Initial comparisons between Multiple Discrete-Continuous Extreme Value (MDCEV) model and Optimization-based Activity Scheduling Integrating Simultaneous choice dimensions (OASIS) framework

Mengyi Wang

Janody Pougala

Michel Bierlaire

Conference paper May 2023

STRC | 23rd Swiss Transport Research Conference
Monte Verità / Ascona, May 10-12, 2023

STRC 2023 conference paper

Initial comparisons between Multiple Discrete-Continuous Extreme Value (MDCEV) model and Optimization-based Activity Scheduling Integrating Simultaneous choice dimensions (OASIS) framework

Mengyi Wang
TRANSP-OR EPFL Lausanne
CH-1015 Lausanne, Switzerland
E-Mail: mengyi.wang@epfl.ch

Janody Pougala
TRANSP-OR EPFL Lausanne
CH-1015 Lausanne, Switzerland
E-Mail: janody.pougala@epfl.ch

Michel Bierlaire
TRANSP-OR EPFL Lausanne
CH-1015 Lausanne, Switzerland
E-Mail: michel.bierlaire@epfl.ch

Abstract

In activity-based travel demand modeling, daily scheduling process is a critical module which can help researchers understand how travel demand is generated from a micro perspective. Multiple Discrete-Continuous Extreme Value (MDCEV) model and Optimization-based Activity Scheduling Integrating Simultaneous choice dimensions (OASIS) framework are both widely used activity scheduling models in recent years. This paper takes an initial attempt to compare the two models by evaluating the forecasting performance in activity participation and activity duration choice dimensions. We estimate the models by the same samples of Lausanne from the 2015 Swiss Mobility and Transport Microcensus (MTMC) and carry out model simulation to test the ability of reproducing real data in both aggregate level and disaggregate level for the models. Because of the limitation to integrate time of day decisions of MDCEV model, more detailed comparative study of the two frameworks will be conducted when the MDCEV model is combined with scheduling model in the future.

Keywords

Activity-based modeling, activity scheduling, MDCEV, OASIS, activity participation, activity duration

Table of contents

List of tables

List of figures

1 Introduction

In recent years, activity-based travel demand models (ABM) have attracted increasing attention because they have following advantages over traditional trip-based models: 1) Focusing on individual activity behaviors and travel demand is taken as "derived"; 2) Considering clear spatial and temporal interdependencies in activity-travel choices; 3) Determining daily activity-travel schedule which can reflect people's quality of life. Therefore, ABMs can better predict future travel pattern at an individual-level and comprehensively assess the effects of transportation demand policies on people's activity and travel behaviors.

As time is treated as an important limited resource, time-use and scheduling process are central components to the activity-based approach (*1-3*). There are various existing methods in time-use field, such as fractional logit model (*4; 5*), hazard-based duration model (*6*), structural equation model (*7*), Tobit model (*8*) and multiple discrete-continuous (MDC) model (*9; 10*). Among these models, MDC model stands out because it is suitable for multiple discreteness choice which is based on utility maximization theory with diminishing marginal returns and a total time budget constraint. The original multiple discrete-continuous extreme value (MDCEV) framework (*9*) has been extended by assuming that the random terms follow different distributions (e.g. MDC nested extreme value (MDCNEV) model (*11*), MDC generalized extreme value (MDCGEV) model (*12*), MDC probit (MDCP) model (*13*) etc.) to accommodate more flexible substitutions among alternatives. In addition, some studies take attempt to adapt the framework to some different situations (e.g. multiple resource constraints (*14*), non-monotonic preferences in time-use decisions (*15*) etc.). As for model simulation, a simple and efficient algorithm which outperforms standard mathematical programming method in efficiency has also been developed (*16*). However, the above models have limitations to integrate time of day choice because they focus on the aggregate duration for an activity type rather than accommodating the time allocation at the episode level. Some work make effort to develop episode-based MDCEV models (*17; 18*), but these models still need to be combined with additional scheduling algorithm when it comes to generating daily schedules.

Optimization-based activity scheduling integrating simultaneous choice dimensions (OASIS) framework has been proposed recently (*19-24*). This modeling approach integrates different daily scheduling choice dimensions including activity participation, location, time of day, duration and travel mode into a single linear mixed integer optimization problem, which can be addressed by standard programming algorithm (*19*). In terms of estimation for model parameters, M-H algorithm is used to sample the choice set to ensure meaningful results, since the original choice set is possibly infinite (*20; 21*). Moreover, due to its high flexibility, the OASIS framework can be extended to jointly model in- and out-of-home activities and interactions among family members at household level (*22; 23*). As the OASIS model can capture trade-offs among multiple related choice dimensions in activity scheduling process, it has a wide range of application prospects in an activity-based context.

The MDCEV model and the OASIS framework have different objectives, with former focusing on the time-use mechanism and latter on the scheduling process. However, the MDCEV model can also help in decision making process of generating activity schedules when it is combined with a scheduling model. Thus, it is valuable to compare the two models to analyze their strengths, weaknesses and application scenarios. Currently, there are is no comparative study of them in the literature. This paper makes an initial attempt to evaluate the forecasting performance in activity participation and activity duration choice dimensions of the two models by using data from 2015 Swiss Mobility and Transport Microcensus (MTMC). Further efforts are required to carry out more detailed comparative experiments when the MDCEV model is combined with extra scheduling algorithm in the future.

