Modelling the impact of activity duration on utility-based scheduling decisions: a comparative analysis

Janody Pougala[1], Tim Hillel[2], and Michel Bierlaire[1]

[1]Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2]Behaviour and Infrastructure Group, University College London, United Kingdom

Short summary

There exist two major categories of activity-based models: on one side, utility-based or econometric models are founded on the principles of random utility maximisation, and use discrete choice modelling techniques to solve activity-based problems. On the other hand, rule based approaches refute the assumption that decision-makers are perfect optimisers and their activity-travel behaviour is the product of context-dependent rules. Recently hybrid models (e.g. OASIS) combine both approaches, to keep the flexibility and theoretical robustness of utility-based models, with the addition of spatio-temporal constraints which increase the behavioural realism and simplify the estimations. However, hybrid models suffer from two main issues: specifying a utility function that accurately reflects the decision-making process, and estimating parameters in a highly complex space due to the constraints. In this paper, we answer these questions within the context of the OASIS framework (Pougala et al., 2022). We first estimate the parameters of two state-of-the-art utility functions (Charypar & Nagel, 2005; Feil, 2010), using data from the Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017) and compare them with OASIS' default linear-in-parameters utility function of the model.
Keywords: activity-based modelling, discrete choice modelling, parameter estimation

1 Introduction

Activity-based models (ABMs) stem from the fundamental assumption that travel demand is derived from the need to perform activities, rooted in a spatiotemporal context, and influenced by personal and environmental factors. By focusing on individuals and explicitly considering these interactions, ABMs aim to be more behaviourally realistic than traditional trip-based models, and to provide more flexible and targeted insights on individual mobility. Successful applications of activity-based models have demonstrated the added value of shifting the focus to individuals and their activities. Two main approaches can be cited: utility-based and rule-based models. The scheduling process is a result of random utility maximisation for the former, and the satisfaction of a set of spatio-temporal rules for the latter. While historically both approaches were considered contradictory, there has been increased research in hybrid models which combine elements from either theory. One limitation of hybrid approaches can be the significant complexity of the models, which require both a robust utility specification able to reflect the dynamics and interactions involved in the decision-making process, and informative constraints that can efficiently reduce the solution space without impacting the completeness of the results. We present the case of the OASIS framework, an integrated framework to simulate daily activity schedules by considering all choice dimensions (activity participation, timing decisions, mode and location choice) simultaneously. The model was designed to accommodate any utility specification and constraints depending on the context, but current implementations have relied on many simplifying assumptions both for the utility specification and formulation of constraints, mainly to compensate for the lack of available data. As a result, the utility function is too simple to properly capture all facets of activity-travel behaviour, which limits possible extensions of the model (including multiday or multiperson scheduling).

The formulation of behaviourally realistic utility functions has been the focus of many works in utility-based models for activity-travel behaviour. In particular, authors have been interested on the impact of timing (start time and duration) on the utility of performing an activity. One common assumption, derived from studies on departure time (Small, 1982), is that individuals have preferences regarding the timings of their activities, and penalise deviations from such preferences (e.g. Pougala et al., 2022; Charypar & Nagel, 2005; Allahviranloo & Axhausen, 2018). These deviations can be asymmetrical (e.g. being late is more penalised than being early). Other authors consider these preferences more implicitly: for example, Joh et al. (2005) formalise an S-shaped utility for duration, which captures the behavioural assumption that a frustration (e.g. being involved in an activity for a duration too short) and satiation (e.g. the activity duration is too long) effects exist. This formulation always considers an increasing utility function, but at different rates depending on whether the threshold of satiation has been reached. This is not the case for formulations that consider explicit scheduling preferences, where the utility (of activity duration) can decrease. Joh et al.'s S-shaped function was adopted by multiple authors, such as Feil (2010) and Ettema & Timmermans (2003). An important gap in the literature is the calibration of utility parameters to observed data, either due to lack of data or due to the complexity of the model.

In this paper, we investigate different utility specifications for activity participation and timing, to be used within the OASIS framework. The specifications presented here are: the OASIS default utility function and two utility functions (Charypar & Nagel, 2005; Feil, 2010) used as scoring functions in the state-of-the-art agent-based microsimulator MATSim. We use the estimation component of the OASIS framework to estimate the parameters of all three utility specifications, using the Mobility and Transport Microcensus (MTMC), a nationwide travel survey of Switzerland (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017). The purpose of this investigation is two-fold:

1. first, to test our methodology for parameter estimation on state-of-the-art utility functions,
2. to empirically investigate the differences of these utility specifications, and their implications in terms of behaviour.

