5.03e-10
Work: short	-0.828	0.13	-6.39	1.7e-10

Summary of statistics

$L(0)$ = -282.4367

$L(\hat{\beta})$ = -199.3681

* Not statistically significant at 95%

the disutility associated with not fulfilling expected work hours, which could impact productivity or income. Furthermore, the improvement in log-likelihood from null log-likelihood signifies that the model's estimated parameters provide a better fit to the observed choices than a model without predictors.

4 Conclusions

This paper discusses implementation requirements for ABMs with intra-household interactions and presents a household-level choice set generation. We build on the Metropolis-Hastings based sampling algorithm developed by Pougala et al. (2021). The important aspects in household choice-set generation can be summarised as: (i) the choice set for individuals in a household are generated in parallel, as they are inter-related, (ii) we move from individual utility function to household utility function, (iii) possible interaction aspects are captured in the utility function. (iv) new operators are introduced to modify choice dimension aspects related to household scheduling, (v) the accepted schedules should remain compliant with household-level constraints, in addition to individual-level validity constraints. This approach can generate an ensemble of high probability schedules, to estimate significant and meaningful parameters, while still containing low probability alternatives to decrease the model bias. Utilising the choice set generation technique, the parameters of a utility-based ABMs, household-level OASIS, (Rezvany et al., 2023) is estimated. The results are both behaviourally sensible and statistically significant.

There are further extensions and improvements of the current work, suggesting paths for future research. The scheduling preferences are assumed to be homogeneous across the sample. Investigating non-homogeneous preferences across individuals can be considered. In the current specification, socio-demographic variations are not considered. In order to investigate more behavioural implications explaining the choice of schedules, socio-demographic variables can be also added in the proposed framework and utility functions. Considering interactions between household demographic variables (e.g. as presence of children, family structure, work characteristics of individuals) and activity participation utility variables is another interesting avenue. Furthermore, exploration of validation techniques can be considered. Validating the approach by estimating parameters with the sampled choice set and measuring their bias with metrics such as the mean absolute error is a possible approach (Lemp & Kockelman, 2012).

References

Allahviranloo, M., Regue, R., & Recker, W. (2014). Pattern Clustering and Activity Inference. In *Transp. res. board 93rd annu. meet.* Washington DC, United States.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis : theory and application to travel demand*. MIT Press.

Danalet, A., & Bierlaire, M. (2015). Strategic sampling for activity path choice. In *15th swiss transp. res. conf.* Ascona, Switzerland.

Department for Transport. (2022). *National Travel Survey, 2002-2021 [data collection].* 16th Edition. UK Data Service. doi: 10.5255/UKDA-SN-5340-12

Ho, C., & Mulley, C. (2013). Tour-based mode choice of joint household travel patterns on weekend and weekday. *Transportation (Amst).*, *40*(4), 789–811. doi: 10.1007/s11116-013-9479-0

Lemp, J. D., & Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. *Transp. Res. Part A Policy Pract.*, *46*(3), 602–613. Retrieved from http://dx.doi.org/10.1016/j.tra.2011.11.004 doi: 10.1016/j.tra.2011.11.004

Pougala, J., Hillel, T., & Bierlaire, M. (2021). Choice set generation for activity-based models. In *Proc. 21st swiss transp. res. conf.* Ascona, Switzerland.

Recker, W. W. (1995). The household activity pattern problem: General formulation and solution. *Transp. Res. Part B*, *29*(1), 61–77. doi: 10.1016/0191-2615(94)00023-S

Rezvany, N., Bierlaire, M., & Hillel, T. (2023). Simulating intra-household interactions for in- and out-of-home activity scheduling. *Transp. Res. Part C Emerg. Technol.*, *157*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0968090X23003522 doi: https://doi.org/10.1016/j.trc.2023.104362

Shakeel, F., Adnan, M., & Bellemans, T. (2022). Joint-Activities Generation among Household Members using a Latent Class Model. *Transp. Res. Procedia*, *62*(2021), 557–564. Retrieved from https://doi.org/10.1016/j.trpro.2022.02.069 doi: 10.1016/j.trpro.2022.02.069

Xu, Z., Kang, J. E., & Chen, R. (2017). A Random Utility Based Estimation Framework for the Household Activity Pattern Problem. *Transp. Res. Procedia*, *23*, 809–826. Retrieved from http://dx.doi.org/10.1016/j.trpro.2017.05.045 doi: 10.1016/j.trpro.2017.05.045