Modelling multiple occurrences of activities during a day: An extension of the MDCEV model

David Palma *[1], Annesha Enam[2], Stephane Hess[1], Chiara Calastri[1], and Romain Crastes dit Sourd[3]

[1]Institute for Transport Studies and Choice Modelling Centre, University of Leeds
[2]Argonne National Laboratory
[3]Centre for Decision Research, Leeds University Business School and Choice Modelling Centre, University of Leeds

2nd March 2021

Abstract

The increased interest in time use among transport researchers has led to a search for flexible but tractable models of time use, such as Bhat's Multiple Discrete Continuous Extreme Value (MDCEV) model. MDCEV formulations typically model aggregate time allocation into different activity types during a given period, such as the amount of time spent working and shopping in a day. While these applications provide valuable insights into activity participation, they ignore disaggregate activity-episodes, that is the fact that people might split their total time spent working in multiple separate blocks, with breaks or other activities in between. Insights into this splitting into episodes are necessary for predicting trips and understanding time use satiation. We propose a modified MDCEV model where an *activity-episode*, rather than an *activity type*, is the basic choice alternative, using a modified utility function to capture the reduced likelihood of individuals performing a very large number of episodes of the same activity. Results from two large revealed preference datasets exhibit equivalent forecast accuracy between the traditional and proposed approach at an aggregate level, but the latter also provides insights on the number and duration of activity-episodes with significant accuracy.

Keywords: time use, MDCEV, episodes, activity modelling, discrete-continuous

*D.Palma@leeds.ac.uk

1 Introduction

Travelling is a necessity that arises due to individuals' activity patterns (Bhat and Koppelman, 1999). People choose how to spend their time and then travel to different locations to carry out their chosen activities. This perspective has gained momentum among transport researchers, who have then been developing models to accurately understand and predict time use decisions.

Time use decisions can be thought of as choosing the activity type (purpose, e.g. work, education, shopping, etc.), number (count by purpose, i.e. number of episodes of a given activity) and duration. In the last decade, the multiple discrete continuous (MDC) structure pioneered by Hanemann (1984) has evolved into an elegant framework to model activity participation and time allocation decisions subject to a budget constraint (Bhat, 2008; Bhat et al., 2013; Liu et al., 2017; Wang and Li, 2011). However, the state-of-the-art MDC models focus on predicting the aggregate duration for an activity type rather than accommodating the time allocation at the episode level (Bhat and Misra, 1999; Calastri et al., 2017; Enam et al., 2018). Hence, the time allocation information obtained from the state-of-the-art MDC models can at best act as a constraint (Bhat et al., 2004); but will seldom be (immediately) useful for the representation of downstream travel choices such as number of trips, mode, destination and route, which rely on episode-level activity participation and time allocation decisions (Auld and Mohammadian, 2009).

Splitting the time invested in each activity into multiple episodes is relevant from a behavioural perspective, as engaging in an extended episode of an activity is different from engaging in multiple episodes of the same type for the same combined duration. For example, working for four hours, having a lunch break and then working for another four hours is not behaviourally equivalent to working for eight hours continuously. For the purpose of travel behaviour analysis, capturing the episode-level activity participation and time allocation decisions allows us to construct the trips that tie together consecutive activity episodes and subsequently model the associated travel decisions such as mode, destination or route choice (Gärling, 1998). Our example above involving two episodes of work separated by a lunch break will lead to four trips (home-work, work-restaurant, restaurant-work, work-home), but simply knowing that an individual works for eight hours a day does not provide information on the particular number of trips performed on top of the first and last (home-work and work-home). This is a crucial shortcoming of the existing modelling approaches, as trip information is necessary to generate the actual demand for a transport system, which is often the end goal of transport planning operations and management.

