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

Negar Rezvany *   Tim Hillel †   Michel Bierlaire *

July 2024

Report TRANSP-OR 240731
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {negar.rezvany,michel.bierlaire}@epfl.ch

†University College London (UCL), Department of Civil, Environmental and Geomatic Engineering (CEGE), United Kingdom, {tim.hillel@ucl.ac.uk}

Abstract

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 the utility functions. However, classical data sources like travel diaries only contain chosen alternatives, not the 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 algorithm 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 the parameters of a household-level scheduling model presented in (Rezvany et al. 2023). This approach aims to generate behaviourally sensible parameter estimates, enhancing the model realism in capturing household dynamics.

Keywords: Activity-based modelling, Choice-set generation, Discrete choice modelling, Intra-household interactions.

1 Introduction

1.1 Motivation and scope

ABMs consider the demand for travel to be driven by participation in spatially and temporally distributed activities. By including why trips are derived, they try to replicate the actual decisions with more behavioural realism compared to the traditional trip-based models focusing on individual trips. 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 the household decision-making perspectives. Hence, models dealing with individual choices need to be revised to take account of the intra-household interactions.

There are two major research streams within the scope of ABMs: (i) rule-based/computational process models (e.g. (Arentze & Timmermans 2004)), and (ii) econometric models (e.g. (Nurul Habib 2018)). Econometric models are based on the assumption that individuals choose their schedule such that the utility they gain is maximised. Activity scheduling and travel behaviour is explained and predicted as a result of discrete choices, treated sequentially or jointly, and solved with econometric methods such as advanced discrete choice models (Bowman & Ben-Akiva 2001) or micro-simulation (e.g. (Bhat 2005)). Thus, using discrete choice models implies the need for calibration of maximum likelihood estimators of the parameters of the utility functions.

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 indeed 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 activity-based models. 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 MH based sampling algorithm introduced by Pougala et al. (2021) Necessary considerations in household choice set generation is noted. Utilising 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 give a brief review of the literature in Section 1.2. In Section 2 the household-level choice set generation methodology is explained. Section 3 presents an empirical investigation to apply the methodology on a real-life case study, followed by analysis of the results. It is followed by discussions on a household-level versus individual-level choice set generation (Section 4). Finally, the concluding remarks and opportunities for future research are presented in Section 5.

1.2 Relevant literature

The scheduling process is central to the activity-based research. Most of the conventional activity-based models in transportation research are based on individual decision-making process where the individuals are treated as isolated agents whose choices are independent of other decision-makers. However, ignoring the interdependence between household members causes biased simulation of activity-travel schedules and lead to inappropriate actions and investment as the schedule of household members are mutually dependent. Capturing interpersonal dependencies between individuals belonging to the same household enhances consistency of predicted choices and behaviour. In (Rezvany et al. 2023), we propose an operational utility-based scheduling framework that explicitly captures multiple intra-household interactions within a single ABMs using a simultaneous approach. The model explicitly accommodates complex interactions among household members such as the allocation of private vehicle to household members, escort duties, joint participation in activities, and sharing rides.

One challenge in the utility-based ABMs is model calibration. There are little work in the field of activity-based modelling specifically tackling estimation of model parameters. Parameter estimation can broadly considered through two approaches; fixed arbitrary parameter values (e.g. (Charypar & Nagel 2005)) or empirical parameter estimation based on data calibration. However, as the traditional surveys such as travel diaries are limited to only revealed preferences, behavioural parameters such as penalties and preferences cannot be easily derived. 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 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). They identify representative patterns from observed activity-travel patterns. Using a genetic algorithm, a pattern is sampled from each of the non-chosen representative pattern clusters such that the information gain is optimised by minimising the D-error of the final sample. A goal-programming is then used to adjust the sampled alternatives according to individuals' spatial and temporal constraints to ensure feasibility of the generated choice set.

Shakeel et al. (2022) focus on modelling potential joint leisure activities within household members using a latent class model. They focus solely on the generation process before the negotiation within household members for scheduling decisions. They establish the linkage between household and individual attributes affecting joint-activity generation. Further research on investigating the generation of joint activities, estimating travel parties involved in joint activity, as well as integrating the model in operational activity-based model are suggested.

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 define terms and notations used in the algorithm and manuscript (Section 2.1). We then give a brief synopsis of the base MH based sampling strategy (Section 2.2) and present the household-level choice set formation and model estimation framework (Section 2.3).

2.1 Definitions

We summarise the a glossary of terms used in the framework in Table 1.