.

This paper is organized as follows. In Section 2, we present the OASIS framework, and in particular the methodology to estimate the parameters of the model based on Metropolis-Hastings sampling. We apply this procedure on the three utility functions described, using a sample of the MTMC data. Finally, we discuss key differences between the specifications, and propose some insights for future investigations.

2 Methodology

OASIS framework

In this section, we briefly introduce the OASIS framework (Pougala et al., 2022). The framework (Fig. 1) is composed of two main elements:

1. A simulation model that outputs distributions of feasible schedules for given individuals, based on a mixed-integer optimisation model where the objective function is the utility of the schedule.
2. An estimation component to calibrate the parameters of the said utility function.

Parameter estimation

In the OASIS framework, the estimation of parameters is two-fold:

1. For a given individual n, generate a set of K feasible alternatives $\tilde{C_n} = \{S_0, ..., S_{k-1}, S_k\}$ which includes the chosen schedule $S_k = S^*$. The non-chosen schedules $S_i \forall i \neq k$ are sampled using the Metropolis-Hastings algorithm (Algorithm 1), where the target distribution is proportional to the utility function of the problem. The parameters of the target distribution are estimated on a random choice set:
   (a) Begin with an initial schedule S_0
   (b) At each iteration i, propose a candidate state S_{new} by modifying the current schedule in one dimension (activity participation, timing, or travel).

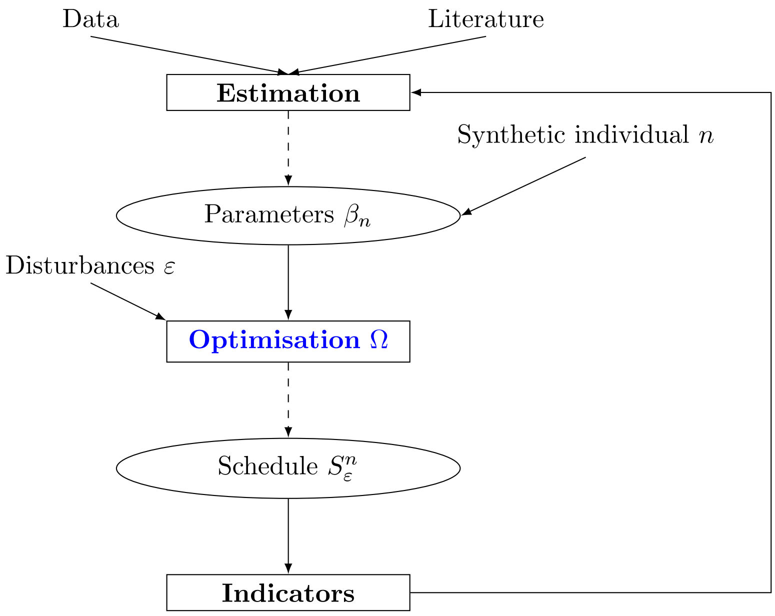

Figure 1: OASIS framework

(c) Accept or reject S_{new} based on the predefined acceptance probability.

2. Considering the choice probability in Eq.1, the maximum likelihood estimators of the parameters $\widehat{\beta}$ are derived from the corresponding likelihood function.

$$P_{in} = P_n(i|\tilde{C}_n) = \frac{e^{\mu V_{in} + \ln P_n(\tilde{C}_n|i)}}{\sum_{j\in\tilde{C}_n} e^{\mu V_{jn} + \ln P_n(\tilde{C}_n|j)}} \tag{1}$$

Given that the parameters are estimated on a sample of alternatives $\tilde{C}_n$, the choice probability is corrected with a term $\ln P_n(\tilde{C}_n|j)$, in order to obtain unbiased estimators (Ben-Akiva & Lerman, 1985). This term corresponds the sampling probability of the choice set $\tilde{C}_n$, and is directly obtained from the Metropolis-Hastings routine.

