Parameter estimation for activity-based models

Janody Pougala

Tim Hillel

Michel Bierlaire

STRC conference paper 2022 **18-20 May 2022**

STRC | **22nd Swiss Transport Research Conference**
Monte Verità / Ascona, May 18-20, 2022

Parameter estimation for activity-based models

Janody Pougala
TRANSP-OR
EPFL
CH-1015 Lausanne
janody.pougala@epfl.ch

Tim Hillel
Human-Centric Laboratory
UCL
London, UK
tim.hillel@ucl.ac.uk

Michel Bierlaire
TRANSP-OR
EPFL
CH-1015 Lausanne
michel.bierlaire@epfl.ch

18-20 May 2022

Abstract

Activity-based models (ABM) have seen a significant increase in research focus in the past decade. Based on the fundamental assumption that travel demand is derived from the need to do activities and time and space constraints (Hägerstraand, 1970; Chapin, 1974). ABM offer a more flexible and behaviorally centered alternative to traditional trip-based approaches. Econometric – or utility-based – activity-based models (e.g. Adler and Ben-Akiva, 1979; Bowman and Ben-Akiva, 2001) postulate that the process of activity generation and scheduling can be modelled as discrete choices. Individuals derive a utility from performing activities, and they schedule them as to maximize the total utility. In this paper, we estimate the parameters of the optimization-based activity-based model developed by Pougala *et al.* (2022), by defining a discrete choice model where the choices for each individual are full daily schedules, each associated with a utility. The maximum likelihood estimators of the parameters (e.g. scheduling penalties, desired start times and durations, constants…) are evaluated on a choice set of daily schedules sampled using the Metropolis-Hastings algorithm (Pougala *et al.*, 2021), derived for sample of individuals from the 2015 Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017). Results show that the proposed methodology significantly improves the calibration of econometric activity-based models.

Keywords

Activity-based modeling, parameter estimation, discrete choice modeling, simulation

1 Introduction

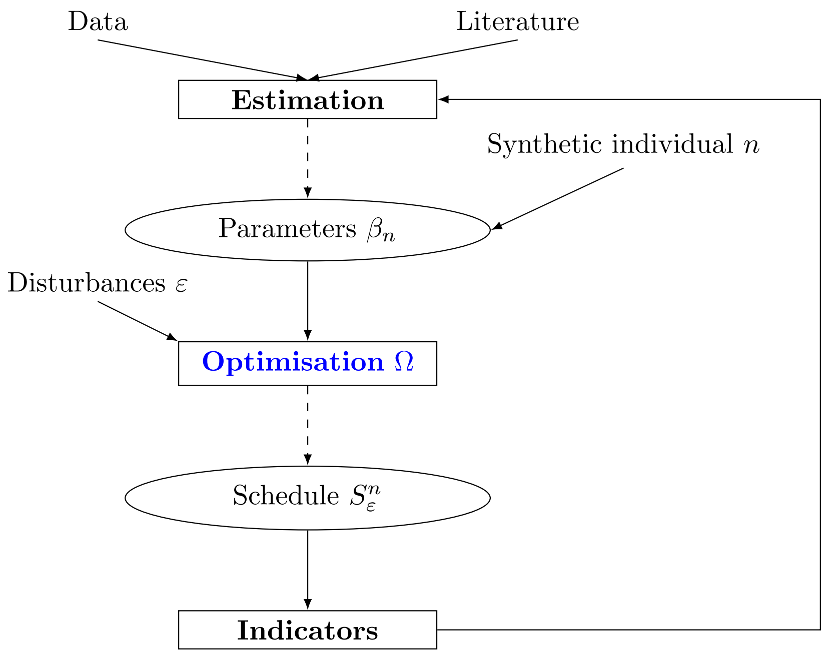
Activity-based models have been the focus of increasing research efforts in a variety of domains: including, but not limited to, transport research, energy demand, or epidemiology. In transportation, they provide a more behaviourally realistic alternative to traditional trip-based models and aggregate analyses.

A significant limitation and challenge in activity-based modelling research is a consistent estimation of stable and significant parameters. This requires an often prohibitive amount of data. In addition, the factors influencing the choice of activity (and related mobility decisions) are difficult to observe because they stem from interactions in time and space within and between cultural, social and physical environment. This information is impossible to obtain from classical data sources (e.g. travel diaries, time use surveys), and there are very few stated preference surveys - which aim to record behavioural data on hypothetical or unseen situations - who are dedicated to activities and daily scheduling, and cover a sufficient population size to calibrate a model.

Another limitation is methodological: many activity-based models in practice are utility-based (i.e. they are based on the random utility maximisation paradigm applied to activity scheduling), and use discrete choice models for their analyses. To estimate parameters of such models, the choice probabilities and likelihood functions must be derived in order to calibrate the maximum likelihood estimators of the parameters. This requires the modeller to assume a finite and enumerable choice set, and universal across the population. This assumption is difficult to justify in the context of activity-based models: if one considers as schedule to be a discrete alternative subject to choice, then the choice set comprised of all possible combinations of this schedule is huge. Of course, individuals are only aware of a fraction of the full choice set, and actually consider an even smaller set when they make their scheduling decisions, after having discarded alternatives that they deem infeasible. The challenge for modellers is therefore to generate a choice set that is both realistic and appropriate to estimate parameters.

In this paper, we estimate the parameters of the utility-based activity-based model presented in Pougala *et al.* (2022). To do so, we apply the methodology to sample a choice set using the Metropolis-Hastings, as presented in Pougala *et al.* (2021). The process is illustrated by Fig. 1. We use data from the Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017) to calibrate two simple models, and discuss the results and their behavioural implications. Finally, we present further avenues that we are currently undergoing.