Table 1: Notations used in the framework

Notation	Name	Description
n	Agent	An individual having decision making capabilities, determined by both preferences and constraints, $n \in \{1, 2, ...N_m\}$.
N_m	Household size	Number of agents in the household.
h	Household	A household, composed a set of N_m agents.
A^n	Considered activity set	An activity set containing all activities a_n that agent n considers performing within her time budget T.
a_n	Activity	Activity a_n that can be performed by agent n.
p_{a_n}	Activity participation mode	A binary variable, indicating engagement mode of activity a_n, which is 1 if performed jointly with other agent(s), and 0 if performed solo.
ℓ_{a_n}	Activity location	Location for activity a_n.
L_{a_n}	Activity location choice set	A discrete and finite location choice set containing all locations ℓ_{a_n} that agent n considers for activity a_n.
m_{a_n}	Transportation mode	The mode to travel from the location of the current activity, ℓ_{a_n}, to the location of the following activity, ℓ_{a+1_n}.
M	Transport mode choice set	A discrete and finite list of considered transport modes.
x_{a_n}	Activity start time	A positive continuous variable representing the start time of activity a_n.
$x^*_{a_n}$	Desired activity start time	An indicative of the desired start time of activity a_n.
τ_{a_n}	Activity duration	A positive continuous variable representing the duration of activity a_n.
$\tau^*_{a_n}$	Desired activity duration	An indicative of the desired duration of activity a_n.
T	Time budget	The time period over which the schedules are generated.
δ	Time block	The schedule is discretised into blocks of duration δ.
δ_{min}	Minimum block duration	Minimum duration of a block.
w_n	Agent priority parameter	Relative weight capturing the priority that is placed on the schedule utility of each agent.
C_h	Choice set	Generated choice set for household h.
i_h	Alternative	Alternative (cluster of agents schedules) i for household h, $i_h \in C_h$.
V_{ih}	Deterministic utility	Deterministic utility of household h for alternative i_h.
X_t	Household state	Household state at step t, which is household schedule comprised of a cluster of schedules of agents in the household; $[X_{1_t}, \ldots, X_{N_{m\,t}}]$.

Continued on next page

Table 1 - Notations used in the framework (Continued)

Notation	Name	Description
X_{n_t}	Agent state	State (schedule) of agent n at step t.
X^*	Neighbouring state	A schedule that can be reached in one step by applying an operator to the current schedule.
ω	Operator	A heuristic that modify specific aspects of the schedule (time, space, participation, or activity participation mode (solo, joint)).
Ω	Set of Operators	A set of possible heuristics that can be used in the algorithm.
$N_{\text{operators}}$	Number of operators	Number of implemented operators to modify the schedules.
P_ω	Operator selection probability	Probability to select operator ω.

2.2 Base Metropolis-Hastings based sampling strategy for ABMs: 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).

2.3 Household-level choice set generation and parameter estimation

2.3.1 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, $[X_{1_t}, \ldots, X_{N_{m_t}}]$. The state of each agent n, X_{n_t}, is her/his schedule within the time budget T (eg. 24 hr), discretised in blocks of duration $\delta \in [\delta_{\text{min}}, 24 - \delta_{\text{min}}]$, 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 ← 0, initialise household state with random household schedule X_t ← S_0
            ▷ Household is comprised of agents 1, ..., n, ..., N_m, with each agent having a state X_{n_t}.
Initialise household utility function with random parameters Û_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) ← APPLYCHANGE(ω, 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*_I feasibility in terms of continuity (no gaps in time or space)
        for n ∈ {1, ..., N_m} \ {I} do
            Choose operator ω with probability P_ω
            X*_n, q(X_{n_t}, X*_n) ← APPLYCHANGE(ω, X_{n_t})
            Check X*_n feasibility in terms of continuity (no gaps in time or space)
            Check X*_n compliance with index agent I
        end for
    end for
    Compute target weight p(X*) = HUF(X*)
    Compute acceptance probability α(X_t, X*) = min(1, p(X*)q(X_t|X*) / p(X_t)q(X*|X_t))
    With probability α(X_t, X*), set X_{t+1} ← X*; else X_{t+1} ← X_t
end for
return C_h: Ensemble containing clusters of schedules for agents 1, ..., 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.

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

2.3.2 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:

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

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:

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

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 $8'560$ individuals, belonging to $4'280$ households of 2 adults, and $44'922$ daily trip diaries ($134'064$ trips). It is a panel data, containing trip diaries of multiple days for the households.

First, we generate choice sets of 10 alternatives for each household using the household-level choice set generation algorithm. We then 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. For this case study, a sample of 2018-2019 daily schedules for $3'126$ households of size 2 with 2 adults, with no missing variables in the data, is used. The sample size is 5466 containing multiple observations per households, over multiple days. 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 2). The scheduling preferences are assumed to be homogeneous across the individuals.

Table 2: 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

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.

3.1 Generated choice set

We run $n_{\text{iter}} = 1000$ iterations of the algorithm for a sample of $5'466$ households of 2 adults. Choice sets of 10 alternatives (including the chosen schedules) is generated 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. 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. The process is performed for a number of iterations, n_{iter}. The initial $n_{\text{warm-up}}$ iterations serve as a warm-up period to stabilise the distribution from which the choice set is sampled. Following this, a set of 9 alternatives is sampled. Table 3 summarises the experimental set-up for household-level choice set generation.

Table 3: Experimental set up for choice set generation.