Algorithm 1 Choice set generation for the ABM with Metropolis-Hastings

$t \leftarrow 0$, initialise state with random schedule $X_t \leftarrow S_0$
Initialise utility function with random parameters $\tilde{U_S}$
for $t = 1, 2, ...$ **do**
  Choose operator ω with probability P_ω
  $X^*, q(X_t, X^*) \leftarrow$ **ApplyChange**(ω, X_t)
  function ApplyChange(ω, state X)
    return new state X', transition probability q(X, X')
  end function
  Compute target weight $p(X^*) = U_S\tilde{(}X^*)$
  Compute acceptance probability $\alpha(X_t, X^*) = \min\left(\frac{p(X^*)q(X_t|X^*)}{p(X_t)q(X^*|X_t)}\right)$
  With probability $\alpha(X_t, X^*)$, $X_{t+1} \leftarrow X^*$, else $X_{t+1} \leftarrow X_t$
end for

Simulation

The scheduling process for a given individual n, a set of activities to be scheduled A_n, and sets of possible modes M_n and $L_{a,n}$ is summarised in 2. The set of estimated parameters $\beta_{a,n}$ is also provided as input.

The objective function of the maximisation problem is the utility function of the schedule. This function is expressed as a sum of a deterministic element and an error term (Eq.2). We assume the distribution of the error term to be known.

$$U_S = V_S + \varepsilon_S \tag{2}$$

The decision variables (activity participation, start time, duration, sequence,...) are chosen such as to maximise the utility of the schedule subject to constrained. For a draw of the error term ε_S^r, the maximisation problem becomes deterministic, and yields one optimal schedule S^r which is a draw from the distribution of schedules for n.
The constraints of the optimisation problem can be classified in two categories:

1. Mathematical constraints: they ensure that the resulting schedule is valid. For example, the time budget T cannot be exceeded, activities cannot overlap,...
2. Context-dependent constraints: they are additional rules that influence the activity-travel behaviour. For example, some activities must take precedence over others (e.g. picking up children from school must happen after they were picked up),...

Algorithm 2 Simulation of activity schedules

Initialise n, β_n, A_n, M_n, L_n
for r = 1,2,...,R **do**
  Draw ε_S^r from distribution of error terms.
  Draw schedule S_n^r by solving $\Omega = \max U_S(X_n, \beta_n, \varepsilon_S^r)$ s.t. constraints
end for

Utility specification

For an individual n, each activity a provides a utility U_{a_n}, composed of the following elements:

1. A *participation* term, which is constant with respect to time.
2. A utility with respect to *activity start time*.
3. A utility with respect to *activity duration*.
4. Utility terms with respect to *travel*, considering the influence of travel time and cost to the activity.

In this paper, we test different specifications of the utility function, more specifically, the utility terms associated with start time and duration.

Default OASIS utility function

In the default OASIS utility function (Eq.3), the influences of start time x_a and duration τ_a are considered by penalising deviations from preferred start time x_a^* (earlier or later starts are penalised, see Eq. 4) and duration τ_a^* (shorter or longer durations are penalised, see Eq. 5). For each activity, we therefore estimate four parameters corresponding to the penalties: $\{\theta_{\text{early}}, \theta_{\text{late}}, \theta_{\text{short}}, \theta_{\text{long}}\}$, as well as activity-specific constants. The assumption is that individuals have asymmetrical penalties for positive and negative deviations from their preferences[1], and penalise differently each activity (implying different ranges of flexibility).
Further assumptions are taken for the *home* activity:

1. The start time is constrained: the day must start at home. There is therefore no utility associated with start time.
2. The duration of the *home* activity is not associated with a preference, but results from the other scheduled activities. There is therefore no utility associated with duration.

[1] The preferred start time and duration can be a fixed point or a continuous range.

$$U_S = U + \sum_{a=0}^{A-1}(U_a^{\text{participation}} + U_a^{\text{start time}} + U_a^{\text{duration}} + \sum_{b=0}^{A-1} U_{a,b}^{\text{travel}}) \quad (3)$$

$$U_a^{\text{start time}} = \theta_a^{\text{early}} \max(0, x_a^* - x_a) + \theta_a^{\text{late}} \max(0, x_a - x_a^*) + \varepsilon_{\text{start time}} \quad (4)$$

$$U_a^{\text{duration}} = \theta_a^{\text{short}} \max(0, \tau_a^* - \tau_a) + \theta_a^{\text{long}} \max(0, \tau_a - \tau_a^*) + \varepsilon_{\text{duration}} \quad (5)$$