The rest of this paper will be organized as follows: The second section reviews the MDCEV and OASIS framework and compares their features. The third section introduces data source and sample formation for the empirical analysis. The fourth section shows the estimation results for the two models by using the same samples of Lausanne from MTMC. Model simulation procedure is conducted then in the fifth section, which represents the predictive performance in activity duration and participation dimensions of the models. Limitations of the present work and directions for future research are summarized in the last section.

2 Methodology review

2.1 MDCEV model

2.1.1 Model structure

The initial econometric MDCEV framework for time use study was proposed by Bhat in 2008 (*10*), which is a classical individual utility maximization problem (the individual subscript n is omitted for clarity):

$$\max\ U(\boldsymbol{t}) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left\{ \left(\frac{t_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right\} \tag{1}$$

$$\text{s.t.} \sum_{k=1}^{K} t_k = T \tag{2}$$

Equation 1 is the objective function of this optimization problem, where $\boldsymbol{t}$ is a $(K \times 1)$ vector which represents the time allocated to K alternatives and t_k is the time allocated to activity type (or purpose) k ($0 \le t_k \le T$ for $\forall k$); $U(\boldsymbol{t})$ is the sum of utilities of all alternatives with respect to the time consumption vector $\boldsymbol{t}$; ψ_k represents the baseline marginal utility of alternative k ($\psi_k > 0$ for $\forall k$); γ_k represents a parameter of alternative k, which is related to zero consumption and satiation ($\gamma_k > 0$ for $\forall k$); α_k represents a parameter of alternative k, which is related to satiation ($0 < \alpha_k < 1$ for $\forall k$). **Equation 2** is the constraint of total time budget, where T represents the total budget ($T > 0$).

To ensure $\psi_k > 0$, ψ_k is further parameterized as:

$$\psi(\boldsymbol{z}_k, \varepsilon_k) = \exp(\boldsymbol{\beta}' \boldsymbol{z}_k + \varepsilon_k) \tag{3}$$

where $\boldsymbol{z}_k$ is a set of attributes characterizing alternative k and the decision maker, and $\boldsymbol{\beta}$ is the corresponding vector of parameters to be estimated; ε_k captures unobserved characteristics that impact the baseline marginal utility for alternative k.

The probability that the individual participates in M of the K activity types is obtained by constructing the Lagrangian function and applying KT first-order conditions for optimal time allocation, which is a simple and elegant closed form expression (See (*10*) for details):

$$P(t_1^*, t_2^*, t_3^*, \dots, t_M^*, 0,0, \dots, 0) = \left[\prod_{i=1}^{M} c_i\right]\left[\sum_{i=1}^{M} \frac{1}{c_i}\right]\left[\frac{\prod_{i=1}^{M} e_i^{V_i}}{\left(\sum_{j=1}^{K} e^{v_j}\right)^M}\right](M-1)! \tag{4}$$

It can be seen from **Equation 1** that the initial MDCEV model adopts a monotonically-increasing form of utility function with respect to time consumption. However, from our daily experience, utility may not always increase with time consumption increases, but may decrease when the individual's desire is satisfied. For example, if one keeps doing an activity for a very long time (e.g. watching two or more movies), he/she will feel uncomfortable. To capture this possible behavioral foundation, Wang and Ye (*15*) have modified the traditional framework by introducing parabolic form of utility functions. In their work, **Equation 1** is changed as below:

$$max \;\; U(\boldsymbol{t}) = \sum_{k=1}^{K} -0.5\psi_k (t_k - m_k)^2 \tag{5}$$

where ψ_k is the parameter that controls the opening size of the parabola ($\psi_k > 0$ for $\forall k$); m_k is the parameter that controls the location of the maximum point for the parabola. To satisfy KT conditions practically, randomness is introduced into the extreme point of the parabola:

$$m_{\;k} = \boldsymbol{\beta}_k{}' \boldsymbol{x}_k + \varepsilon_k \tag{6}$$

Similarly, ψ_k can be further parameterized as follows to ensure it is greater than zero:

$$\psi_k = \exp(\boldsymbol{\gamma}_k{}' \boldsymbol{z}_k) \tag{7}$$

The final probability expression can be derived in the same way as traditional model, which includes an integral and can be approximated by Gauss-Hermite technique (See (*15*) for details):

$$P(t_1^*, t_2^*, t_3^*, \dots, t_M^*, 0,0, \dots, 0) = \left(1 + \psi_1 \sum_{i=2}^{M} \psi_i^{-1}\right) \int_{-\infty}^{+\infty} g\,(\varepsilon_1) \cdot \prod_{i=2}^{M} g\,[\psi_i^{-1}(V_1 - V_i + \psi_1 \varepsilon_1)]$$

$$\cdot \prod_{s=M+1}^{K} G\,[\psi_s^{-1}(V_1 - V_s + \psi_1 \varepsilon_1)] d\varepsilon_1 \tag{8}$$

2.1.2 Forecasting algorithm

As for model simulation, instead of mathematical programming algorithm, Pinjari and Bhat proposed a simple and efficient algorithm to execute the forecasting procedure for the traditional model (*16*). This algorithm follows the model property that the marginal utility ψ_k of a chosen alternative is always greater than that of an alternative that is not chosen. The value of Lagrangian multiplier λ is somewhere between them (See (*16*) for detailed derivation).