Feature	Definition	Value
Ω	Set of operators	Block, Assign, Anchor, Swap, Inf/Def, Partic_mode, Meta
$N_{operators}$	Number of operators	7
$P_{operators}$	Operator selection probability	$1/N_{operators}$
n_{iter}	Number of iterations	$1'000$
$n_{warm-up}$	Warm-up iterations	50

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 pick 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

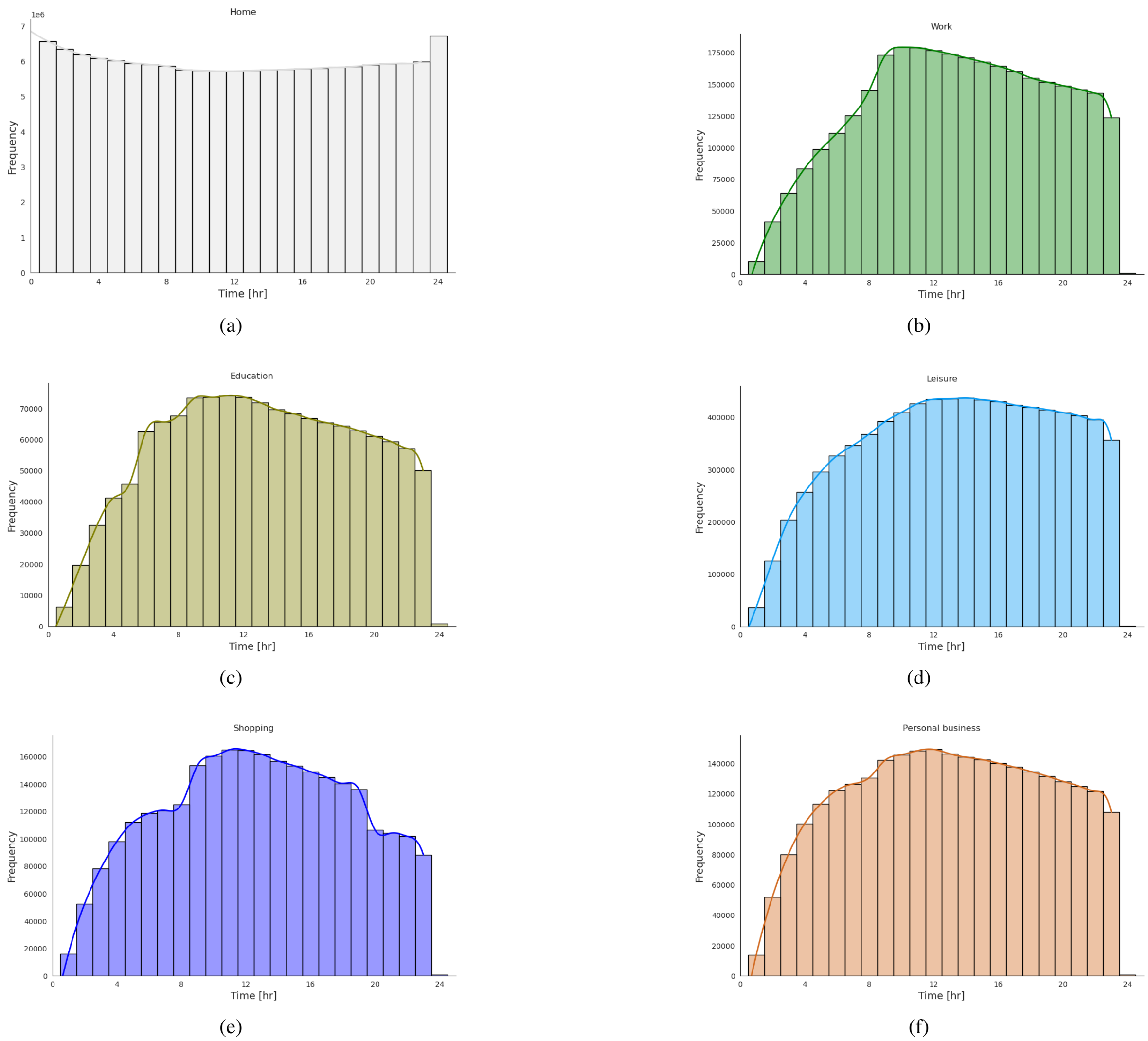


Figure 1: Distribution of activity participation across different hours of day in generated choice sets

after the daily activities. Figures 1c and 1b indicate distinct peak activity times for education and work with concentrated density during typical school and 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 household 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. The magnitudes and signs of the other constants are relative to the baseline behaviour which is staying at home. As precise location information is not available in the data, travel parameters are not estimated. For estimation of travel parameters location and network data are required to compute attributes for chosen and unchosen alternatives. The estimation solely focus on activity scheduling parameters. The models are estimated with PandasBiogeme (Bierlaire 2020). As choice set might consist multiple observations per each household, panel specifications has been considered. It is notable that as we have panel data in the sampled choice set, there is serial correlation, as the error terms associated with the observations obtained from the same households share a great deal of unobserved variables. Therefore, panel effect is considered in the estimation procedure of models. Model specifications differing in specification of the utility function have been tested. In this paper, a model containing only activity- and scheduling-specific attributes (Section 3.2.1), as well as model specification with socio-economic characteristics (Section 3.2.2) are presented.

3.2.1 Base model with activity-specific parameters

In this specification, the attributes used in the model are related to the activity-specific constants and parameters, as well as scheduling deviation penalties. For each alternative, the household utility function is defined as follows:

$$\text{HUF} = \sum_{n=1}^{n=N_m} w_n \, U_n \tag{4}$$

where n presents an agent having decision-making capabilities in the household. N_m is the number of agents in the household. w_n is the agent priority parameter, which captures the heterogeneous influence of household members on household decisions by accounting for how much relative priority is placed on the utility of each individual. In this case-study w_n is set to 1 for all agents in the household, indicating identical relative influence for household agents.