Our approach, even though not directly applicable to agent-based simulation models (ABM), can be a useful input if combined with a scheduling algorithm. Understanding time use behaviour is an integral part of any ABM. ABMs are an increasingly popular tool and can be broadly categorised into two groups, namely (i) tour-based models where the home-based tours (trip chains) are considered as the building blocks of the day, and (ii) activity-based models, where single activities are considered the building blocks of the day. In both the tour-based and the activity-based models, the analyst needs to estimate the amount of time a person invests in a stop (trip-end) or in an activity episode. Given the potential for the time spent in an activity to be split across multiple non-adjacent episodes (e.g. shopping in the morning and again in the afternoon), it is not sufficient to only know the aggregate amount of time a person spends in an activity type. Therefore, the approach undertaken in the current paper is a step towards evaluating the amount of time spent in an activity episode rather than in an activity type (which might involve multiple episodes per day). Both Garikapati et al. (2014) and Enam and Konduri (2017) tried to model time allocation in such a way that their prediction was suitable for ABMs. However, the frameworks they propose are more aligned with the tour-based approach of ABM and not the activity-based approach of ABM, and their methods are not applicable for finding the time allocation at the activity-episode level which is the focus of the current paper. While providing a step forward, the approach proposed in this paper only predicts the number and duration of activity-episodes, but not their sequence, and it would take an additional scheduling algorithm to make the output of the proposed approach directly applicable for ABM.

Other approaches to deal with the "episodic consumption of time" have been proposed in the literature. Pinjari and Bhat (2010) suggest splitting the day into periods (e.g. night, morning, afternoon and evening) and estimating a MDCEV model for each. While having the benefit of providing a rough schedule or at least a time window inside which each episode is performed, this method imposes fairly arbitrary definitions of the time periods.

More recently, Saxena et al. (2019) suggested an approach which has many similarities with our own. Both approaches predict the number and length of episodes from each activity during a day (or any other period of time), but do not predict their chronological order, i.e. they do not generate a schedule or a sequence. Both models are based on the MDCEV formulation, using activity-episodes as the basic alternative. The main difference is that Saxena et al. (2019) enforces the ordering of the episodes in the model formulation, therefore ensuring that a forecast will never allocate time to episode 2 of an activity unless episode 1 of the same activity has been allocated time to (i.e. episode 2 cannot be performed unless episode 1 has, see Section 2.4 for more details). While desirable, this property comes at the cost of a different, and more complex, likelihood function, requiring specialised estimation and forecasting software (e.g. a custom *R* or *Matlab* script). Our approach, on the other hand, does not enforce the order of the episodes, instead assuming independence between all episodes (implications are discussed in section 2.4). In exchange for this simplifying assumption, our method retains the simpler MDCEV formulation, which is readily available in multiple software packages (Hess and Palma, 2019; Lloyd-Smith, 2020a).

The objective of the current research is to expand on the activity participation and time allocation research based on MDC formulations with an episode-based framework. The proposed framework can produce time allocation choices at the activity-episode level, shedding light on time use behaviour at a more granular level. Its results can be readily used for trip generation models, as each change from one activity-episode to the next requires the individual to travel. This approach considers an activity-episode to be the basic choice alternative of the MDCEV model, in contrast with an activity type. Additionally, the proposed formulation accounts for the increasingly lower likelihood of performing later episodes of an activity type compared to the first. We demonstrate its potential using two large-scale household travel survey datasets, one from Leeds, UK and the other from the Puget Sound Region (PSR), USA.

The remainder of the paper is organised as follows: Section 2 presents the estimation and forecasting methodology. Section 3 describes the data sources used for our empirical examples and Section 4 presents the estimation and forecasting results. Section 5 provides a summary of the work and concludes the paper.

2 Estimation and forecasting methodology

In this section, we introduce the MDCEV model (Bhat, 2008) and describe two modelling approaches to time use data based on this model. First, we discuss the traditional or *aggregate* approach, used by most time use applications. Secondly, we describe the *episode-based* approach which we propose in the present paper.