Figure 1: Full activity-based framework (Pougala *et al.*, 2022)



2 Background

Activity-based models originally emerged in the 1970s as a response to the shortcomings of traditional 4-step models (Vovsha *et al.*, 2005; Castiglione *et al.*, 2014), namely:

1. trips are the unit of analysis and are assumed independent, meaning that correlations between different trips made by the same individual are not accounted for properly within the model;
2. models tend to suffer from biases due to unrealistic aggregations in time, space, and within the population; and
3. space and time constraints are usually not included.

The early works of Hägerstraand (1970) and Chapin (1974) established the fundamental assumption of activity-based models, that the need to do activities drives the travel demand in space and time. Consequently, mobility is modelled as a multidimensional system rather than a set of discrete observations. Unlike traditional trip-based models, ABMs focus on overall behavioural patterns: decisions are analysed at the level of the household as opposed to seemingly independent individuals, and dependencies between events are taken into account (Timmermans, 2003; Pas, 1985). Specifically, modellers are interested in the link between activities and travel, often considered within a given timeframe. Typically, a single day is used as the unit of analysis. The resulting goal of studies in the literature is therefore to replicate as accurately as possible the interactions and considerations involved in the development of a daily schedule by an individual.

While the scheduling process is central to the activity-based research, there is no clear consensus on the representation and modelling of the daily scheduling process in utility-based frameworks. Typically, individuals are assumed to schedule activities by maximising the utility they can expect to gain. The timeframe is often introduced as a time budget that constrains the overall time expenditure. The scheduling decisions can be modelled as discrete choices: sequential discrete choice models consider a series of choices done consecutively with varying amounts of feedback between each step. On the other hand, joint models also integrate correlations between each aspect of the scheduling decision by evaluating them simultaneously. Other models do not consider the choice as fully discrete, but an hybrid consumption of discrete and continuous "goods".

Little work in the field of activity-based modelling specifically tackles estimation of model parameters. Usually, the parameters of the utility function are either estimated empirically by calibrating them to the available data or fixed to some arbitrary values (e.g. Charypar and Nagel (2005)). The issue with calibration on data is that the available travel surveys traditionally used for activity-based research are limited to revealed-preferences, and thus parameters that are inherently linked to the behaviour (such as penalties and preferences) cannot easily be derived.

Bowman and Ben-Akiva (2001) have estimated the parameters of each tier of their models (destination and mode choice, time of day, activity pattern model) simultaneously within the tier by estimating the corresponding discrete choice models (multinomial logit with alternative sampling for destination and mode choice, multinomial logit for time of day, nested logit for activity pattern), and sequentially across each tier. Specifically for the activity pattern model, they estimate different parameters for two subsets of the population (workers and non-workers). They find that the sequential estimation procedure leads to consistent parameters estimates but inconsistent standard errors. The estimates are also different than those obtained with simultaneous estimations. Another notable example is the work of Arentze *et al.* (2011) who estimate the parameters of a multiday dynamic activity generation model based on one-day travel observations. They consider a mixed logit framework, and have demonstrated that standard log-likelihood methods can be used provided the one-day choice probabilities can be derived as a function of the models parameters. They managed to estimate all of the model parameters, with the exception of the scale of the error term, which would still require longitudinal data in order to be properly estimated.

Allahviranloo and Axhausen (2018) use a bi-level optimisation model to estimate the mean and variance of the utility function for independent and joint activities, assuming these utility functions to be normally distributed across the population. They also estimate the penalties for schedule deviations (start time and duration). The upper level of their model has an objective function that maximises the accuracy of their estimation. The lower level is composed of the utility functions of all considered activities, also to be maximised. The objective functions of the lower level serve as constraints for the upper level. The value of the parameters are updated at each iteration using a Genetic Algorithm, and the model runs until one of the stopping criteria is reached. Their model allows them to successfully derive probability distributions for the mean and variance of the utility functions for each type of activity (i.e. school, work, leisure, etc.) and level of household interaction (independent vs. joint).

Although the parameters of utility functions are crucial elements of any models, methods to estimate them have not been greatly explored yet for activity-based applications. Therefore, there is a significant potential for contribution.

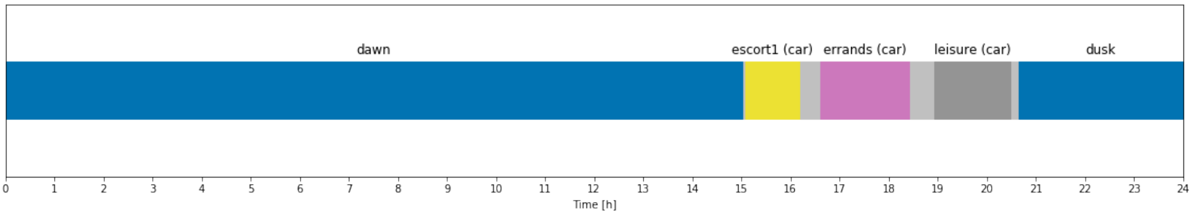
3 Model

We model the discrete choice of a full daily schedule for a given individual. A schedule (e.g. Fig. 2) is the outcome of one's decisions with respect to activity participation, activity location, activity scheduling, and transportation mode choice. It is defined as a sequence of *activities*, starting and ending at home, and lasting 24h. An activity a is characterised by:

- a location l_a where the activity can be performed,
- a start time x_a,
- a duration τ^a,
- a cost c_a for participating in the activity,
- an outbound trip, performed with mode of transportation m_a to the location of the next activity. For the last activity of the day, and activities taking place at the same location, the duration of this trip is 0.