For each agent n, the utility function for each alternative is defined as follows:

$$U_n = \sum_{a_n \in A^n} U_{a_n} \tag{5}$$

where U_n is the utility associated with schedule for agent n. U_n is made up of utility components linked to the performed activities (U_{a_n}). The activity-specific utility function for each activity a_n of agent n is defined as follows:

$$\begin{aligned} U_{a_n}^{\text{act. sp.}} = \gamma_{a_n} \quad &+ \quad \theta_{a_n}^{\text{early}} \max(0, x_{a_n}^* - x_{a_n}) \quad + \quad \theta_{a_n}^{\text{late}} \max(0, x_{a_n} - x_{a_n}^*) \\ &+ \quad \theta_{a_n}^{\text{short}} \max(0, \tau_{a_n}^* - \tau_{a_n}) \quad + \quad \theta_{a_n}^{\text{long}} \max(0, \tau_{a_n} - \tau_{a_n}^*) \quad + \quad \theta_{a_n}^{\text{jnt}} \, p_{a_n} + \quad \epsilon_{S_n} \end{aligned} \tag{6}$$

where γ_{a_n} is the activity-specific constants, $\theta_{a_n}^{\text{early}}$ and $\theta_{a_n}^{\text{late}}$ are start time penalty parameters for deviations from preference, $\theta_{a_n}^{\text{short}}$ and $\theta_{a_n}^{\text{long}}$ are duration penalty parameters for deviations from preference. x_{a_n} is start time of activity a_n. $x_{a_n}^*$ is preferred start time for activity a_n. τ_{a_n} and $\tau_{a_n}^*$ are duration and preferred duration of activity a_n, respectively. $\theta_{a_n}^{\text{jnt}}$ is joint activity participation parameter for activity a_n, capturing the (dis)utility of joint activity engagement. p_{a_n} is the participation mode of activity a_n, which is 1 if the agent performs the activity jointly with other agent(s), and 0 otherwise. ϵ_{S_n} is an error term capturing unobserved variables in the utility of the schedule of agent n.

Table 4 summarises the estimation results. Home activity is set as a reference, thus magnitudes and signs of coefficients are relative to the home baseline. The estimated parameters are all 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.

Most parameter estimates are significant at 95% confidence interval.

The penalty parameters have a negative sign, indicating a decline in utility when deviating from their preference. For example the significant negative coefficient for education earlier than preferred suggests individuals find less utility in participating to education activities earlier than their preferred timing. This can be as starting education earlier can lead to reduced sleep, as well as that it may also limit individuals from engaging in

Table 4: Estimation results for base model with activity-specific parameters on households of 2 adults

Name	Value	Rob. Std err	Rob. t-test	Rob. p-value
Education:constant	2.41	0.615	3.91	9.26e-05
Education:early	-2.15	0.659	-3.27	0.00108
Education:late	-0.455	0.16	-2.84	0.00449
Education:long	-1.25	0.224	-5.59	2.33e-08
Education:short	-0.734	0.135	-5.45	4.94e-08
Leisure:constant	3.22	0.146	22.1	0
Leisure:early	-0.455	0.032	-14.2	0
Leisure:joint_partic	0.892	0.0893	10	0.318
Leisure:late	-0.175	0.0169	-10.4	0
Leisure:long	-0.322	0.0186	-17.3	0
Leisure:short	-0.482	0.0602	-8.0	1.33e-15
Personal business:constant	3.75	0.239	15.7	0
Personal business:early	-0.748	0.106	-7.03	2.14e-12
Personal business:late	-0.324	0.0495	-6.53	3.9e-11
Personal business:long	-0.531	0.0495	-10.7	0
Personal business:short	-3.65	0.855	-4.27	1.99e-05
Shopping:constant	5.6	0.206	27.2	0
Shopping:early	-1.32	0.13	-10.2	0
Shopping:late	-0.237	0.0395	-6	1.92e-09
Shopping:long	-0.631	0.0437	-14.4	0
Shopping:short	-4.63	0.646	-7.17	7.68e-13
Work:constant	5.69	0.233	24.4	0
Work:early	-0.743	0.0841	-8.83	0
Work:late	-0.423	0.0563	-7.51	5.88e-14
Work:long	-0.749	0.0504	-14.9	0
Work:short	-0.58	0.0428	-13.6	0

Summary of statistics
Number of estimated parameters = 26
Sample size = 3126
Number of observations = 5466
$\mathcal{L}(0)$ = -12010.03
$\mathcal{L}(\hat{\beta})$ = -1557.137
AIC = 3166.275

other activities that provide greater utility or satisfaction during those earlier hours for them such as personal care, exercise, or leisure activities that contribute to a balanced lifestyle. The negative and significant estimate for shorter work activities than preferred may reflect the disutility associated with not fulfilling expected work hours, which could impact productivity or income. Shorter durations than expected are penalised about 7 times more than longer for shopping and personal business activities. These negative and significant estimates may reflect the disutility associated with not fulfilling individual and household needs, impacting overall household satisfaction and well-being. 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.