2.1 The MDCEV model

The MDCEV model is derived from a classical individual utility maximisation problem (individual subscript n has been omitted for clarity), as follows:

$$\begin{aligned} Max_{x_n} \quad & \sum_{k=1}^{K} \frac{\gamma_k}{\alpha} \psi_k \left(\left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha} - 1 \right) \\ s.t. \quad & \sum_{k=1}^{K} x_k p_k = B \end{aligned} \tag{1}$$

where x_k is the amount of alternative k consumed (i.e. the time allocated to activity k). The utility function in (1) fulfils the requirements of additive separability and is driven by two different sets of parameters, ψ_k and γ_k. The ψ_k parameters (one for each choice alternative) represents the marginal *base utility* of alternative k, while γ_k relates to the associated level of satiation, with a larger value implying a lower satiation for alternative k, i.e. a higher consumption when chosen. The α parameter is also related to satiation, but is the same across all alternatives. We use this particular parameterisation of the MDCEV model (generic α with alternative-specific γ_k) as it leads to the most efficient forecasting algorithm (Pinjari and Bhat, 2011). Consumption is subject to a budget constraint B, expressed either in money or time unit (24 hours in the present case), with p_k representing the price per unit of alternative k.

Stochasticity is included in the model through a random error term ϵ_k in the base utility of each alternative. Both the base utility ψ_k and the satiation parameter γ_k need to be positive, and can be further parameterised as follows:

$$\psi_k = e^{\delta_k + \beta_k z_k + \epsilon_k} \tag{2}$$

$$\gamma_k = \theta_k + \lambda_k z_k \tag{3}$$

where γ_k and θ_k are constants for alternative k for the baseline utility and satiation parameters, z_k is a vector of attributes of the alternative and/or characteristics of the individual (e.g. a measure of the activity attractiveness, age of the individual, whether this observation was during a weekend, etc.), and β_k and λ_k are estimated parameters capturing the impact of z_k. Many implementations of MDCEV use an exponential transformation in the definition of γ_k to ensure positivity, but we have found that this often leads to slow model convergence and inferior solutions, while unconstrained estimation generally still yields positive estimates. If ϵ_k is assumed to follow an independent and identical $Gumbel(0, \sigma)$ distribution across individuals and alternatives, then the following closed form expression for the likelihood of a choice (i.e. time allocation throughout a day) can be derived.

$$L(x_1, x_2, ..., x_K) = \frac{1}{p_1 \sigma^{M-1}} \left(\prod_{k=1}^{M} f_k \right) \left(\sum_{k=1}^{M} \frac{p_k}{f_k} \right) \frac{\prod_{k=1}^{M} e^{\frac{V_k}{\sigma}}}{\left(\sum_{k=1}^{K} e^{\frac{V_k}{\sigma}} \right)^M} (M-1)! \tag{4}$$

where $f_k = \frac{1}{x_k + \gamma_k}$ and $V_k = \gamma_k + \beta_k z_k + (\alpha - 1)\, ln\left(\frac{x_k}{\gamma_k} + 1\right) - ln\,(p_k)$. Alternatives are ordered in such a way that the first M are consumed. This formulation corresponds to the MDCEV model without an outside good (Bhat, 2008). In the case of time use applications, the cost attribute p_k for each alternative is expressed as a single time unit. The fact that the *cost* attribute is thus the same across alternatives simplifies the equations and makes (4) independent of which alternative is labelled as the first one. In the context of this paper, a time unit corresponds to an hour.

Forecasting with the MDCEV can be done efficiently by using the algorithm proposed by Pinjari and Bhat (2011). Even though this method is proposed for MDCEV models with an outside good, it is easy to generalise it to the case without an outside good, by taking the alternative with the highest base utility (given ϵ_k) as the first consumed alternative.

2.2 Approach 1: Aggregate

The traditional approach to using MDCEV in a time use context disregards the number of episodes of each activity, focusing only on the total amount of time spent performing each activity during a given day. One observation corresponds to one day of data, the number of available alternatives in the model is equal to the number of different activities, and M is equal to the number of activities that have a non-zero consumption on a given day. The duration of each activity is the sum of the time spent in all episodes of that activity.