The boundary conditions (start and end of the schedule at home), are modelled as two dummy activities "dawn" and "dusk".

Figure 2: Example output



Individuals are assumed to be rational and time sensitive, and to select the preferred schedules among all possible feasible schedules. The preferences of the individual for each schedule S are captured by a utility function U_S. The time sensitivity is modelled by considering scheduling preferences: a time interval when the agent prefers to start the activity: $[x_a^-, x_a^+]$, where $x_a^- \leq x_a^+$, and a range of desired durations $[\tau_a^-, \tau_a^+]$, where $\tau_a^- \leq \tau_a^+$.

We define the schedule utility U_S as the sum of a generic utility U associated with the whole schedule and utility components capturing the activity-travel behaviour:

$$U_{\S} = U + \sum_{a=0}^{A-1} (U_a^1 + U_a^2 + U_a^3 + \sum_{b=0}^{A-1} (U_{a,b}^4 + U_{a,b}^5)). \tag{1}$$

The components and the associated assumptions are defined as follows:

1. A generic utility U that captures aspects of the schedule that are not associated with any activity.
2. The utility U_a^1 associated with the participation of the activity a, irrespective of its starting time and duration.

$$U_a^1 = \beta_{\text{cost}} * c_a + \varepsilon_1 \tag{2}$$

3. the utility U_a^2 associated with starting time. This term captures the perceived penalty created by deviations from the preferred starting time.

$$U_a^2 = V_a^2 \tag{3}$$

with:

$$V_a^2 = \theta_a^e \max(0, x_a^- - x_a) + \theta_a^\ell \max(0, x_a - x_a^+), \tag{4}$$

where $\theta_a^e \leq 0$ and $\theta_a^\ell \leq 0$ are unknown parameters to be estimated from data. The first (resp. second) term captures the disutility of starting the activity earlier (resp. later) than the preferred starting time.

4. the utility U_a^3 associated with duration. This term captures the perceived penalty created by deviations from the preferred duration. Here, we illustrate this using a deterministic (dis)utility:

$$U_a^3 = V_a^3 \tag{5}$$

with:

$$V_a^3 = \beta_{a_a}^e \max(0, \tau_a^- - \tau_a) + \beta_a^\ell \max(0, \tau_a - \tau_a^+), \tag{6}$$

where $\beta_a^e \leq 0$ and $\beta_a^\ell \leq 0$ are unknown parameters to be estimated from data. Similarly to the specification of start time, the first (resp. second) term captures the disutility of performing the activity for a shorter (resp. longer) duration than the preferred one,

5. For each pair of locations (ℓ_a, ℓ_b), respectively, the locations of activities a and b with $a \neq b$, the utility $U_{a,b}^4$ associated with the trip from ℓ_a to ℓ_b, irrespective of the travel time. This term may include variables such as cost, level of service, etc. Here, we illustrate the framework with a specification involving travel cost. It may also include an error term, capturing the unobserved variables.

$$U_{a,b}^4 = \beta_{t,\text{cost}} * c_t + \varepsilon_4 \tag{7}$$

6. For each pair of locations (ℓ_a, ℓ_b), the utility $U_{a,b}^5$, which captures the penalty associated with the travel time from ℓ_a to ℓ_b. Here, it is assumed to be deterministic:

$$U_{a,b}^5 = V_{a,b}^5 \tag{8}$$

with

$$V_{a,b}^5 = \theta_t \rho_{ab}, \tag{9}$$

where θ_t is an unknown parameter to be estimated from data, and ρ_{ab} is the travel time to the next location.

Table 1 summarises the parameters of the problem, that will be estimated. Each parameter is associated with a given variable. Indices S, a, and n denote respectively a schedule, an activity and an individual.

Table 1: Parameters to be estimated

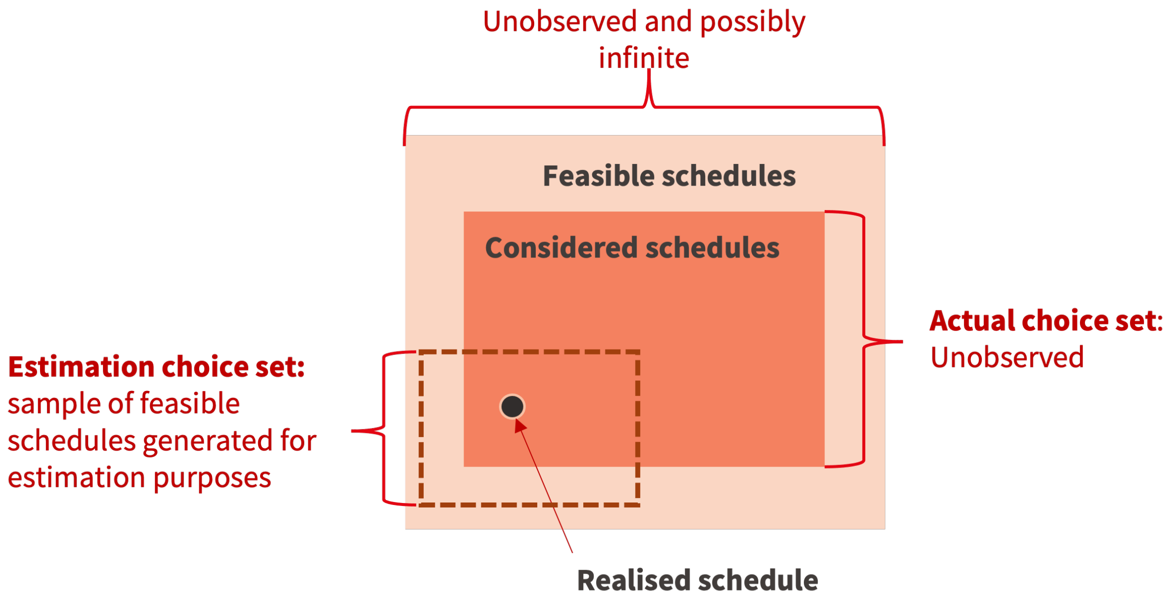
Parameter	Notation	Associated variable
Alternative-specific constants	$ASC_{S,n}$	-
Activity-specific constant	$ASC_{a,n}$	-
Cost of activity participation	β_{cost_a}	Cost c_a
Penalty start time (early)	θ_a^e	Deviation start time δ_{e,x_a}
Penalty start time (late)	θ_a^ℓ	Deviation start time δ_{ℓ,x_a}
Penalty duration (short)	β_s^e	Deviation duration δ_{s,τ_a}
Penalty duration (long)	β_a^ℓ	Deviation duration δ_{ℓ,τ_a}
Travel cost	$\beta_{t,\text{cost}}$	Cost c_t
Travel time	$\beta_{t,\text{time}}$	Time ρ_{ab}