3.2.2 Model with socio-economic attributes

The previous model includes only variables that are attributes of the alternatives (schedules), assuming a homogeneous population, where the taste parameters are shared by everyone. It is reasonable to assume that people have different tastes. In the context of choice models, it means that the value of the parameters may depend on the socio-economic characteristics of the decision-makers. Socio-economic characteristics do not vary across alternatives. Therefore, their role in the model is to capture the heterogeneity of taste. We now investigate the possible heterogeneity of tastes in the population. The following model specifications are discussed in this paper:

1. **Model 1: model specification with Gender as socio-economic attribute:**
   We now introduce socio-economic variables Gender_1, and Gender_2 which indicates the gender of agent 1 and agent 2 in a household, respectively. The variables are categorical and equals 1 if the gender of the agent is male and 0 if female. We consider segments in the population, characterized by the socio-economic characteristics and we associate a different set of taste parameters to each of the segments. Here, we interact gender with participation in each of the activities. In other words, we assume that different genders have different tastes in participating in activities.

   We define the segmented version of a parameter θ as follows:

   $$\theta_{\text{segmented}} = \theta_{\text{baseline}} + \text{Gender}_n \theta_{\text{male}} \tag{7}$$

   where $\text{Gender}_n = 1$ when agent n is male, and $\text{Gender}_n = 0$ when female.

   The activity-specific utility functions for each agent n are given by Equation 6. Segmented model parameters (Equation 7) are used in the specification. For each alternative, the household utility is defined as Equation 4.

   The estimated parameters are summarised in Table 5. The signs of the parameter estimates are as expected. The parameters associated with earlier start times of activities Education and Leisure are significant at 5% confidence interval. For Education, the parameter for earlier start time deviations for males is significant and negative, while the baseline start time deviation parameter is also negative. This would lead to more negative start time deviation parameters for males for Education compared to females. This can be interpreted that males are more sensitive to deviation in early start time of Education activities compared to women, all else being equal. For Leisure, the parameter for earlier start time deviations for males is significant and positive, while the baseline start time deviation parameter is negative. This would lead to less negative start time deviation parameter for males for Leisure activities compared to females. This can be interpreted that males are less sensitive to deviation in earlier start time of Leisure activities compared to females, all else being equal.

   Comparing to the base model (Section 3.2.1), the results show that there is an improvement in the final log likelihood (from -1557.137 to -1546.605), but an increase of Akaike Information Criterion ratio index (from 3166.275 to 3175.211). It is a sign that the improvement of the fit may not be sufficient for the number of parameters involved. Using a likelihood ratio test (with a test statistics of $21.064 < \chi^2_{0.95,15} = 24.996$), we can conclude that the base model without taste heterogeneity across genders cannot be rejected at the 5% significance level.

2. **Model 2: model specification containing the interaction between number of household cars with activity participation mode:**
   In this model specification, we include the number of household cars or light vans (including landrover, jeep, minibus), 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. The number of household cars is considered through three binary dummy variables; noCar, OneCar, and TwoMoreCar. For each agent n, terms interacting activity participation mode with number of household cars are added to the utility term.

$$\begin{aligned} U_{a_n}^{\text{car int.}} = U_{a_n}^{\text{act. sp.}} + \theta_{\text{joint\_partic\_NoCar}} \cdot p_{a_n} \cdot \text{NoCar} + \\ \theta_{\text{joint\_partic\_OneCar}} \cdot p_{a_n} \cdot \text{OneCar} + \\ \theta_{\text{joint\_partic\_TwomoreCar}} \cdot p_{a_n} \cdot \text{TwomoreCar}, \quad \forall a_n \text{ in ['Leisure']} \end{aligned} \tag{8}$$

where $U_{a_n}^{\text{act. sp.}}$ is activity-specific utility function for each activity a_n of agent n (Equation 6). p_{a_n} is a binary variable indicating participation mode for activity a_n. noCar, OneCar, and TwoMoreCar are binary variables indicative of a household with no cars, one car, and two or more cars, respectively. The $\theta_{\text{joint\_partic\_NoCar}}$, $\theta_{\text{joint\_partic\_OneCar}}$, and $\theta_{\text{joint\_partic\_TwomoreCar}}$ are the associated parameters capturing interaction between the number of household cars and activity participation mode. $\theta_{\text{joint\_partic\_OneCar}}$ is normalised to zero for identification purposes.

The estimated parameters are summarised in Table 6. All parameter estimates are 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 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.

Comparing to the base model (Section 3.2.1), the results show that there is an improvement in the final log likelihood (from -1557.137 to -1553.337). Using a likelihood ratio test (with a test statistics of $7.6 > \chi^2_{0.95,2} = 5.991$), we can conclude that the model considering the interaction between number of household cars and joint activity participation has a better fit at the 5% significance level.

3.3 Simulation results

Using the estimated parameters Section 3.2, we simulate the schedules for the sample dataset. A detailed explanation of the simulation procedure can be found in Rezvany et al. (2023). For each household in the sample, 100 realisations are drawn from the underlying schedule distribution.

