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

Negar Rezvany^{*1}, Tim Hillel², and Michel Bierlaire³

¹Doctoral assistant, School of Architecture, Civil and Environmental Engineering (ENAC),

Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

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

³Professor, 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. Using a Metropolis-Hastings sampling approach, we sample an ensemble containing clusters of schedules for all agents in a household. Alternatives for all household agents are generated in parallel, encompassing 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, Intrahousehold interactions.

1 INTRODUCTION

1.1 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, 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 (MH) based sampling algorithm can be a smart move to strategically sample alternatives, to calibrate econometric ABMs. This functionality adopts the MH based sampling algorithm introduced by Pougala et al. (2021). As intrahousehold 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 MH 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. We first give a brief review of the relevant literature. In Section 2, the household-level choice set generation methodology is explained. Section 3 presents an empirical investigation, applying the methodology on a real-life case study. Concluding remarks and future research are discussed in Section 4.

1.2 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 activity-based model, 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 MH algorithm is used. This functionality adopts the MH based sampling algorithm introduced by Pougala et al. (2021). In the remainder of this section, we first give a brief synopsis of the base MH based sampling strategy. Then we present the household-level choice set formation and model estimation framework.

2.1 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 MH 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 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 Meta-operator can be defined to combine the actions of two or more operators. A set of validity constraints should be checked for the generated states to ensure that the choice set only contains feasible schedules.

A detailed explanation of the MH sampling strategy for ABMs can be found in (Pougala et al., 2021).

2.2 Household-level choice set generation and parameter estimation

2.2.1 Choice set generation

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.

In the household-level choice-set generation technique, the choice set of all agents in a household are generated in parallel. This ensures compatibility between schedules of agents in a household in generated alternatives. The household state at step t, X_t , is household schedule comprised of a cluster of schedules of agents in the household, $[X_{1_t}, \ldots, X_{N_{mt}}]$. The state of each agent n, X_{n_t} , is her/his schedule within the time budget T (e.g. 24 hr), discretised in blocks of duration $\delta \in [\delta_{\min}, 24 - \delta_{\min}]$, where δ_{\min} is the minimum block duration.

The algorithm is initialised with a random household schedule X_0 (e.g. ensemble of reported schedules of all agents in the household). An agent I from the household, is selected as index. The protocol to choose the index person is decided by the modeller (e.g. random selection, rule-based selection based on agent employment type, etc). The combinatorial solution space of the index agent is explored using the MH algorithm.

The candidate state of the index agent is used as the benchmark for ensuring schedule synchronisation with other agents in the household. Solution space of other household agents is explored using the MH technique, ensuring being compliant with household-level, as well as individual-level 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.

The output of the generator is an ensemble containing clusters of schedules for all individuals in a household. The household choice-set formation procedure is summarised in Algorithm 1. It is notable that socio-demographic characteristics of individuals and their household (e.g. household structure, employment characteristics of individuals) are preserved in the choice set generation procedure. The socio-demographic characteristics are captured and included in the generated alternatives in the choice set. This feature prevents information loss and enables investigating more behavioural implications explaining the choice of schedules through estimating model specifications with socio-demographic variables.

Algorithm 1 Household choice-set generation for ABMs with MH

 $t \leftarrow 0$, initialise household state with random household schedule $X_t \leftarrow S_0$ \triangleright Household is comprised of agents $1, \ldots, n, \ldots, N_m$, with each agent having a state X_{n_t} Initialise household utility function with random parameters \hat{U}_S for t = 1, 2, ... do Choose agent I as index for n = I do Choose operator ω with probability P_{ω} $X_I^*, q(X_{I_t}, X_I^*) \leftarrow \text{APPLYCHANGE}(\omega, X_{I_t})$ function APPLYCHANGE(ω , state X_n) **return** new state X'_n , transition probability $q(X_n, X'_n)$ end function Check X_T^* feasibility in terms of continuity (no gaps in time or space) for $n \in \{1, \ldots, N_m\} \setminus \{I\}$ do Choose operator ω with probability P_{ω} $X_n^*, q(X_{n_t}, X_n^*) \leftarrow \text{APPLYCHANGE}(\omega, X_{n_t})$ Check X_n^* feasibility in terms of continuity (no gaps in time or space) Check X_n^* compliance with index agent Iend for end for Compute target weight $p(X^*) = \text{HUF}(X^*)$ Compute acceptance probability $\alpha(X_t, X^*) = \min\left(1, \frac{p(X^*)q(X_t|X^*)}{p(X_t)q(X^*|X_t)}\right)$ With probability $\alpha(X_t, X^*)$, set $X_{t+1} \leftarrow X^*$; else $X_{t+1} \leftarrow X_t$ end for return C_h : Ensemble containing clusters of schedules for agents $1, \ldots, N_m$ in household h

Operators, $\omega \in \Omega$, are heuristics that modify the current state of agents to create new candidate states. Operators are created according to modeller's needs. Dedicated operators should be implemented for the household context. For instance, participation mode operator $\omega_{\text{partic}_mode}$ changes whether an activity is performed jointly with other member(s) of the household or alone. 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.