3.1 Choice set

The universal choice set of the problem (all possible daily schedules for each individual) is combinatorial and cannot be enumerated. Therefore, we propose strategies to generate a choice set that can be used to estimate parameters. This choice set must include only feasible alternatives, regardless of whether they are actually considered by the individual or not, as illustrated in Fig. 3.

Using a smaller subset of alternatives allows to evaluate the maximum likelihood function, but biases the estimators of the parameters. In order to obtain unbiased parameters, Ben-Akiva and Lerman (1985) introduce an alternative specific correction term to the choice probability. Eq. (10) defines the probability that an individual n chooses an alternative $i_n \in C_n$, associated with a deterministic utility V_{in}. C_n is the choice set generated with the Metropolis-Hastings strategic sampling.

Figure 3: Choice sets in ABM (Pougala *et al.*, 2021)



$$P\left(i_n|\mathcal{C}_n\right) = \frac{\exp\left[V_{in} + \ln q(\mathcal{C}_n|i_n)\right]}{\sum_{j\in\mathcal{C}_n}\exp\left[V_{jn} + \ln q(\mathcal{C}_n|j)\right]} \tag{10}$$

The alternative-specific correction term $\ln q(\mathcal{C}_n|i_n)$ is the logarithm of the conditional probability of sampling the choice set C_n given that i_n is the chosen alternative. This formulation implies that if every alternative has equal probability of being chosen, $\ln q(\mathcal{C}_n|i_n) = 0$ and the estimation of the model on the subset is the same as the estimation on the full choice set (Frejinger *et al.*, 2009).

3.1.1 Random generation

Given a set A_n of activities considered by an individual n, N schedules are randomly generated following the procedure described in Algorithm 1. Each activity $a \in A_n$ is associated with a start time x_a and a duration τ_a, and a travel time $\rho_{(a,b)}$ between a pair of activities (a, b), where $b \neq a$. T is the time horizon (e.g. 24h).

The algorithm easily generates a given number of schedules, with uniform probabilities of

Algorithm 1 Random choice set generation

Choice set $C_n \leftarrow \emptyset$, final size N
Current start time $x_a \leftarrow 0$, Remaining time $\tau_r \leftarrow T$
Iteration $n \leftarrow 0$
while $n \leq N-1$ **do**
  while $\tau_r \geq 0$ **do**
    Choose *next* activity a^* from set A_n
    if first activity of schedule **then**
      Choose *next* start time x^*_a such that $x_a \leq x^*_a \leq T - \rho_{(a^*,\text{home})}$
    else
      Assign $x^*_a \leftarrow x_a + \tau_a + \rho_{(a,a^*)}$
    end if
    Choose *next* duration τ^*_a such that $\tau^*_a \leq \tau_r$
    Update $x_a \leftarrow x^*_a$, $\tau_r \leftarrow T - x^*_a - \tau^*_a$, $n \leftarrow n+1$
  end while
  Add schedule to C_n
end while
Add *chosen* schedule to C_n

selecting them. Given the size of the solution space, the limitation of the random generation is that there is a high chance of generating only schedules that would never be considered by the individual, and therefore uninformative for the estimation of parameters.

To overcome this challenge, we propose a second approach to generate choice sets with schedules that have high probabilities of being chosen by the individual based on strategic sampling with the Metropolis-Hastings algorithm.

3.1.2 Strategic sampling with Metropolis-Hastings

We apply the methodology described in Pougala *et al.* (2021) to strategically sample the alternatives from the universal choice set using the Metropolis-Hastings algorithm, and adding the chosen alternative to the final sample. Algorithm 2 summarises one iteration of the random walk.

Following Ben-Akiva and Lerman (1985), we define in Equation (11) the alternative specific corrective term for a choice set C_n of size $J+1$ with J unique alternatives. Each alternative j is sampled from the target distribution of the Metropolis-Hastings algorithm with probability q_{jn}, such that $q_{jn} = 0$ if $j \notin C_n$.