Table 5: Estimation results for Model1 with gender as socio-economic characteristic on households of 2 adults

Name	Value	Rob. Std err	Rob. t-test	Rob. p-value
Education:constant	2.2	0.715	3.07	0.00211
Education:constant_MALE	0.716	0.887	0.808	0.419
Education:early	-1.63	0.471	-3.46	0.00053
Education:early_MALE	-2.93	1.28	-2.28	0.0225
Education:late	-0.304	0.198	-1.53	0.126
Education:late_MALE	-0.315	0.265	-1.19	0.235
Education:long	-1.24	0.226	-5.5	3.73e-08
Education:short	-0.753	0.131	-5.77	8.13e-09
Leisure:constant	3.31	0.198	16.7	0
Leisure:constant_MALE	-0.166	0.265	-0.626	0.532
Leisure:early	-0.545	0.0519	-10.5	0
Leisure:early_MALE	0.18	0.0731	2.47	0.0136
Leisure:joint_partic	0.0942	0.0886	1.06	0.288
Leisure:late	-0.183	0.0322	-5.69	1.29e-08
Leisure:late_MALE	0.0163	0.0509	0.32	0.749
Leisure:long	-0.322	0.0187	-17.2	0
Leisure:short	-0.482	0.0597	-8.07	6.66e-16
Personal business:constant	3.78	0.284	13.3	0
Personal business:constant_MALE	-0.0964	0.36	-0.268	0.789
Personal business:early	-0.825	0.148	-5.57	2.53e-08
Personal business:early_MALE	0.191	0.188	1.02	0.309
Personal business:late	-0.299	0.0698	-4.29	1.81e-05
Personal business:late_MALE	-0.0372	0.0967	-0.384	0.701
Personal business:long	-0.535	0.0495	-10.8	0
Personal business:short	-3.6	0.855	-4.21	2.6e-05
Shopping:constant	5.66	0.242	23.4	0
Shopping:constant_MALE	-0.0518	0.291	-0.178	0.859
Shopping:early	-1.3	0.147	-8.86	0
Shopping:early_MALE	-0.0715	0.257	-0.278	0.781
Shopping:late	-0.193	0.0518	-3.73	0.000193
Shopping:late_MALE	-0.0895	0.0773	-1.16	0.247
Shopping:long	-0.633	0.044	-14.4	0
Shopping:short	-4.69	0.651	-7.2	5.9e-13
Work:constant	5.99	0.31	19.4	0
Work:constant_MALE	-0.532	0.344	-1.55	0.121
Work:early	-0.806	0.136	-5.95	2.73e-09
Work:early_MALE	0.114	0.165	0.694	0.488
Work:late	-0.513	0.194	-5.62	1.95e-08
Work:late_MALE	0.164	0.101	1.62	0.104
Work:long	-0.754	0.0499	-15.1	0
Work:short	-0.584	0.0416	-14	0

Summary of statistics
Number of estimated parameters = 41
Sample size = 3126
Number of observations = 5466
$\mathcal{L}(0)$ = -12010.03
$\mathcal{L}(\hat{\beta})$ = -1546.605
AIC = 3175.211

Table 6: Estimation results for Model2 considering interaction between number of cars (categorical) and joint activity participation on households of 2 adult

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

Number of estimated parameters = 28

Sample size = 3126

Number of observations = 5466

$\mathcal{L}(0)$ = -12010.03

$\mathcal{L}(\hat{\beta})$ = -1553.337

AIC = 3162.674

4 Discussions: Household-level vs individual-level choice set generation

In this section, we compare and discuss the household-level with individual-level choice set formation technique. Within-household interactions lead to additional complexities in the household scheduling. In the household-level choice-set generation technique, these aspects can be broadly classified as: (i) additional choice dimensions; activity participation mode; whether an individual participates in an activity solo or jointly with another household member, (ii) time arrangements; schedule synchronisation between participating agents in joint activities, (iii) constraints; such as resource availability and limitation, (iv) group decision-making mechanism; moving from schedule utility of isolated individuals to household utility function, reflected in the MH algorithm through the target distribution and target weight of each candidate state (state = cluster of schedules of individuals in a household). Table 7 presents a summary comparison of household-level and individual-level choice set formation.

Table 7: Summary comparison of household-level vs individual-level choice-set formation

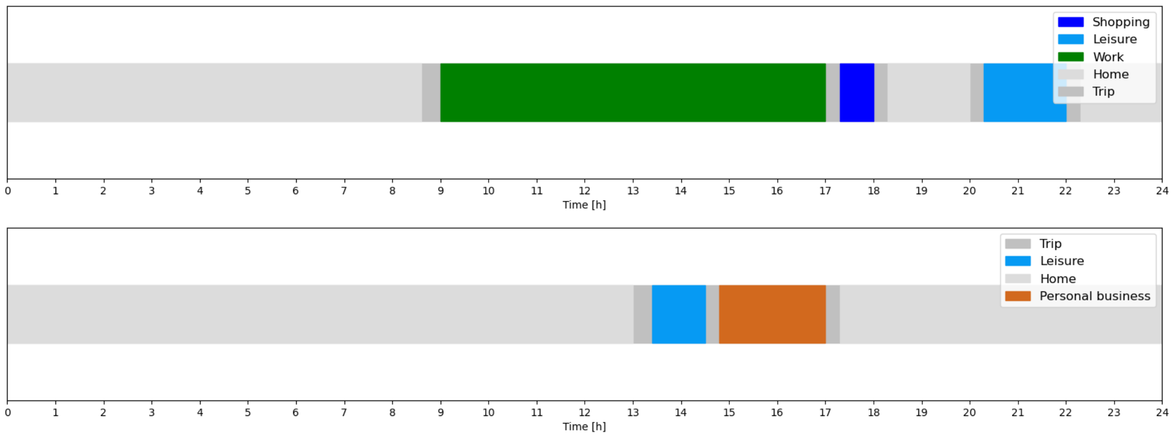
Feature	Individual-level	Household-level
Initialisation	Agent schedule	Schedule of all household agents
Generation procedure	Separate	In parallel
Target weight	Schedule utility	Household utility
Decision variables	Specific to activities schedule and location	Specific to activities schedule, location and companionship
Constraints	Schedule continuity	Schedule continuity and household synchronisations
Operators	Specific to scheduling dimensions	Specific to companionship and scheduling dimensions
Output	Ensemble of schedules	Ensemble of cluster of schedules