MATSIM scoring function

The utility function used in MATSIM was formalised by Charypar & Nagel (2005). The utility of activity duration has a logarithm form (Eq. 7), which implies a decreasing marginal utility. In addition, a too short duration is penalised. For start time (Eq. 8), schedule deviations such as being late or early are penalised.
The parameters are: a parameter common to all activities β_{act}, a typical duration τ_a^* (considered known), a scaling factor A and a priority term ρ. $\beta^{\text{short}}, \beta^{\text{early}}, \beta^{\text{late}}$ penalise schedule deviations (δ).

$$U_S = U \sum_{a=0}^{A-1}(U_a^{\text{duration}} + U_a^{\text{start time}} + U_a^{\text{travel}}) \quad (6)$$

$$U_a^{\text{duration}} = \max\left[0, \beta_{\text{act}} \tau_a^* \ln\left(\frac{\tau_a}{\tau_a^* \exp(-A/(\rho\tau_a^*))}\right)\right] + \beta_a^{\text{short}} \delta_a^{\text{short}} \quad (7)$$

$$U_a^{\text{start time}} = \beta_a^{\text{early}} \delta_a^{\text{early}} + \beta_a^{\text{late}} \delta_a^{\text{late}} \quad (8)$$

PlanomatX utility function

We test the utility specification proposed by Feil (2010), which is a modification of the MATSIM utility function (Section 2). The utility function considers the impact of activity duration with an asymmetric S-shaped curve with an inflection point, as formalised by Joh et al. (2005) (Eq. 10). The parameters of the S-shape are: the inflection point α_a, the slope β_a, and the relative vertical position of the inflection point γ_a. When $\gamma_a = 1$, α_a can be considered as the duration where the utility reaches its maximum. They do not consider start time in their utility function.

$$U_S = \sum_{a=0}^{A-1}(U_a^{\text{act}} + U^{\text{travel}}) \quad (9)$$

$$U_a^{\text{act}} = U_a^{\text{min}} + \frac{U_a^{\text{max}} - U_a^{\text{min}}}{(1 + \gamma_a \exp \beta_a [\alpha_a - \tau_a])^{1/\gamma_a}} \quad (10)$$

3 Results and discussion

Case study

We use the Mobility and Transport Microcensus (MTMC), 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. In order to illustrate a real-life application of the simulator, we focus on the sample of full-time students residing in Lausanne (236 individuals).

Following the methodology described in Section 2, we start by generating the choice sets of daily schedules for each individual in the sample. Each choice set is composed of 10 alternatives, including the chosen (recorded) schedule.

The models are estimated with PandasBiogeme (Bierlaire, 2020). The estimation process is done using 70% of observations in the sample data, where one observation is the daily schedule of one individual.

Estimation results

Tables 1, 2, and 3 present the parameter estimates for the OASIS utility function, the MATSIM scoring function, and the PlanomatX function, respectively. We have chosen to display only significant parameters at a 5% level. For the estimation of the MATSIM function, we have considered the same assumptions as described by Charypar & Nagel (2005) for the values of the scaling parameter ($A = -200$) and the priorities for each activity ($\rho_a = 1$ for $a \in \{\text{home, education, work}\}$ and $\rho_a = 3$ otherwise).

Similarly, we have assumed for the estimation of the PlanomatX function that $U_a^{\min} = 0$ and $\gamma_a = 1 \ \forall a$, as described by Feil (2010).

Finally, given that in the OASIS context, *home* is the reference alternative and therefore associated with a null utility, we do not have estimated any parameter for this activity. The magnitudes and signs of the other coefficients should therefore be considered relative to the home baseline.

OASIS

For *education*, being early seems to be slightly more penalised than being late, although the penalties are almost symmetrical. For duration, cutting the activity short is associated with a negative penalty, whereas a long duration is not regarded negatively (the associated parameter is not significantly different from 0). Surprisingly, for *work* the penalty for being late is not statistically significant, while being early or a short duration are significantly penalised.

For *leisure*, only being late is penalised. This can be explained by the fact that, while leisure is usually considered as a discretionary activity, it is likely constrained by the participation of other individuals or feasible times (e.g. opening hours of facilities The penalty for a short duration for both leisure and shopping is not significant, which implies that these activities are more flexible than education or work for scheduling trade-offs.

Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
Education: ASC	7.62	1.26	6.04	1.55e-09
Education: early	-1.15	0.282	-4.07	4.62e-09
Education: late	-0.89	0.214	-4.15	3.28e-05
Education: short	-0.452	0.23	-1.97	0.0493
Leisure: ASC	5.02	0.679	7.38	1.57e-13
Leisure: late	-0.747	0.135	-5.54	3.1e-08
Leisure: long	-0.137	0.0497	-2.75	0.00593
Shopping: ASC	5.04	0.807	6.25	4.07e-10
Shopping: early	-0.652	0.144	-4.53	5.88e-06
Shopping: late	-0.534	0.0944	-5.65	1.57e-08
Shopping: long	-0.17	0.06	-2.84	0.00456
Work: ASC	4.34	1.53	2.84	0.00448
Work: early	-0.71	0.223	-3.19	0.00145
Work: short	-1.37	0.481	-2.86	0.00423
Summary statistics				
$L(0) = -454.1869$				
$L(\hat{\beta}) = -152.0466$				
$\bar{\rho}^2 = 0.621$				

Table 1: Estimation results for OASIS utility function. Only statistically significant parameters were included.

MATSIM

A similar behaviour is implied by the estimates of these parameters. For *education* both start time deviations are penalised (being early slightly more than being late) in comparable magnitudes. Being early at a leisure activity is not associated with a statistically significant penalty, as opposed to being late. For *work*, we have once again an insignificant parameter for being late.

Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
β_{act}	0.0514	0.00974	5.27	1.34e-07
Education: early	-1.6	0.449	-3.57	0.00036
Education: late	-1.01	0.291	-3.48	0.00051
Leisure: late	-0.467	0.122	-3.84	0.00012
Shopping: early	-0.476	0.119	-4.01	6.04e-05
Shopping: late	-0.293	0.0842	-3.48	0.00049
Work: early	-2.75	0.712	-3.87	0.000111
Work: short	-1.59	0.493	-3.22	0.00126
Summary statistics				
$L(0) = -593.8925$				
$L(\hat{\beta}) = -248.568$				
$\bar{\rho}^2 = 0.56$				

Table 2: Estimation results for MATSIM utility function. Only statistically significant parameters were included.

PlanomatX

On the other hand, the parameters estimates for the S-shaped utility function are more difficult to interpret, especially the values of the inflection point α, which is the duration when the utility function reaches its maximum. For *education* and *work*, this parameter is around 2 hours, which means that beyond this duration, the utility increases at a decreasing rate (satiation effect). The fact that longer durations are usually scheduled for these activities suggests that the time allocation for education and work is more constraint-driven than utility-driven. For *shopping*, we observe the opposite. The inflection point is at a very high duration as compared to the typical values in the dataset. However, the negative slope suggests a decreasing utility.

Comparison of utilities

Figures 2-5 illustrate the utilities as functions of activity duration for education, work, leisure and shopping.

Given the similarity of their specifications, the OASIS and MATSIM utility functions have comparable trends for the *education* and *work* activities. The OASIS utility seems to converge towards the PlanomatX utility for very long durations.

For *leisure* and *shopping*, the results are more heterogeneous. For the MATSIM specification, the impact of duration seems negligible as the utility varies very little with duration. For these two activities the OASIS function and PlanomatX are in concordance: about the inflection point for leisure, and the overall decreasing trend for shopping.

These results show that a linear-in-parameter specification is able to capture overall utility trends. An in-depth investigation of simulation results is now required to understand the impacts of the utility specifications on results (accuracy and interpretability), and on model performance.

Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
Education: $U^{\max}$	4.79	0.443	10.8	0.00
Education: α	1.57	0.202	7.75	9.1e-15
Education: β	7.56	4.84	1.56	0.119
Leisure: $U^{\max}$	4.47	0.379	4.50	9.1e-15
Leisure: α	0.668	0.213	3.13	0.00172
Leisure: β	2.53	0.686	3.69	0.000225
Shopping: $U^{\max}$	2.12	0.333	6.36	2.04e-10
Shopping: α	3.66	0.975	3.75	0.000175
Shopping: β	-4.85	2.3	-2.1	0.0353
Work: $U^{\max}$	3.31	0.637	5.19	2.08e-07
Work: α	2.07	0.0459	45.	0.00
Work: β	11.5	0.792	14.5	0.00
Summary statistics				
$L(0) = -454.1869$				
$L(\hat{\beta}) = -187.871$				
$\bar{\rho}^2 = 0.56$				

Table 3: Estimation results for PlanomatX utility function, considering $U_a^{\min} = 0$, and $\gamma_a = 1$.