Algorithm 2 Choice set generation with Metropolis-Hastings

$n \leftarrow 0$, initialise state with random schedule $X_n \leftarrow S_0$
while $n \leq n_{iter}$ **do**
  Choose operator ω
  With probability P_ω, $X^* \leftarrow \textbf{Operator}(X_n)$
  Compute acceptance probability $\alpha(X_n, X^*) = \min\left(\frac{b(X^*)q(X_n|X^*)}{b(X_n)q(X^*|X_n)}\right)$
  With probability $\alpha(X_n, X^*)$, $X_{n+1} \leftarrow X^*$, else $X_{n+1} \leftarrow X_n$
end while

$$q(\mathcal{C}_n|i_n) = \frac{1}{q_{in}} \prod_{j \in C_n} \left(\sum_{j \in C_n} q_{jn} \right)^{J+1-J} \tag{11}$$

4 Case study

The Mobility and Transport Microcensus (MTMC) is a Swiss nationwide survey gathering insights on the mobility behaviours of local residents (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017). Respondents provide their socio-economic characteristics (e.g. age, gender, income) and those of the other members of their household. Information on their daily mobility habits and detailed records of their trips during a reference period (1 day) are also available. The 2015 edition of the MTMC contains 57'090 individuals, and 43'630 trip diaries. For this study, we estimate the parameters of a sample of schedules for the residents of Lausanne.

We consider three samples:

1. Workers, including full- and part-time workers,
2. Students,
3. All inhabitants, including workers and students.

For each of these samples, we calibrate two models:

1. Model 1, where we classify activities according to two levels of flexibility, and estimate the corresponding parameters for both categories.

2. Model 2, where we estimate all activity-specific parameters and constants, as defined in Table 1.

We consider 7 different activities: home, work, education, leisure, shopping (buying non-essential goods), errands and services (buying essential goods and groceries, or using services e.g. medical appointments, etc.), escort (accompanying someone to an activity). Following the definition of Pougala *et al.* (2022), travel is not considered as a standalone activity.

We make additional simplifications:

- We do not estimate travel parameters, and consider them null in Eq. (1),
- We do not estimate scheduling preferences (desired start time and durations). Instead, we use the modal start times and durations for each activity from the distribution across the full population of Lausanne (Table 2). Therefore, we consider them homogeneous across the population. Note that the *home* activity is not associated with a desired start time or duration.

Table 2: Modal times for Lausanne

Activity	Start time [hh:mm]	Duration [hh:mm]
Work	7:45	4:40
Education	7:32	2:19
Leisure	12:05	0:0
Shopping	12:05	0:05
Errands, services	14:05	0:05
Escort	18:00	0:0

We generate for each individual choice sets of sizes $N = 10, 100$ with the random generation and the Metropolis-Hastings algorithm.

4.1 Model 1

In the first model, we consider *activity-specific* constants, and *flexibility-specific* parameters. Each activity is classified in one of two possible degrees of flexibility $k \in \{\text{Flexible}, \text{Not Flexible}\}$. The assumption is that schedule deviations are penalised differently depending on the flexibility towards the activities. Intuitively, we expect non-flexible activities to have a more negative impact on the utility function of the schedule. For the sake of simplification, the initial implementation of model 1 ignores the travel component to only focus on activity parameters. Table 3 summarises the parameters estimated in the first model, for a total of 14 parameters. For identification purposes, the constant associated with the activity *home* is set to 0.

Table 3: Model 1: Parameters to be estimated

Parameter	Notation	Activities
Activity-specific constant	$ASC_{a,n}$	All *(reference: home)*
Penalty start time (early)	θ^e_{F}	Flexible (Errands, Leisure, Shopping, Home)
Penalty start time (late)	θ^ℓ_{F}	
Penalty duration (short)	β^s_{F}	
Penalty duration (long)	β^ℓ_{F}	
Penalty start time (early)	θ^e_{NF}	Not flexible (Education, Work, Escort)
Penalty start time (late)	θ^ℓ_{NF}	
Penalty duration (short)	β^s_{NF}	
Penalty duration (long)	β^ℓ_{NF}	

4.2 Model 2

In the second model, we consider *activity-specific* constants and schedule deviation parameters. The assumption is that schedule deviations are penalised differently for each activity. Similarly to the previous model, we focus only on activity parameters and do not estimate travel penalties. Table 4 summarises the parameters estimated in the second model, for a total of 30 parameters. For identification purposes, the constant and penalties associated with the activity *home* are set to 0.

Table 4: Model 2: Parameters to be estimated

Parameter	Notation	Activities
Activity-specific constant	$ASC_{a,n}$	All (*reference: home*)
Penalty start time (early)	θ^e_{F}	
Penalty start time (late)	θ^ℓ_{F}	
Penalty duration (short)	β^s_{F}	
Penalty duration (long)	β^ℓ_{F}	

4.3 Results

The model estimates are reported in the appendix, Appendix A. We only report the estimates of statistically significant parameters, at a significance level of 5%. In addition, for model 2 (Table 8 to Table 10), we only report estimates for the activities *work*, *education* (non-flexible activities) and *leisure* (flexible activity).