In the context of household-level ABMs, each state is a household schedule, and the target weight is the household utility function with parameters calibrated on a randomly generated choice set. To derive the total utility for the household, the utility of individual household agents should be combined, depending on the nature of the group decision-making strategy. For example, in Utilitarianism/Additive-type household, 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 (Equation 1). The weights w_n , capture the relative "power" of each individual in the household-oriented decisions.

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

2.2.2 Parameter estimation

The household scheduling process is defined as a discrete choice problem. Each alternative is a household daily schedule, containing full daily schedules of all household agents. Each alternative is associated with a utility, capturing the household utility. The scheduling model parameters can be estimated with maximum likelihood estimation on the sampled choice set. The likelihood function is evaluated for each alternative of the choice set. The parameters are derived such that the likelihood function is maximised.

As the evaluation is carried out on a sample of the full universal choice set, the likelihood function is corrected with probability of sampling the choice set given the chosen alternatives (Ben-Akiva & Lerman, 1985). C_h is the generated choice set for household h. Thus, the probability that a household h chooses alternative $i_h \in C_h$, associated with a deterministic utility V_{ih} , is defined as follows:

$$P(i_h|C_h) = \frac{\exp\left[V_{ih} + \ln q(C_h|i_h)\right]}{\sum_{j_h \in C_h} \exp\left[V_{j_h} + \ln q(C_n|j_h)\right]}$$
(2)

 C_h is the choice set for household h, which contains clusters of schedules for all agents in the household. V_{ih} is the deterministic utility of the total household for alternative i_h . The alternative specific correction term take into account sampling biases defined as:

$$q(C_h|i_h) = \frac{1}{q_{ih}} \prod_{j_h \in C_h} \left(\sum_{j_h \in C_h} q_{jh} \right)^{J+1-\tilde{J}}$$
(3)

where C_h is the household choice set of size J + 1 with \hat{J} unique alternatives for household h. Unique alternatives are identified based on the combination of schedules of all household agents. j_h represents alternative sampled from the target distribution of the MH algorithm with probability q_{jh} . For each household and each alternative in their respective choice sets, the sample correction term is evaluated to be added to the utility function.

3 Empirical investigation

The data from the 2018-2019 UK National Travel Survey (NTS) (Department for Transport, 2022) is used to apply the methodology on a real-life case study. The NTS is a household survey containing information on daily trips and socio-economic characteristics of individuals and their household within the UK. The 2018-2019 version of the data contains 31'773 individuals, belonging to 13'418 households, and 140'879 trip diaries.

First, we select a sample of 5'466 2-membered households of 2 adults from the 2018-2019 UK NTS. We then generate choice sets of 10 alternatives for each household using the household-level choice set generation algorithm. After, we estimate the parameters of the utility function of a household-level activity-based model (Rezvany et al., 2023) for the sample.

We initially process the data to convert the trip diaries to daily activity schedules. Data points with missing information are excluded. We group the activities into 6 categories: Home, Work, Education, Leisure, Shopping, and Personal business (eg. eat/drink, using services like medical appointments).

The mode of start times and durations for each activity from the distribution across households of 2 with 2 adults, are used as indicators for desired start and duration times in the model (Table 1). The scheduling preferences are assumed to be homogeneous across the individuals.

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 42% of Leisure activities are performed jointly. Thus, in our choice set generation, we consider Leisure activities to have the possibility to be done either jointly or alone.

Table 1: Scheduling preferences

Activity	Desired start time [hh:mm]	Desired duration[hh:mm]
Work	08:00	08:30
Education	08:45	7:15
Leisure	10:30	02:20
Shopping	10:10	00:30
Personal business	10:30	00:30

3.1 Generated choice set: analysis and discussions

We run 1000 iterations of the algorithm, generating choice sets of size N = 10 alternatives for each household. The ensemble of observed schedules of household agents is used as the initial state of the random walk. A set of operators are implemented to modify the schedules to generate new states in the random walk. We consider Block, Assign, Swap, Anchor, Partic_mode, and Combination Meta-operators for the random walk. Each operator has equal probability of being chosen, denoted as $P_{\text{operators}}$. The target distribution of the random walk is the household utility function (Equation 1), with parameters calibrated on a randomly generated choice set. The accepted schedules are sampled after a warm-up period.