In the algorithm, the core loop is the update of λ and the comparison of its value with the ψ_k value of the next alternative until all the chosen alternatives are determined. Similar to that, in the parabolic MDCEV model, the sorting of alternatives is based on the product value of ψ_k and m_k (*15*). This is the main difference between the two algorithms.

2.2 OASIS model

2.2.1 Model structure

The initial OASIS framework was proposed by Janody et al in 2022 (*19*). Individual considers a set of activities and each activity is associated with location, preferred starting time, preferred duration, travel mode etc. The individual is assumed to select one valid schedule with the highest utility. U_S is defined as the utility of schedule *S*, which is the sum of a generic utility U associated with the whole schedule and utility components capturing activity-travel behavior:

$$U_S = U + \sum_{a=0}^{A-1}\left(U_a^1 + U_a^2 + U_a^3 + \sum_{b=0}^{A-1}\left(U_{a,b}^4 + U_{a,b}^5\right)\right) \tag{9}$$

where:

U represents generic utility which captures aspects of the schedule that are not associated with any activity;

U_a^1 is the utility associated with participation of activity a, and it may include an error term:

$$U_a^1 = \beta_{\text{cost}} * c_a + \varepsilon_1 \quad (10)\ ;$$

U_a^2 is the utility associated with starting time of activity a, which captures the perceived penalty created by deviations from preferred starting time:

$$U_a^2 = \theta_a^e max(0, x_a^- - x_a) + \theta_a^\ell max(0, x_a - x_a^+) \tag{11}$$

In **Equation 11**, $\theta_a^e \leq 0$ and $\theta_a^\ell \leq 0$ are unknown parameters to be estimated from data;

With similar specification, U_a^3 is the utility associated with duration of activity a, which captures the perceived penalty created by deviations from preferred duration:

$$U_a^3 = \beta_{a_a}^e max(0, \tau_a^- - \tau_a) + \beta_a^\ell max(0, \tau_a - \tau_a^+) \tag{12}$$

$U_{a,b}^4$ is the utility associated with trip from location of activity a to location of activity b, irrespective of travel time (e.g. travel cost), and it may include an error term:

$$U_{a,b}^4 = \beta_{t,\mathrm{cost}} * c_t + \varepsilon_4 \tag{13}$$

$U_{a,b}^5$ is the utility which captures penalty associated with travel time from activity a to b:

$$U_{a,b}^5 = \theta_t \rho_{ab} \tag{14}$$

In **Equation 14**, θ_t is an unknown parameter to be estimated from data; ρ_{ab} is the travel time.

As all feasible daily schedules for each individual cannot be enumerated, it is necessary to rely on samples of alternatives to estimate the model. M-H algorithm is used to generate schedules with high probabilities of being chosen by the individual so that meaningful estimation can be ensured (See (*21*) for details). The likelihood function of OASIS model is the same as the multinomial logit model, in which an alternative-specific correction term is introduced in order to obtain unbiased parameters (See (*20*) for details):

$$P(i_n \mid \mathcal{C}_n) = \frac{\exp[V_{in} + \ln\ q(\mathcal{C}_n \mid i_n)]}{\sum_{j \in \mathcal{C}_n} \exp\left[V_{jn} + \ln\ q(\mathcal{C}_n \mid j)\right]} \tag{15}$$

2.2.2 Forecasting algorithm

As for model simulation, individuals schedule their day by solving a linear mixed integer optimization problem to maximize their overall utility under multiple constraints. The objective function is derived from **Equation 9** with decision variables such as whether activity a is selected, whether activity b is scheduled immediately after activity a, starting time of activity a, duration of activity a and indicator variable of the availability of private mode. The constraints aimed to ensure the generated schedule is a valid one, including total budget constraint (both time and cost), minimal duration constraint, activity sequence constraint, time consistency constraint, activity selection constraint, mode choice constraint and so on (see (*19*) for details).