4.3.1 Model 1

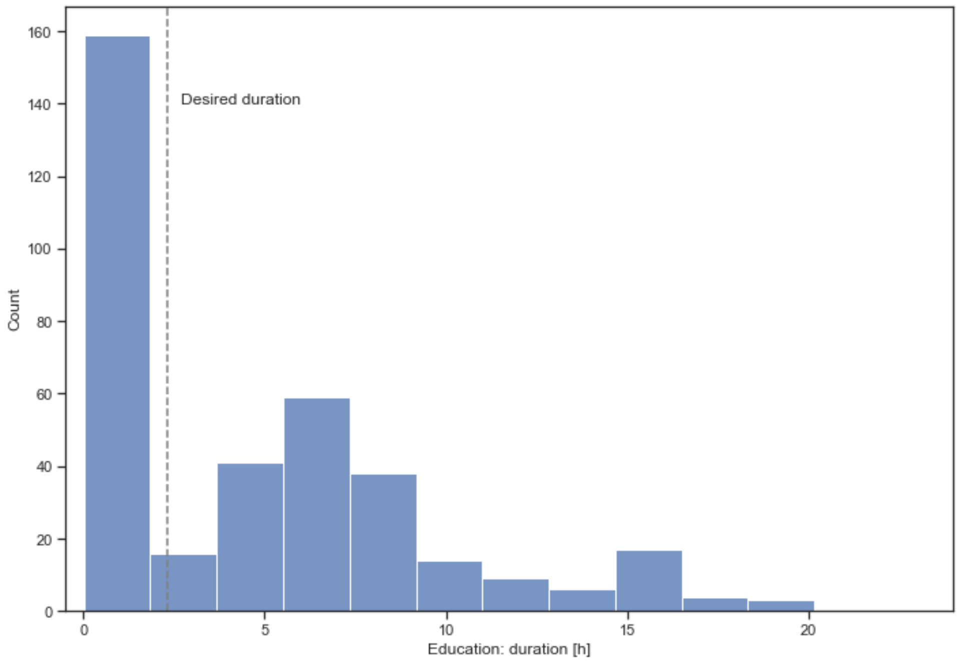
For all three samples, we obtain sensible results: all activity-specific constants are positive as compared to staying at home. It translates the fact that, all else being equal, individuals prefer performing activities rather than staying at home[1] This can also be explained by the specification of the utility function which only include parameters that are expected to be negative (e.g. penalties for schedule deviations).

As expected, all penalties are negative. For workers, the ratios between early and late penalties are in line with the literature of departure time choice (e.g. (Small, 1982)): being late is more penalised than being early, both for flexible and non-flexible activities. For duration, flexible activities running for longer than desired are penalised, but the parameter for shorter durations is negligible. On the other hand, when the activity is not flexible, shorter durations than expected are penalised about 4 times more than longer. This behaviour is also observed among students, although they tend to not penalise flexible activities starting earlier (the parameter is not statistically significant). With the exception of the penalty for short durations, the penalties associated with students are usually smaller in magnitude than for workers. The values of the constants are comparable.

Interestingly, when we look at the entire Lausanne sample, late activities are slightly less penalised than early, for both flexible and non-flexible activities. Shorter activities are also more penalised than longer ones.

[1] Note that here we interpret *home* as an absence of activity, and do not take into account activities performed at home. This assumption can be relaxed with the condition of a more exhaustive dataset with records of in-home activities, such as time use surveys.

Figure 4: Distribution of duration for education in generated students sample

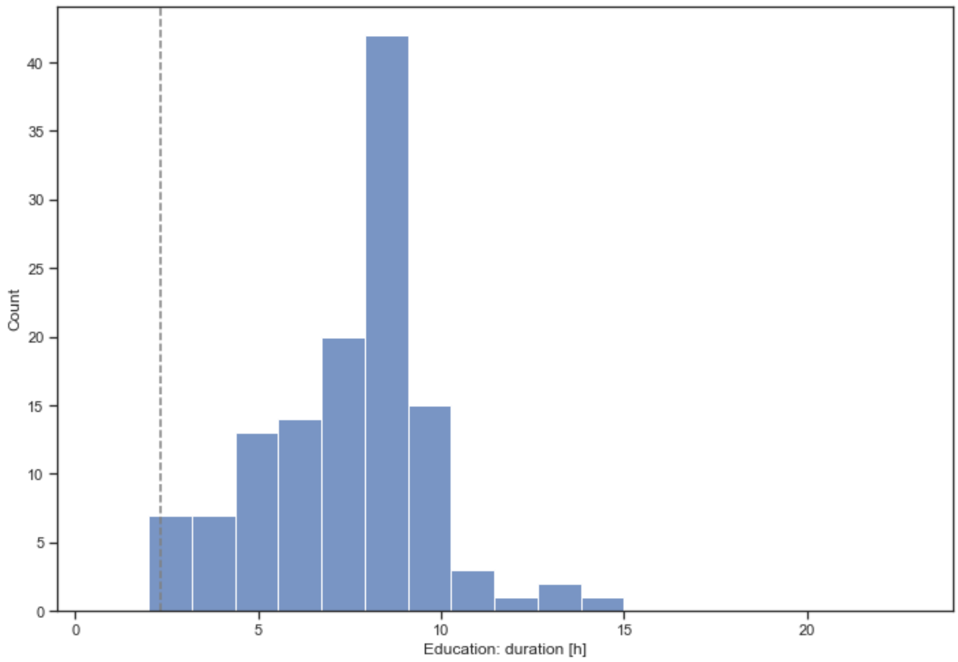


4.3.2 Model 2

In the second model, we include activity-specific penalties in addition to the constants. The *education* activity for both students and workers is very highly penalised when the duration is shorter than the preference. Given the magnitude difference with the other activities, this seems to be indicative of an imbalance in the choice set. As illustrated in Fig. 4, in most of the schedules where education is present, its duration is shorter than the input desired duration, although very close to it. By comparing it with the real distribution (Fig. 5), we can see that the mean duration is much longer. These results highlight the limitation of using homogeneous and fixed desired times across the entire population.