Figure 1 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 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 1b 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.

3.2 Parameter estimation: Model specifications and results

Using the generated choice set, the scheduling model has been estimated for the sample. 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. As precise location information is not available in the data, travel parameters are not estimated. The estimation solely focus on activity scheduling parameters. In this paper, a model containing activity- and scheduling-specific attributes, as well as socio-economic characteristics is presented. We include number of household cars, in the utility function. We want to test whether the household car ownership can potentially affect the agents' tendency toward participating in leisure activities jointly with other household agents. In this model specification, we include terms interacting the number of household cars with activity participation mode for the leisure activity. 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. Work activities bring the most utility per time unit followed by Shopping, Personal business, Leisure and Education activities. All parameter estimates are statistically significant. The estimated parameter for joint participation in leisure are significant. The estimation results indicates a tendency towards joint participation in leisure activities for agents in single car households. This indicates that doing leisure activities with other household agent(s) is 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.

Participating in activities jointly also requires coordination with other agent(s). As coordinating



Figure 1: Distribution of activity participation across different hours of day in generated choice sets

Name	Value	Rob. Std Err	Rob. t-test	Rob. p-value
Education:constant	2.35	0.617	3.81	0.000139
Education:early	-2.13	0.645	-3.3	0.000979
Education:late	-0.457	0.166	-2.86	0.00422
Education:long	-1.21	0.224	-5.41	6.14e-08
Education:short	-0.728	0.133	-5.47	4.54e-08
Leisure:constant	3.22	0.146	22	0
Leisure:early	-0.459	0.0324	-14.2	0
Leisure:joint_partic	0.244	0.109	2.25	0.0246
Leisure:joint_partic_no_car	-0.364	0.214	-1.7	0.0885
Leisure:joint_partic_two_or_more_car	-0.262	0.123	-2.13	0.0328
Leisure:late	-0.176	0.0169	-10.4	0
Leisure:long	-0.322	0.0188	-17.2	0
Leisure:short	-0.486	0.0607	-8	1.33e-15
Personal business:constant	3.77	0.239	15.8	0
Personal business:early	-0.75	0.107	-7.03	2.06e-12
Personal business:late	-0.326	0.0492	-6.62	3.51e-11
Personal business:long	-0.533	0.0497	-10.7	0
Personal business:short	-3.6	0.853	-4.22	2.44e-05
Shopping:constant	5.61	0.207	27.1	0
Shopping:early	-1.32	0.13	-10.2	0
Shopping:late	-0.237	0.0395	-6	2.02e-09
Shopping:long	-0.634	0.0438	-14.5	0
Shopping:short	-4.67	0.654	-7.14	9.34e-13
Work:constant	5.67	0.231	24.5	0
Work:early	-0.738	0.0839	-8.8	0
Work:late	-0.423	0.0559	-7.56	4.04e-14
Work:long	-0.747	0.0501	-14.9	0
Work:short	-0.576	0.0426	-13.5	0
Summary of statistics $L(0) = -12010.03$				
$L(\hat{eta}) = -1553.337$				

Table 2: Estimation results

with others might mean compromising on ones interests, coordination costs can decrease the tendency to participate in activities jointly. We can observe that households with no cars are less likely to do leisure activities jointly. This can be interpreted as they should use active or public transport modes for their travels, synchronising their schedules with other agents might be an extra effort which make them less inclined to coordinate their schedules for joint activity participation. Moreover, in households with two or more cars, agents have more tendency to have their independent schedules and avoid deviating from their preferences.

The penalty parameters have a negative sign, indicating a decline in utility when deviating from their preference. 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 sam-

pling 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 remain compliant with household-level constraints, in addition to individual-level validity constraints, (vi) the algorithm returns an ensemble containing clusters of schedules for individuals in household, and (vii) individual and household socio-demographic characteristics are preserved and reported in the generated choice-set. This feature enables testing model specifications containing socio-demographic variables. This procedure generates household-level choice set containing sufficiently varied alternatives for behaviourally sensible parameter estimates. 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. Moreover, complex travel-related interaction dimensions within household members such as resource constraints (e.g. car availability) and escort duties can be considered in the framework. The travel-related parameters can be estimated having access to the required data (e.g. location and network data). Furthermore, exploration of validation techniques can be considered. Validating the approach by estimating parameters with the sampled choice set, embedding the estimated parameters in the household-level OASIS (Rezvany et al., 2023) to simulate household daily schedules, and comparing the simulated schedule distributions with observed distribution from the dataset can be investigated..

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
- 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. Proceedia*, 23, 809–826. Retrieved from http://dx.doi.org/10.1016/j.trpro.2017.05.045 doi: 10.1016/j.trpro.2017.05.045