2.3 Comparisons of model features

From the above review, the features of the two models can be summarized and compared as follows:

- The two models are both constrained optimization problems. Specifically, MDCEV model is a non-linear optimization problem with one linear equation constraint; OASIS is a linear mixed integer optimization problem with multiple linear constraints.
- They are both based on random utility theory. In MDCEV model, individual is assumed to maximize the overall utility of all activity alternatives with respect to time consumption; In OASIS model, individual is assumed to select the valid schedule with the highest utility.
- They have different inputs and outputs. The input of MDCEV model are some explanatory variables while that of OASIS model is a set of considered activities with locations, transport modes and scheduling preferences. The output of MDCEV model includes activity participation and activity duration dimensions while that of OASIS model is a whole schedule including activity participation, start time, duration, location and travel mode.
- They have different utility forms of activity duration. MDCEV uses more sophisticated non-linear utility functions of activity duration as it focuses on exploring time use mechanism while OASIS model currently uses a linear form of function. In addition, the parabolic MDCEV and OASIS both have non-monotonic utilities.
- They have different choice sets. The choice set of MDCEV model is a set of pre-determined and finite activity types or episodes. But the choice set of OASIS consists of all possible valid schedules, which cannot be enumerated.
- They both use maximum likelihood estimation to estimate the model parameters. However, because the choice set is possibly infinite in OASIS model, it is necessary to rely on samples of alternatives to estimate the model.
- They use forecasting algorithms based on different principles. MDCEV has its specific simulation algorithm according to model property; OASIS uses the standard mathematical algorithm in model simulation. Because of the simple and efficient algorithm and fewer modeling dimensions of the MDCEV model, the time needed for simulation is expected to be shorter than OASIS model.

3 Data source and sample formation

The Mobility and Transport Microcensus (MTMC) is a Swiss nationwide survey gathering insights on the mobility behaviors of local residents (*25*). Socio-economic characteristics of respondents and those of their family members are provided. Information on their daily travel habits and detailed trip diary (1 day) are also recorded. The 2015 edition of the MTMC contains 57090 individuals, and 43630 trip diaries. In this paper, we focus on full samples of Lausanne.

The sample formation involved following steps: Firstly, we selected samples from Lausanne; Secondly, we removed those who didn't finish their daily schedule at home; Thirdly, samples who have negative activity durations or total duration is more than 24 hours were deleted; Finally, we removed those who are at home all the day. This is because we mainly focus on outdoor activities in this paper. The final data for empirical analysis consists of 1016 individuals. Among those, 700 individuals were used to estimate parameters of MDCEV and OASIS models, and the remaining 316 individuals were used to test forecasting performance of the models.

As for activity types, after some deletion and merging work, we get 8 classifications in total: home, work, education, shopping (buying non-essential goods), errands and services (buying essential goods and groceries, or using services e.g. medical appointments, etc.), business trip, leisure and escort (accompanying someone to an activity).

4 Model estimation

4.1 MDCEV model

Traditional MDCEV model mainly have two profiles: one is γ-profile by assuming $\alpha_k = 0$ for all the alternatives; the other one is α-profile by assuming $\gamma_k = 1$ for all the alternatives. We take both profiles and the parabolic form into account in this paper.

To simplify the models, we don't consider at home activity since it is the type that almost everyone participates in and usually takes up a significantly longer time than other activities so that it will need an extra "outside good" specification in the model (*10*). Therefore, we focus on the following five activity types: working, education, shopping, leisure, others (errands and services, business trip and escort).

Since we our main objective is to compare the forecasting performance of the two models with basic specification, instead of analyzing estimation results, we didn't put explanatory variables such as socio-demographic characteristics into MDCEV model but just estimated constants and satiation parameters. In addition, by doing this can also ensure some kind of comparability because the current specification of OASIS model only considers activity-travel related attributes. However, we should recognize that the current work is just an initial effort, more detailed study which includes those critical variables is needed in the future. The estimation results of MDCEV models are presented in **Table 1** and **Table 2**.

Table 1: Estimation results of traditional MDCEV model

Parameters	γ-Profile		α-Profile	
	Value	**t-Statistic**	**Value**	**t-Statistic**
ψ_k component				
Constants				
Working	0.6141	6.043	0.5892	5.903
Education	- 0.7964	- 6.087	- 0.7504	- 5.770
Shopping	0.5426	5.500	0.4390	4.601
Leisure	0.9144	9.558	1.0627	11.186
Others	—	—	—	—
Satiation parameters				
Working	1.0000	(fixed)	1.0000	(fixed)
Education	1.0000	(fixed)	1.0000	(fixed)
Shopping	0.6557	17.771	1.0000	(fixed)
Leisure	1.8053	14.018	0.4257	1.971
Others	0.8302	12.758	1.0000	(fixed)
Summary statistics				
Number of cases	700		700	
Final Log-likelihood	- 2775.5		- 2703.6	

Table 2: Estimation results of parabolic MDCEV model

Parameters	**Value**	**t-Statistic**
m_k component		
Constants		
Working	- 1.9531	- 14.027
Education	- 1.1320	- 21.744
Shopping	- 0.7094	- 33.436
Leisure	0.0000	(fixed)
Others	- 1.1403	- 15.374
ψ_k component		
Constants		
Working	- 1.0348	- 48.763
Education	1.6480	8.197
Shopping	2.2063	12.400

Leisure	—	—
Others	2.2156	5.337
Summary statistics		
Number of cases	700	
Final Log-likelihood	- 4571.119	

4.2 OASIS model

The estimation procedure of OASIS model in this paper follows the earlier work in 2022 (*24*). Five activities are considered: home, working, education, leisure and shopping. Home activities are included here because we need to ensure a complete schedule as the output. Other activities like errands and services, escort and business trips are not included here since their frequency and duration are quite small and almost will not affect the output. We define activity-specific parameters and constants in the utility functions. The parameters were estimated with a choice set of 25 alternatives per observation.