The same phenomenon can be observed in the parameters of the *workers* sample, but we seem to obtain more stable estimates in the full Lausanne sample. In the latter, the penalties for start times (resp. duration) are almost symmetrical: being early or late (resp. shorter or longer) are equally penalised. This is not the case for the sub-populations of workers and students, for whom the asymmetry is observed, especially for start time. As previously mentioned, we can assume that the asymmetry for the duration penalties is mainly affected by the bias in the choice set.

Figure 5: Distribution of duration for education in real students sample



5 Discussion and further work

In this paper, we have presented a procedure to estimate the parameters of the activity-based model introduced in Pougala *et al.* (2022). The process includes the generation of a choice set for parameter estimation, with a good variety of alternatives to ensure unbiased and stable parameter estimates, and with tractable sample probabilities. We have applied our methodology on a simple, time-dependent and linear in parameters utility function, and yielded statistically significant and behaviourally sensible results even with a small number of alternatives in the choice set (10). These results are very promising, and the model can be further improved by:

- Considering desired timings for each socio-economic category, and generate choice sets accordingly. Assuming these variables to be randomly distributed instead of fixed would also allow for greater flexibility.
- Investigating a more complex utility specification by adding non-linear relationships. For instance, the utility function defined by Feil (2010), which includes a non-linear effect of activity duration can be considered.
- Similarly, more variables can be added to explain the choice of schedules - such as socio-economic characteristics.
- Investigating more complex model structures than multinomial logit models. For instance, the mixed logit model (e.g. by considering randomly distributed parameters

across the population) seems well suited to the problem.

Finally, we have not discussed the matter of validation of the parameters, which is complex and limited without dedicated resources (e.g. stated-preference survey). Calibrating the model on a synthetic population would allow to evaluate the performance of the model with know control variates.

6 References

Adler, T. and M. Ben-Akiva (1979) A theoretical and empirical model of trip chaining behavior, *Transportation Research Part B: Methodological*, **13** (3) 243 – 257, ISSN 0191-2615.

Allahviranloo, M. and K. Axhausen (2018) An optimization model to measure utility of joint and solo activities, *Transportation Research Part B: Methodological*.

Arentze, T., D. Ettema and H. Timmermans (2011) Estimating a model of dynamic activity generation based on one-day observations: Method and results, *Transportation Research Part B: Methodological*, **45** (2) 447–460, feb 2011.

Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete choice analysis : theory and application to travel demand*, Mit Press series in transportation studies 9, the MIT Press, Cambridge, ISBN 9780262536400.

Bowman, J. L. and M. E. Ben-Akiva (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A: Policy and Practice*, **35** (1) 1–28, jan 2001.

Castiglione, J., M. Bradley and J. Gliebe (2014) *Activity-Based Travel Demand Models: A Primer*, Transportation Research Board, Washington, D.C., mar 2014.

Chapin, F. S. (1974) *Human activity patterns in the city: Things people do in time and in space*, vol. 13, Wiley-Interscience.

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

Feil, M. (2010) Choosing the Daily Schedule: Expanding Activity-Based Travel Demand Modeling, Ph.D. Thesis, ETH Zürich.

Frejinger, E., M. Bierlaire and M. Ben-Akiva (2009) Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological*, **43** (10) 984–994.

Hägerstraand, T. (1970) What About People in Regional Science?, *Papers in Regional Science*.

Office fédéral de la statistique and Office fédéral du développement Territorial (2017) Comportement de la population en matière de transports. Résultats du microrecensement mobilité et transports 2015, *Technical Report*, Neuchâtel, Berne.

Pas, E. (1985) State of the art and research opportunities in travel demand: Another perspective, *Transportation Research Part A: General*, **19** (5-6) 460–464.

Pougala, J., T. Hillel and M. Bierlaire (2021) Choice set generation for activity-based models, Ascona, Switzerland.

Pougala, J., T. Hillel and M. Bierlaire (2022) Capturing trade-offs between daily scheduling choices, *Journal of Choice Modelling*, **43** (100354). Accepted on Mar 19, 2022.

Small, K. A. (1982) The scheduling of consumer activities: work trips., *American Economic Review*.

Timmermans, H. (2003) Recent and expected directions in modelling transportation, *SATC 2003*.

Vovsha, P., M. Bradley and J. Bowman (2005) Activity-based travel forecasting models in the united states: progress since 1995 and prospects for the future, paper presented at the *EIRASS Conference on Progress in Activity-Based Analysis*.

A Model estimates

A.1 Model 1

Table 5: Estimation results for Model 1 on student population

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	F late	-0.474	0.108	-4.38	1.21e-05
2	F long	-0.48	0.145	-3.3	0.951e-04
3	NF early	-0.217	0.118	-1.84	0.0665
4	NF late	-0.447	0.238	-1.88	0.0605
5	NF short	-4.04	1.86	-2.17	0.0303
6	ASC_Education	7.77	1.54	5.03	4.92e-07
7	ASC_Errands	4.79	1.46	3.27	0.00106
8	ASC_Escort	4.97	1.46	3.41	0.651e-04
9	ASC_Leisure	10.4	1.57	6.62	3.54e-11
10	ASC_Shopping	5.86	1.0	5.84	5.33e-09
11	ASC_Work	3.86	1.44	2.69	0.00725

Summary statistics
Number of observations=198
$L(0) = -651.9881$
$L(\hat{\beta}) = -273.1573$
$\bar{\rho}^2 = 0.56$