2.3 Approach 2: Episode-based

Like in the aggregate approach, one observation of the *episode-based* approach corresponds to a day of data, but episodes of each activity are no longer aggregated. Instead, multiple *episodes* of each activity are available. In contrast with the *aggregate* approach, M is now the number of activity-episodes that are conducted by a given individual on a given day. While in the *aggregate* approach, the number of available alternatives K is simply equal to the number of activity types, K now depends on the maximum number of episodes that the analyst defines *a priori*, which must be at least as large as the maximum number of episodes observed in the data for each activity. This leads to a substantial increase in the number of alternatives - now equal to $\sum_k E_k$, where E_k is the maximum number of instances in which activity k is performed by anyone in the data (see Figure 1 for an example on how the alternatives are coded). To avoid an excessive number of alternatives, it is theoretically possible to define a maximum number of episodes smaller than the one observed in the database. For example, even though there might be some outliers in the database who perform seven episodes of *drop-off/pick-up*, an analyst may decide to consider only up to five episodes. In such a situation, a pragmatic approach is for the analyst to aggregate the time spent in episodes five, six and seven, into a single episode. Alternatively, if there is only a small amount of observations with more than five episodes, the analyst could simply drop those observations. If too many observations are in this situation, then the analysts should consider a higher number of episodes in their modelling.

The resulting high number of alternatives poses a problem in terms of parameterisation. As time use studies generally use large datasets where participants engage with a potentially large number of episodes of each activity, it quickly becomes infeasible to use different alternative specific constants (ASC) parameters δ for each alternative (activity-episode). However, it is clearly important to allow for variability in the base utility across the episodes of the same activity type, for two key reasons. First, engaging in too many episodes of the same activity is likely to be undesirable as the amount allocated to each episode would become too small to be *productive*. Second, a large number of non-adjacent episodes would also imply more travel between activities that are geographically not in the same place, where this would then affect the amount of time left for non-travel activities. Additionally, different activities may have different average number

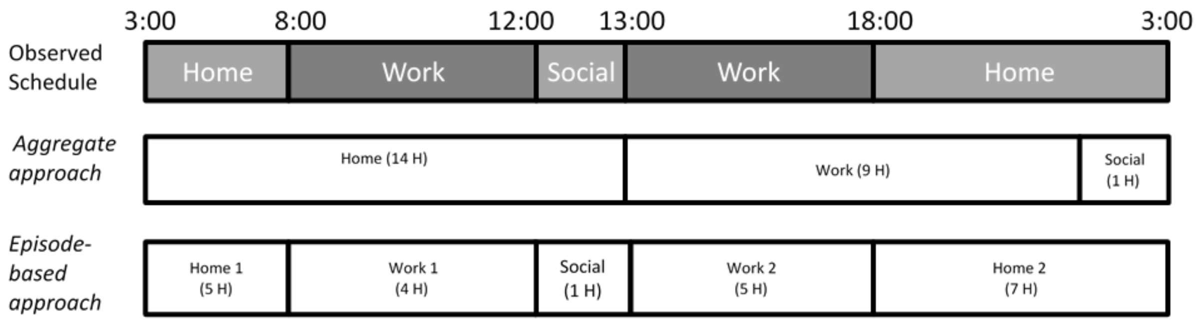

Figure 1: Example of alternative coding for traditional and episode approach

of episodes. For example, if we assume that a day starts at 3 AM as in Figure 1, being *at home* will probably have at least two episodes during the day (morning and evening), while *getting petrol* would probably be performed no more than once a day. This phenomena requires episode penalties to be different across activities.