In addition, travel parameters are considered null in **Equation 9** in this paper, and the desired start times and durations of different activities for each person were drawn from pre-determined distributions which were fitted across the Lausanne population. The estimation results are showed in **Table 3**.

Table 3: Estimation results of OASIS model

Parameters	**Value**	**t-Statistic**
Constants		
Working	12.800	8.54
Education	9.160	8.69
Shopping	10.100	13.20
Leisure	10.400	11.00
Early		
Working	- 1.380	- 4.95
Education	- 1.490	- 7.85
Shopping	- 1.060	- 12.00
Leisure	- 0.254	- 1.46
Late		
Working	- 0.532	- 3.01
Education	- 0.575	- 3.42
Shopping	- 0.685	- 4.46
Leisure	- 1.030	- 8.75
Long		
Working	- 0.746	- 3.95

Education	- 5.290	- 4.37
Shopping	- 1.150	- 4.56
Leisure	- 0.434	- 5.13
Short		
Working	- 1.240	- 4.85
Education	- 0.336	- 2.10
Shopping	- 7.910	- 2.67
Leisure	- 0.566	- 2.32
Summary statistics		
Number of observations = 700		
$L(0)$ = -1875.052		
$L(\beta)$ = -218.583		
$\bar{\boldsymbol{\rho}}^{\mathbf{2}}$ = 0.873		

5 Model simulation

5.1 Statistical analysis

By using the estimated parameters shown above and the same set of 316 samples, model simulation was conducted for the three MDCEV models and the OASIS model, respectively. As the common dimensions are activity duration and activity participation of the two different frameworks when MDCEV model is not combined with scheduling model, we only focus on their predictive performance of these two choices. Also, we only concern four common activities contained in the models: working, education, leisure and shopping.

In each model, forecasting procedure was carried out 20 times for every person. All the simulation results were recorded for statistical analysis. For simulated samples and those observed in the dataset, two descriptive statistics are compared: average time spent on each activity and proportion of schedule containing each activity. Note that these statistics are derived exclusively for schedules which contain at least one activity out-of-home. **Table 4** and **Table 5** summarize the results of all the models.

Table 4: Average time spent out-of-home, in hh:min (multiple schedules for each sample)

Activity	**Data**	**MDCEV**			**OASIS**
		γ-Profile	α-Profile	Parabolic	
Working	02:36	01:15	01:15	02:53	01:03
Education	01:07	00:21	00:22	00:20	00:31
Shopping	00:13	00:42	01:07	00:27	00:14
Leisure	01:25	02:55	02:19	01:50	01:30

From **Table 4**, it can be seen that among MDCEV models, the parabolic model generates average durations that are closer to the observed ones than the other two profiles. OASIS model has better predictive results of education, shopping and leisure activities compared to the parabolic MDCEV model except for working activity, with an underestimate of more than one hour. To get a more accurate comparison, we calculate the mean squared errors (MSE) across four activity types for the four models. The values for γ-profile model, α-profile model, parabolic model and OASIS model are 1.22, 1.00, 0.23,0.69, respectively. Therefore, we can say under the current situation the parabolic MDCEV model has an overall better forecasting performance in activity duration dimension.

Table 5: Proportion of schedules containing each activity (multiple schedules for each sample)

Activity	**Data**	**MDCEV**			**OASIS**
		γ-Profile	α-Profile	Parabolic	
Working	0.33	0.54	0.45	0.77	0.21
Education	0.18	0.19	0.16	0.33	0.14
Shopping	0.33	0.49	0.42	0.43	0.17
Leisure	0.58	0.73	0.64	0.86	0.75

As showed in **Table 5**, OASIS model has an overall underestimation of activity participation, which is similar to the finding of earlier work (*24*). On the contrary, MDCEV models overestimate activity participation in general. The MSE indicators for the four models are 0.023, 0.0066, 0.076, 0.018. Overall, the α-profile MDCEV model has a better performance in terms of predicting activity participation. Interestingly, the parabolic MDCEV model has the worst value this time, compared to the other three models. Taking the two dimensions above together, OASIS model might have better forecasting performance at an aggregate level.

The forecasting performance is also tested at a disaggregate level. We randomly select one simulated schedule for each person (excluding full day at home schedule) and compare again the two statistics in **Table 6** and **Table 7**. The results show that the values are almost constant, indicating all of the models have some kind of stability.