Table 6: Estimation results for Model 1 on worker population

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	F early	-0.813	0.16	-5.09	3.53e-07
2	F late	-1.12	0.138	-8.08	6.66e-16
3	F long	-0.569	0.165	-3.45	0.554e-04
4	NF early	-0.827	0.160	-5.15	2.58e-07
5	NF late	-1.26	0.236	-5.31	1.08e-07
6	NF long	-0.789	0.229	-3.45	0.57e-04
7	NF short	-3.24	0.555	-5.84	5.30e-09
8	ASC_Education	10.8	2.50	4.33	1.50e-05
9	ASC_Errands	7.63	1.28	5.97	2.32e-09
10	ASC_Escort	9.79	1.45	6.77	1.31e-11
11	ASC_Leisure	15.3	1.38	11.1	0.0
12	ASC_Shopping	12.5	1.38	9.05	0.0
13	ASC_Work	18.5	2.00	9.28	0.0

Summary statistics
Number of observations=528
$L(0) = -2030.859$
$L(\hat{\beta}) = -462.2077$
$\bar{\rho}^2 = 0.77$

Table 7: Estimation results for Model 1 on Lausanne sample

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	F early	-1.01	0.0858	-11.8	0.0
2	F late	-0.722	0.0596	-12.1	0.0
3	F long	-0.433	0.165	-3.45	0.554e-04
4	F short	-9.79	1.45	-6.74	1.56e-11
5	NF early	-1.47	0.148	-9.96	0.0
6	NF late	-1.07	0.105	-10.2	0.0
7	NF long	-1.36	0.136	-9.99	0.0
8	NF short	-1.46	0.154	-9.49	0.0
9	ASC_Education	14.1	0.948	14.8	0.0
10	ASC_Errands	7.27	0.625	11.6	0.0
11	ASC_Escort	10.2	0.76	13.5	0.0
12	ASC_Leisure	14.5	0.629	23.0	0.0
13	ASC_Shopping	11.6	0.529	22.0	0.0
14	ASC_Work	16.5	0.854	19.3	0.0

Summary statistics
Number of observations=909
$L(0) = -8093.839$
$L(\hat{\beta}) = -2610.456$
$\bar{\rho}^2 = 0.68$

A.2 Model 2

Table 8: Estimation results for Model 2 on student population

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	Education: ASC	17.2	4.54	3.78	1.58e-04
2	Education: early	-1.89	0.781	-2.43	0.0153
3	Education: late	-2.91	0.948	-3.07	0.00216
4	Education: long	-0.195	0.132	-1.48	0.140
5	Education: short	-146.0	50.9	-2.87	0.0604
6	Leisure: ASC	15.3	4.48	3.41	6.62e-04
7	Leisure: late	-1.88	1.02	-1.84	0.0658
8	Leisure: long	-0.828	0.482	-1.72	0.086
9	Leisure: short	-3.52	1.97	-1.79	0.074
10	Work: ASC	8.45	2.17	3.90	9.82e-05
11	Work: early	-0.684	0.217	-3.15	0.00163
12	Work: late	-6.92	1.96	-3.53	4.19e-04
13	Work: long	0.177	0.065	2.72	0.0066
14	Work: short	-10.4	3.70	-2.82	0.00479

Summary statistics
Number of observations=198
$L(0) = -651.9881$
$L(\hat{\beta}) = -172.0351$
$\bar{\rho}^2 = 0.69$

Table 9: Estimation results for Model 2 on worker population

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	Education: ASC	51.4	10.3	4.99	6.17e-07
2	Education: early	-32.9	7.88	-4.17	3.02e-05
3	Education: late	-7.56	1.72	-4.40	1.09e-05
4	Education: long	-3.96	0.921	-4.30	1.71e-05
5	Education: short	-29.6	5.81	-5.09	3.52e-07
6	Leisure: ASC	16.1	2.30	7.00	2.49e-12
7	Leisure: early	-1.67	0.621	-2.69	0.0072
8	Leisure: late	-1.35	0.248	-5.43	5.64e-08
9	Leisure: long	-0.193	0.0958	-2.01	0.0444
10	Work: ASC	23.6	4.77	4.94	7.70e-07
11	Work: early	-2.48	0.683	-3.64	2.74e-04
12	Work: late	-1.58	0.452	-3.49	4.87e-04
13	Work: long	-1.49	0.575	-2.59	0.00963
14	Work: short	-3.51	1.32	-2.67	0.00758

Summary statistics

Number of observations=528

$L(0) = -2030.859$

$L(\hat{\beta}) = -390.7576$

$\bar{\rho}^2 = 0.79$

Table 10: Estimation results for Model 2 on Lausanne sample

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	Education: ASC	17.4	1.44	12.1	0.0
2	Education: early	-2.22	0.561	-3.96	7.50e-05
3	Education: late	-2.00	0.334	-6.00	2.01e-09
4	Education: long	-1.58	0.255	-6.17	6.74e-10
5	Education: short	-1.86	0.212	-8.80	0.0
6	Leisure: ASC	13.5	0.709	19.0	0.0
7	Leisure: early	-0.942	0.111	-8.52	0.0
8	Leisure: late	-0.59	0.055	-6.8	1.02e-11
9	Leisure: long	-0.271	0.055	-4.93	8.22e-07
10	Leisure: short	-8.62	1.57	-5.47	4.44e-08
11	Work: ASC	17.6	1.23	14.3	0.0
12	Work: early	-1.49	0.3	-4.95	7.41e-07
13	Work: late	-1.39	0.224	-6.23	4.56e-10
14	Work: long	-1.7	0.205	-8.32	0.0
15	Work: short	-1.34	0.179	-7.5	6.28e-14

Summary statistics
Number of observations=909
$L(0) = -8093.839$
$L(\hat{\beta}) = -2313.796$
$\bar{\rho}^2 = 0.71$