Choice-set generation technique for household scheduling, generates an ensemble of schedules with consistent alternatives for all household members, forming choice set of all individuals in a household in parallel. This ensures inter-agent validity of alternatives in the choice-set, enhancing model realism in capturing household dynamics. Whereas 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.

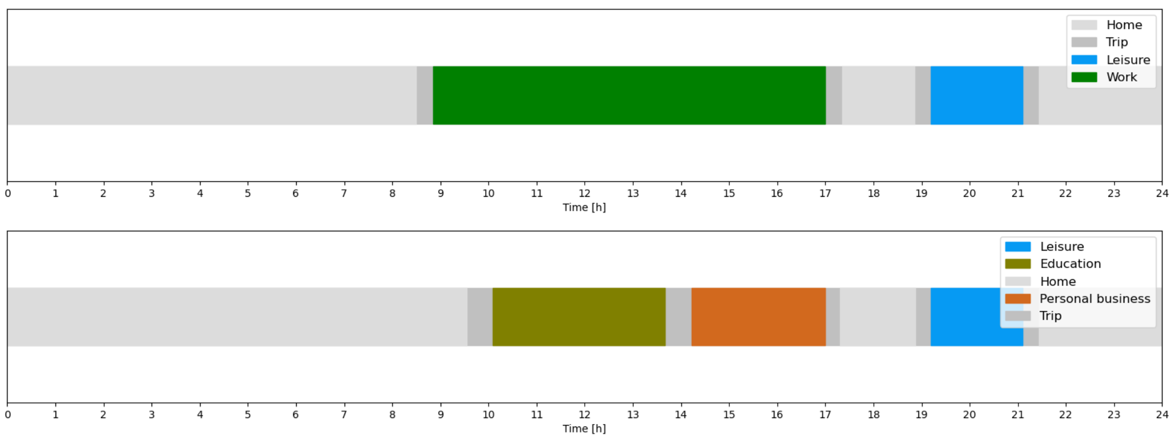
For instance, Figure 2 presents an example showcasing compatibility of generated alternatives in the choice set with household-level algorithm. Figure 2a shows the initial schedules of the 2 agents in a household of 2. Figure 2b presents the schedule of the 2 agents in an example generated alternative. The synchronisation between the schedules of agent 1 and 2 for the joint Leisure activity engagement can be observed in the generated schedules (Figure 2b). Furthermore, the effect of various heuristics that modify the initial schedules to generate choice set alternatives can be observed in the presented example. The results are indicative of the capability of the algorithm to generate compatible schedules for the agents in multi-member households considering interactions within members.

Analysing the generated choice-set with the household-level algorithm, the frequency of leisure activities with activity participation type chosen as joint, is identical for both agents in the household. This equality is not valid for the generated choice-set with individual-level choice-set formation technique. The observed compatibility between the generated schedules in the choice-set, both through observations from randomly selected alternatives and also aggregated checks on the whole choice set, ensures the soundness of the algorithm logic.

As an empirical investigation of the added value of estimating parameters with the household-level model compared to the individual-level choice set generation model, we simulate schedules for a sample test data using parameter estimates from both algorithms. Figure 3 presents the distribution of activities in the course of a day, in the data (Figure 3a), and resulting from the simulator framework using individual-level model (Figure 3b), and the household-level algorithm (Figure 3c). The distributions are for schedules including at least one activity out of home. The height of each bar represents the proportion of the sample that is participating in each activity at a given moment of time. From the results, the household-level model seems to provide more realistic results, closer to the observations in the data.