Table 6: Average time spent out-of-home, in hh:min (one schedule for each sample)

Activity	**Data**	**MDCEV**			**OASIS**
		γ-Profile	α-Profile	Parabolic	
Working	02:36	01:09	01:18	02:56	01:13
Education	01:07	00:22	00:19	00:23	00:30

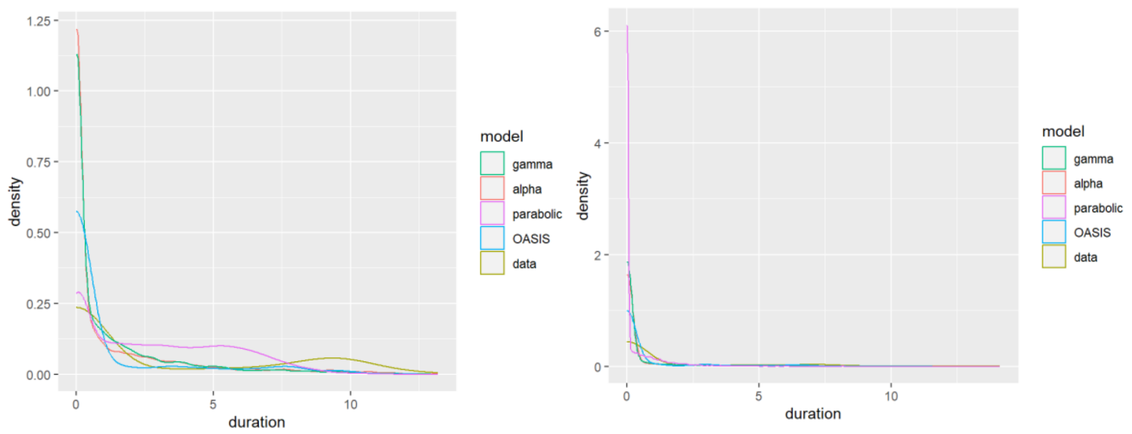
Shopping	00:13	00:39	01:09	00:27	00:18
Leisure	01:25	03:09	02:10	01:48	01:19

Table 7: Proportion of schedules containing each activity (one schedule for each sample)

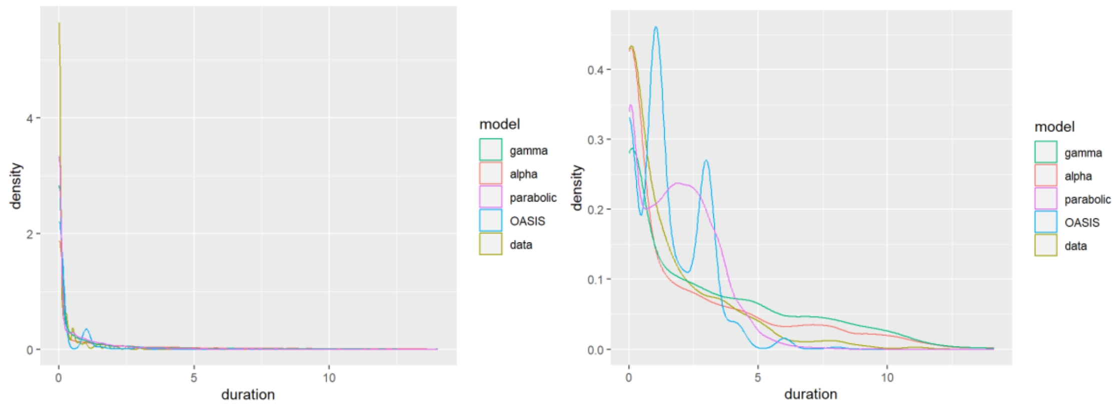
Activity	Data	MDCEV			OASIS
		γ-Profile	α-Profile	Parabolic	
Working	0.33	0.47	0.47	0.77	0.23
Education	0.18	0.19	0.16	0.35	0.14
Shopping	0.33	0.47	0.43	0.46	0.20
Leisure	0.58	0.74	0.64	0.84	0.66

5.2 Duration distribution

Finally, in order to analyze the features of distribution, density curves of the simulated durations per model and activity are compared in **Figure 1** as below. We focus mainly on working and leisure activities. As for working activity, it can be seen from figure (a) that the curve of parabolic model is the closest one to the real data at the beginning. While later, the other three models perform better until the duration is about 8 hours or so, after which the four models perform nearly the same. In terms of leisure activity, figure (d) represents that the α-profile MDCEV model is the one who has the closest curve to the observed one, compared to the other three models.