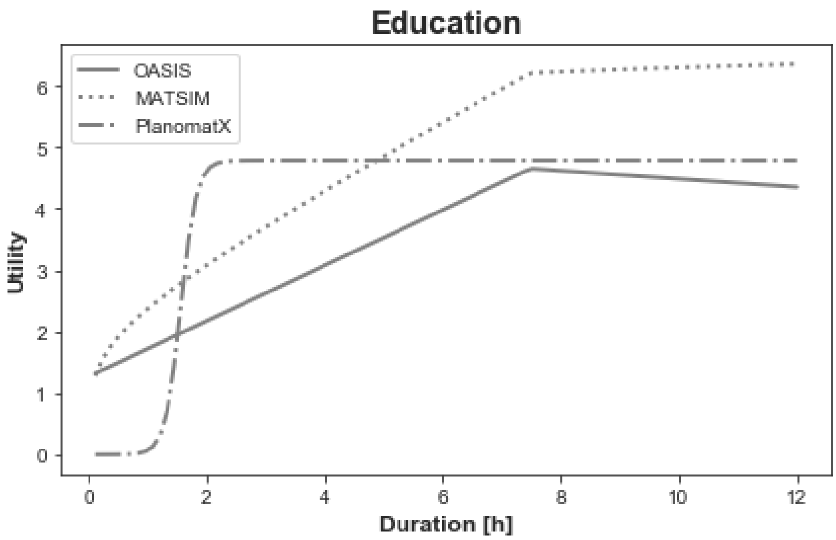

Figure 2: Utility of activity duration for education

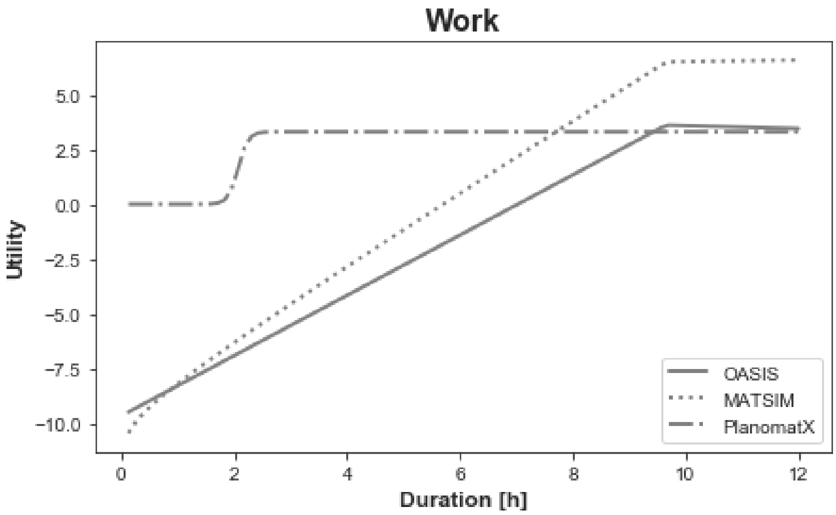

Figure 3: Utility of activity duration for work

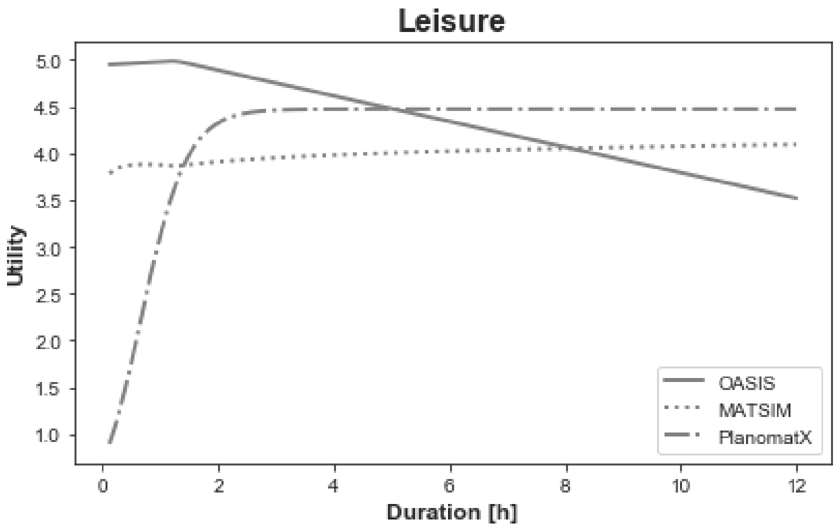

Figure 4: Utility of activity duration for leisure

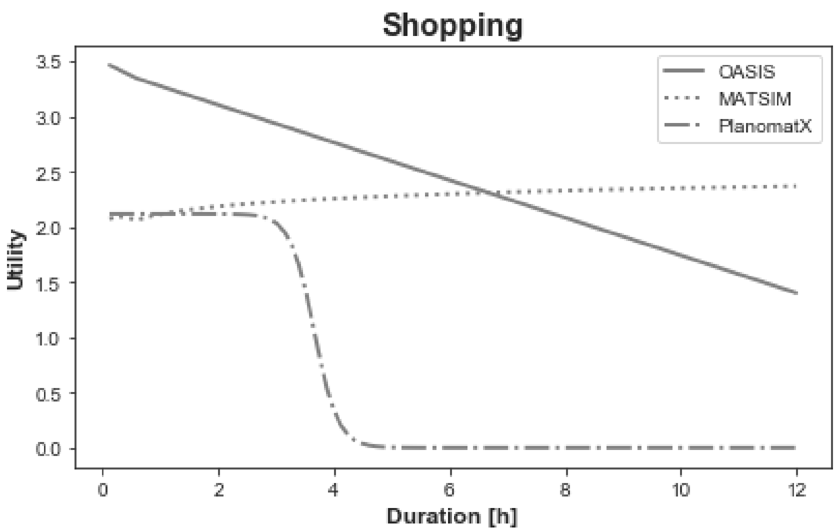

Figure 5: Utility of activity duration for shopping

4 Conclusions

In this work, we have tested three different utility specifications to explain activity-travel behaviour, with a particular emphasis on the sensitivity to activity duration. The contributions of this work are a demonstration of the estimation methodology of the OASIS framework applied to different utility functions. In addition, we have compared some behavioural insights provided by the three models, and started identifying focus areas for further investigations. Therefore, future work will include:

1. A comparative analysis of simulation outputs (schedule distributions) using the three utility functions. This analysis will be based on dedicated indicators of performance.
2. A sensitivity analysis to the model inputs, in particular to the size of the choice set, and its composition (e.g. diversity of the alternatives).
3. A relaxation of some assumptions on the values of the parameters. For instance, estimating the parameters of the PlanomatX function without the assumption of $\gamma_a = 1$.
4. An investigation of the influence of other variables and their effect on the utility of activities. For example, travel time and cost or socio-demographic characteristics.

References