(a) Initial schedules for agent 1 (top) and agent 2 (bottom) in a household

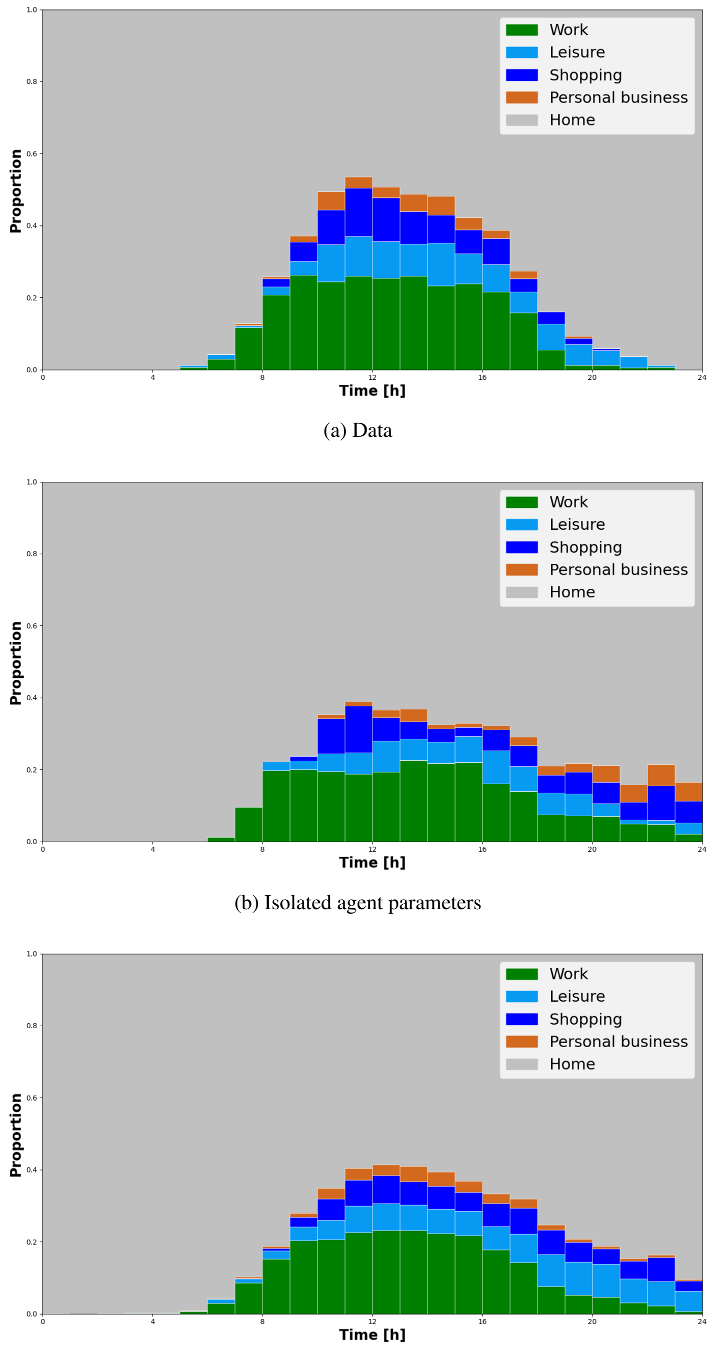


(b) Example generated schedules for the household; agent 1 (top) and agent 2 (bottom)

Figure 2: Example alternatives from household-level choice-set

5 Conclusions

In this paper, implementation requirements for ABMs with intra-household interactions is discussed. We propose a procedure to generate household-level choice set containing sufficiently varied alternatives for behaviourally sensible parameter estimates. A parameter estimation process for household-level ABMs, using discrete choice modelling, is then presented. Our household-level choice set generation methodology build on he MH based sampling algorithm developed by Pougala et al. (2021). The main characteristics of our household choice-set generation framework 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) new operators are introduced to modify choice dimension aspects related to household scheduling, (iv) the accepted schedules remain compliant with household-level constraints, in addition to individual-level validity constraints, (v) the algorithm returns an ensemble containing clusters of schedules for individuals in household, and (vi) 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. 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, even with a relatively small number of alternatives in the choice set.



(a) Data

(b) Isolated agent parameters

(c) Household parameters

Figure 3: Time of day activity frequency

There are further extensions and improvements of the current work, suggesting avenues for future research. The scheduling preferences are assumed to be homogeneous across the sample. Investigating non-homogeneous preferences across individuals can be considered. For example, for each activity, a distribution across the population can be fitted. For each individual, desired start times and durations can be then drawn from these distributions. Another interesting extension to the choice set generation algorithm, is capturing correlations between day-to-day scheduling for multi-day analysis. 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.

Arentze, T. A. & Timmermans, H. J. (2004), 'A learning-based transportation oriented simulation system', *Transp. Res. Part B Methodol.* **38**(7), 613–633.

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

Bhat, C. R. (2005), 'A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions', *Transp. Res. Part B Methodol.* **39**(8), 679–707.

Bierlaire, M. (2020), A short introduction to PandasBiogeme, Technical report.

Bowman, J. L. & Ben-Akiva, M. E. (2001), 'Activity-based disaggregate travel demand model system with activity schedules', *Transp. Res. Part A Policy Pract.* **35**(1), 1–28.

Charypar, D. & Nagel, K. (2005), 'Generating complete all-day activity plans with genetic algorithms', *Transportation (Amst).* **32**(4), 369–397.

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]'.

Ho, C. & Mulley, C. (2013), 'Tour-based mode choice of joint household travel patterns on weekend and weekday', *Transportation (Amst).* **40**(4), 789–811.

Nurul Habib, K. (2018), 'A comprehensive utility-based system of activity-travel scheduling options modelling (CUSTOM) for worker's daily activity scheduling processes', *Transp. A Transp. Sci.* **14**(4), 292–315.

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.

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**.
URL: *https://www.sciencedirect.com/science/article/pii/S0968090X23003522*

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.
URL: *https://doi.org/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.
URL: *http://dx.doi.org/10.1016/j.trpro.2017.05.045*