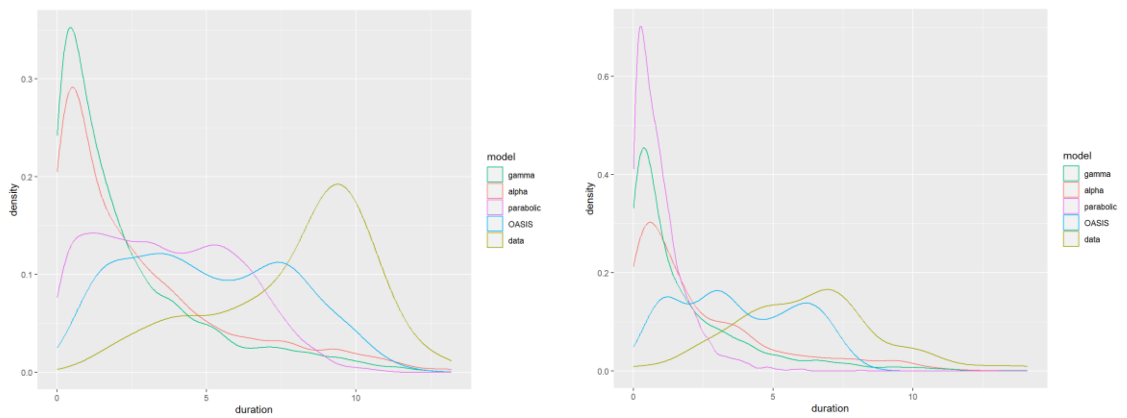
(a) Work (b) Education



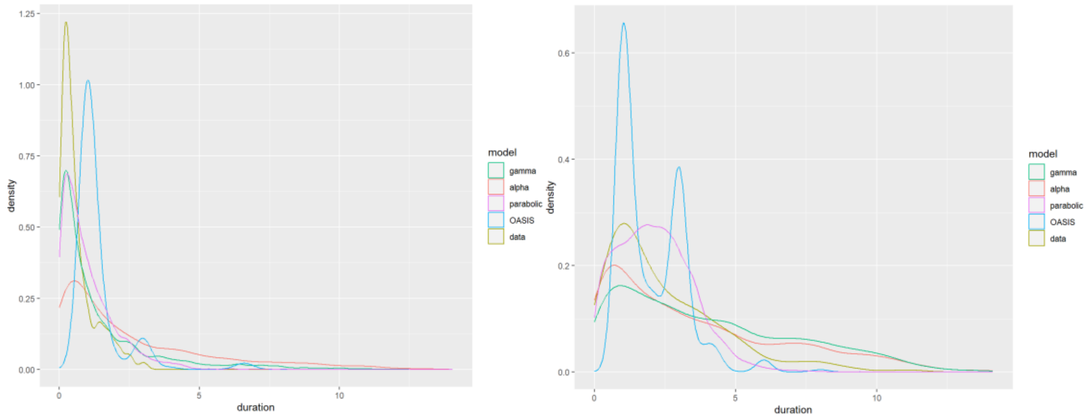
(c) Shopping (d) Leisure

Figure 1: Simulated durations (hours) among all the samples, per model and activity

The statistics and distributions of activity duration presented before are derived from the schedules that include zeros. In order to eliminate the effects from those who don't participate, density curves of duration only among those who participate are also drawn in **Figure 2**. For working activity, we can see from figure (a) that the trends of distributions of parabolic MDCEV model and OASIS model both are closer to the real one, while OASIS model performs even better due to the closer mean value. As shows by figure (b), it is obvious that the OASIS model stands out in terms of both trend and mean value in education activity. In figure (c), the mean values of shopping duration from all the models are similar, in which OASIS model has the closest trend to the observed data. Finally, when it comes to leisure activity, it can be seen form figure (d) that the parabolic MDCEV model performs the best when considering both the trend forecasting and mean value fitting.



(a) Work (b) Education



(c) Shopping (d) Leisure

Figure 2: Simulated durations (hours) among those who participate, per model and activity

6 Limitations and future directions

This paper takes an initial attempt to compare the forecasting performance in activity duration and participation choice dimensions of MDCEV model and OASIS framework. There are a few limitations of the current work which correspond to directions for future research:

Firstly, socio-demographics and more sophisticated utility functions are to be considered. Actually, MDCEV model focuses on exploring time allocation mechanism and describing influence factors by estimation results, in which individual and household variables are critical attributes. However, the present work didn't put these factors into the model, which might result in some bias in the results. Besides, the MDCEV models in this paper also didn't include at-home activity as the OASIS model, which may also have some impacts on the conclusions. Additionally, the current specification of utility in OASIS model includes only activity-travel specific variables. A linear impact on the utility has been assumed for each of them. Such a simple formulation may not be enough to capture complex behaviors and interactions.

Secondly, the consistency of model output is to be ensured. As the output of current MDCEV framework is not the same with that of OASIS model, we only chose the common activity participation and duration dimensions to compare in this paper. It is expected to conduct a full detailed comparative study when MDCEV model can also generate daily activity schedules like OASIS model. This can be either improving the MDCEV model or combing it with other models. Some studies have taken a few steps in this direction. For example, Pinjari and Bhat proposed a MDCNEV model structure which considers several time intervals of the day they modeled the activity duration of different activity types (*11*); Eluru et al. used the MDCEV model to model activity type, duration, time of day and mode choice, and then used multinomial logit (MNL) model to model destination choice (*26*); Sindi informed a conceptual framework that set the activity generation components and activity scheduling structures together (*27*) etc. Once the objective is achieved, the content of the comparative experiment can be the ability of reproducing actual data and behavioral response to external changes in different application scenarios.

Finally, more datasets are to be tested. The sample size of Lausanne is relatively small, or there are some other issues of the data that might cause identification problems in MDCEV model, as we can see from the estimation results. Anyway, other datasets are needed for extensive testing to provide more solid evidence of research conclusion in the future.

7 Reference list

1. Habib, K. N., and E. J. Miller. Modelling duration of work/school episodes using activity diary data for the specification of activity-travel scheduler.In *Proceedings of PROCESSUS Second International Colloquium on the Behavioral Foundation of Land-Use Transportation Models: Frameworks, Models and Application*, Citeseer, Toronto, Canada, 2005.