The situation is similar when considering the duration (and satiation) of different episodes. For most activities, later episodes will likely be shorter than earlier ones. For example, a third episode of *education* in the evening will likely be shorter than the previous ones due to fatigue. But behaviour can change across activities. For example, the second *at home* episode might be longer than the first one, depending on commuting time. Using different θ parameters for each activity-episode could help capture these effects, but again, this is unpractical given the potentially high number of activity-episodes. This point about satiation also explains why, despite there being a disutility associated with engaging in an additional episode, splitting the time for major activities across events might still be beneficial. For example, engaging in two four-hour episodes of work with a rest in between is probably more desirable than engaging in one eight-hour episode. However, the disutility from satiation needs to be offset against the disutility of conducting additional episodes, and two four-hour episodes are likely to be preferred to eight one-hour episodes.

To capture these effects, we use a generic baseline constant for each activity type (δ and θ parameters) and add a polynomial episode penalty to the base utility ψ and satiation γ of each alternative, where the value of this depends on the number of the episode. To avoid identification issues, the first episode of each activity does not have any penalty; instead, the penalties apply from the second episode onwards. Penalties can be used inside both ψ_{ki} and γ_{ki}. The analyst needs to decide on what degree of polynomials to include, being mindful that a high degree will provide more flexibility to the penalty, at the cost of a higher number of parameters to estimate, and potential issues with multicollinearity. For example, a second degree polynomial will allow for a parabola-shaped penalty, which can have a single maximum or minimum. A fourth degree polynomial would allow for two local maxima or minima, and so forth. The analyst may decide what is the best degree based on the histogram of episode consumption for each activity (see, for example, Figures 3 and 4). The equations of the baseline utility and satiation parameters result as follows:

$$\psi_{ki} = exp\left(\delta_k + \beta_k z_k + \sum_{p=1}^{P_{\psi k}} \pi_{\psi kp}(i-1)^p + \varepsilon_{ki}\right) \tag{5}$$

$$\gamma_{ki} = \theta_k + \lambda_k z_k + \sum_{p=1}^{P_{\gamma k}} \pi_{\gamma kp}(i-1)^p \tag{6}$$

where i enumerates the episode of activity k, $P_{\psi k}$ and $P_{\gamma k}$ represent the number of polynomial terms used for ψ_k and γ_k, and $\pi_{\psi kp}$ $(p = 1, ..., P_{\psi k})$ and $\pi_{\gamma kp}$ $(p = 1, ..., P_{\gamma k})$ are the associated penalty parameters to be estimated. The penalty is only within activities, i.e. there is no penalty at the day level in terms of episode counts across all activities. Indeed, such a penalty term would require us to model the order between activities as well, which we are not doing.

To understand the effect of penalties more clearly, consider the case of two activities: *at home* and *get petrol*. As the first activity is usually performed twice a day, while the second one is usually performed only once, we expect penalties for *getting petrol* to be much more negative than for *at home*. Such values would make a second episode of the *getting petrol* activity much less likely than a second episode of the *at home* activity.

Revisiting the comparison between approaches, and using the notation in (5), Saxena et al. (2019) assumes that $\psi_{ki} > \psi_{k(i+1)}$, where ψ_{ki} is the marginal utility of episode i of activity k when consumption is zero. These assumptions lead to a closed-form likelihood function that is conditional on the order of episodes. This likelihood is different to the one from a traditional MDCEV, therefore requiring specialised software for model estimation.

2.4 Forecasting

In both approaches, the forecast for each observation is calculated by solving the optimisation problem in (1) multiple times, each time with different values of ε_{ki} drawn from the corresponding distribution. The final forecast is the average across solutions for all different ε_{ki}.

Pinjari and Bhat (2010) propose an efficient algorithm to solve the optimisation problem based on an iterative process. First, the price-normalised baseline marginal utilities ($\frac{\psi_k}{p_k}$) are sorted in descending order of magnitude and one alternative is incrementally added to the consumption set, until the choice set is exhausted or the magnitude of the baseline marginal utility of the next alternative in line becomes less than the Lagrangian multiplier. While the original algorithm is proposed for models with an outside good, it is easy to generalise to a model without it, by assuming that the alternative with the highest price-normalised base utility is consumed, and then proceed to calculate the optimal consumption of the remaining alternatives. More formally, for each observation in the dataset:

1. Draw a complete set of ε_k $(k = 1, ..., K)$ from the appropriate distribution.
2. Sort alternatives in decreasing order of magnitude $r_k = \frac{\psi_k}{p_k}$ Let this new ordering be indexed by m, and set M=1.
3. Compute $\lambda = \left(\frac{B + \sum_{m=1}^{M} p_m \gamma_m}{\sum_{m=1}^{M} p_m \gamma_m r_k^{\frac{1}{1-\alpha}}} \right)^{\alpha - 1}$
4. If $\lambda \leq r_{M+1}$ and M is smaller than the total number of alternatives, then make $M = M + 1$ and go back to step 3.
5. Calculate optimal consumption as $x_k = \left(\left(\frac{\lambda}{r_k} \right)^{\frac{1}{\alpha_k - 1}} - 1 \right) \gamma_k$

Both the *aggregate* and *episode-based* approaches use exactly the same forecasting algorithm. In the aggregate approach, each activity type constitutes an alternative, while in the *episode-based* approach an event of an activity type is considered to be an alternative. The only change in the described algorithm is that the index k should be replaced by the composite index ki.

When forecasting with the *episode-based* approach, nothing a priori forces an individual to choose event i before event $i + 1$ for a given activity k. For example, episode 2 will be consumed before episode 1 of activity k if $\varepsilon_{k1} < \varepsilon_{k2} + \sum_{p=1}^{P_{\psi k}} \pi_{\psi kp}$. This is a consequence of all ε_{ki} being independent from each other. However, despite the ordering of episode consumption not being guaranteed for a given set of ε_{ki} draws, this is respected when averaging across a sufficiently large number of draws. If a large number of ε_{ki} sets are used (i.e. if the forecasting algorithm is applied numerous times to each individual using different draws for each occasion) then the mean forecast across these draws will show a decreasing probability of engaging in a higher number of episodes. This is caused by the penalty terms $\pi_{\psi ki}$, which make each episode less (more) likely to be consumed if $\pi_{\psi ki}$ is negative (positive). Similarly, negative (positive) values of $\pi_{\gamma ki}$ will make later episodes more likely to be shorter (longer) in the average forecast. As usual when working with random draws, our advice is to calculate the forecast multiple times, with an increasing number of draws each time, and stop when further increases do not yield a significant change in prediction. The particular number of draws will depend on the dataset being analysed, meaning it is not possible to recommend a specific number.

A limitation when forecasting with the episode-based approach is that the maximum number of possible episodes is defined a priori by the analyst, preventing the model from predicting more than that number of episodes. Despite these limitations, the forecasting algorithm is efficient, and no adaptations are required for it to be applied to the *episode-based* approach. This allows an analyst to use standard MDCEV software to implement the *episode-based* approach, such as Apollo (Hess and Palma, 2019) or rmdcev (Lloyd-Smith, 2020b).

3 Data

In order to demonstrate the proposed approach and compare it to traditional time use modelling using the MDC framework, we use two Revealed Preferences data sources, one collected in Leeds, UK, and the other in the Puget Sound Region (PGS), USA.

3.1 Leeds dataset

The Leeds dataset was collected in 2017 as part of the ERC-funded project "DECISIONS". Time use was only one of the aspects on which the data collection effort was focused, see Calastri et al. (2020) for more details. The study participants first completed a background survey providing data on their socio-demographics, commuting behaviour, and attitudes. At a later stage, they were asked to install the mobility tracking application *rMove* (Resource Systems Group, 2017) on their smartphones. *rMove* recorded participants' location for two weeks through their phone's GPS. Every time the application detected the end of a trip, it would prompt a short survey asking the participant for the trip purpose, mode, cost (if any), and who else was part of the trip. At the end of each day of tracking, participants saw a summary of all their daily trips, giving them the opportunity to correct or complete the information if there was any error.