Allahviranloo, M., & Axhausen, K. (2018). An optimization model to measure utility of joint and solo activities. *Transportation Research Part B: Methodological*. doi: 10.1016/j.trb.2017.12.004

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). Cambridge, MA, USA: MIT press.

Bierlaire, M. (2020). A short introduction to pandasbiogeme. *A short introduction to PandasBiogeme*.

Charypar, D., & Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation*, *32*(4), 369–397. Retrieved from `https://doi.org/10.1007/s11116-004-8287-y` doi: 10.1007/s11116-004-8287-y

Ettema, D., & Timmermans, H. (2003). Modeling Departure Time Choice in the Context of Activity Scheduling Behavior. In *Transportation research record* (pp. 39–46). National Research Council. doi: 10.3141/1831-05

Feil, M. (2010). *Choosing the Daily Schedule: Expanding Activity-Based Travel Demand Modeling* (Doctoral dissertation, ETH Zürich). doi: 10.3929/ethz-a-006200573

Joh, C. H., Arentze, T. A., & Timmermans, H. J. (2005). A utility-based analysis of activity time allocation decisions underlying segmented daily activity-travel patterns. *Environment and Planning A*. doi: 10.1068/a36214

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* (Tech. Rep.). Neuchâtel, Berne. Retrieved from `www.are.admin.ch/mtmc`

Pougala, J., Hillel, T., & Bierlaire, M. (2022, jun). Capturing trade-offs between daily scheduling choices. *Journal of Choice Modelling*, *43*, 100354. doi: 10.1016/J.JOCM.2022.100354

Small, K. A. (1982). The scheduling of consumer activities: work trips. *American Economic Review*. doi: 10.2307/1831545