2. Habib, K. M. N., and E. J. Miller. Modeling skeletal components of workers' daily activity schedules. *Transportation Research Record*, Vol. 1985, No. 1, 2006, pp. 88-97.

3. Nurul Habib, K. M., and E. J. Miller. Modelling activity generation: a utility-based model for activity-agenda formation. *Transportmetrica,* Vol. 5, No. 1, 2009, pp. 3-23.

4. Ye, X., and R. M. Pendyala. A model of daily time use allocation using fractional logit methodology.In *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory*, University of Maryland, College Park, 2005.

5. Cardoso, A. R., E. Fontainha, and C. Monfardini. Children's and parents' time use: empirical evidence on investment in human capital in France, Germany and Italy. *Review of Economics of the Household*, Vol. 8, 2010, pp. 479-504.

6. Lee, B., and H. J. Timmermans. A latent class accelerated hazard model of activity episode durations. *Transportation Research Part B: Methodological,* Vol. 41, No. 4, 2007, pp. 426-447.

7. Kitamura, R., J. Robinson, T. Golob, M. Bradley, J. Leonard, and T. van der Hoorn. A comparative analysis of time use data in the Netherlands and California. 1992.

8. Liu, C., Y. O. Susilo, and A. Karlström. Jointly modelling individual's daily activity-travel time use and mode share by a nested multivariate Tobit model system. *Transportmetrica A Transport Science,* Vol. 13, No. 6, 2017, pp. 491-518.

9. Bhat, C. R. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological,* Vol. 39, No. 8, 2005, pp. 679-707.

10. Bhat, C. R. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological,* Vol. 42, No. 3, 2008, pp. 274-303.

11. Pinjari, A. R., and C. Bhat. A multiple discrete–continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological*, Vol. 44, No. 4, 2010, pp. 562-583.

12. Pinjari, A. R. Generalized extreme value (GEV)-based error structures for multiple discrete-continuous choice models. *Transportation Research Part B: Methodological,* Vol. 45, No. 3, 2011, pp. 474-489.

13. Astroza, S., V. M. Garikapati, D. You, A. R. Pinjari, C. R. Bhat, and R. M. Pendyala. A Multivariate Multiple Discrete Continuous Probit Model of Time Allocation to Commuting Modes and Physical Activity.In, Technical report, Department of Civil and Environmental Engineering, 2015.

14. Mondal, A., and C. R. Bhat. A new closed form multiple discrete-continuous extreme value (MDCEV) choice model with multiple linear constraints. *Transportation Research Part B: Methodological,* Vol. 147, 2021, pp. 42-66.

15. Wang, M., and X. Ye. Development of Multiple Discrete–Continuous Extreme Value (MDCEV) Model with Non-monotonic Utilities: Formulation and Application to Outdoor Non-

mandatory Timeuse Decisions. Presented at *Transportation Research Board 102th Annual Meeting*, Washington, D.C., 2023.

16. Pinjari, A. R., and C. Bhat. Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. *Journal of choice modelling,* Vol. 39, 2021, p. 100283.

17. Palma, D., A. Enam, S. Hess, C. Calastri, and R. Crastes dit Sourd. Modelling multiple occurrences of activities during a day: an extension of the MDCEV model. *Transportmetrica B: Transport Dynamics,* Vol. 9, No. 1, 2021, pp. 456-478.

18. Saxena, S., A. R. Pinjari, A. Roy, and R. Paleti. Multiple discrete-continuous choice models with bounds on consumptions. *Transportation Research Part A: Policy and Practice,* Vol. 149, 2021, pp. 237-265.

19. Pougala, J., T. Hillel, and M. Bierlaire. Capturing trade-offs between daily scheduling choices. *Journal of choice modelling*, Vol. 43, 2022, p. 100354.

20. Pougala, J., T. Hillel, and M. Bierlaire. Parameter estimation for activity-based models. Presented at *22rd Swiss Transport Research Conference (STRC 2022)*, Ascona, Switzerland, 2022.

21. Pougala, J., T. Hillel, and M. Bierlaire. Choice set generation for activity-based models. Presented at *21st Swiss Transport Research Conference (STRC 2021)*, Ascona, Switzerland 2021.

22. Rezvany, N., T. Hillel, and M. Bierlaire. Integrated in-and out-of-home scheduling framework: A utility optimization-based approach. 2022.

23. Rezvany, N., T. Hillel, and M. Bierlaire. Simulating multiple intra-household interactions. Presented at *1st EPFL Symposium on Transportation Research*, 2023.

24. Pougala, J., T. Hillel, and M. Bierlaire. OASIS: Optimisation-based Activity Scheduling with Integrated Simultaneous choice dimensions. 2022.

25. 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.

26. Eluru, N., A. Pinjari, R. Pendyala, and C. Bhat. An econometric multi-dimensional choice model of activity-travel behavior. *Transportation Letters,* Vol. 2, No. 4, 2010, pp. 217-230.

27. Sindi, A. Improved activity-based model for mega-events.In, Carleton University, 2017.