A total of 449 respondents successfully completed the two weeks of tracking, providing full information for at least 95% of all their trips. Most participants lived in the greater Leeds area, yet the sample is not representative of this area's population. Women (58%) and University graduates (69%) are over-represented. Most participants (30%) are between 30 and 39 years old, with under-25 participants representing only 15% of the sample. About 25% reports an income between £20K and £30K a year (see Table 1).

On the basis of the recorded trips and their stated purposes, it is possible to construct a daily activity schedule for each participant, which we use to model time use. We aggregated trip purposes into eleven activities: *home* (i.e. being at home), *work* (either at main location or elsewhere), *leisure or social* (e.g. meeting friends, going to the cinema, eating out, etc.), *drop-off/pick-up* (i.e. driving someone else to their activity location, e.g. taking children to school), *exercise* (e.g. spending time at a gym), *shopping* (both maintenance, such as grocery shopping,

Table 1: Summary of Leeds database: sample socio-demographics

		Female	Male	Total
Participants		260	189	449
Bachelor degree		182	126	308
Age	18-24	41	26	67
	25-29	29	14	43
	30-39	81	53	134
	40-49	57	36	93
	50-59	41	40	81
	60-65	7	12	19
	66-75	3	8	11
	>75	1	0	1
Personal income (thousands of £)	Missing	12	11	23
	<10	43	21	64
	10-20	65	24	89
	20-30	70	44	114
	30-40	51	44	95
	40-50	12	27	39
	50-75	4	16	20
	>75	3	2	5

and non-maintenance, such as leisure shopping), *private business* (e.g. doctor's appointment), *petrol* (i.e. buying petrol for a vehicle), and *education* (e.g. school or university classes). We considered two additional activities: *travelling*, i.e. travelling to an activity location, and an *other/unknown* activity, used in the presence of errors in the tracking (e.g. participant did not provide the purpose of a trip or the end location of a trip did not match - within a tolerance - the beginning of the next trip).

Key to our approach is the observation that people can engage in the same activity across multiple episodes throughout a day. The Leeds data contains 28,839 episodes in total for all activities. Among these, *at home* is the one participants engage with more often and for longer, and the only one in the dataset with an average of more than two daily episodes. *Travelling* follows as the second activity in terms of number of episodes, but *work* is the second highest in terms of time spent. Table 2 presents a summary of average daily time use in the Leeds dataset.

As the data was collected using geographical tracking, travelling is a pre-requisite to record a new activity. For this reason, *travelling* is a very common activity in our sample, and we did not split it into episodes as their number would have been perfectly correlated with the total number of episodes in a day (given that a new activity only starts after travelling). We did not disaggregate *other/unknown* into episodes either, as this activity mostly represents errors in data collection. We simply decided to retain it in the model to make sure that the 24-hour daily budget would be satisfied.

We limited the number of episodes per activity to five in the Leeds dataset by aggregating subsequent episodes into the fifth one. We chose not to remove observations with more than five episodes per activity, as this would have implied dropping more than 5% of the sample.

Table 2: Summary of daily activity engagement and time consumption in the Leeds sample

	Fraction of sample who Engage (%)	Time spent when engaged (Hr)		Number of episodes (#)				Length of episodes when engaged (Hr)	
				Min.	Max.	when engaged			
		Average	s.d.			Average	s.d.	Average	s.d.
At home	98	15.38	5.57	2	13	2.21	0.91	6.96	5.22
Work	46	6.66	3.30	0	18	1.78	1.12	3.74	3.38
Exercise	17	3.84	4.76	0	7	1.40	0.80	2.74	3.72
Education	4	3.55	2.93	0	5	1.43	0.79	2.49	2.59
Leisure	40	3.25	3.35	0	11	1.66	1.01	1.96	2.56
Other/unknown*	4	3.06	4.71	0	1				
Travelling*	91	2.34	2.49	0	1				
Drop-off/Pick-up	20	2.20	3.92	0	10	1.68	0.98	1.31	2.77
Private Business	25	1.98	3.28	0	10	1.49	0.88	1.33	2.58
Shopping	34	1.54	2.94	0	9	1.55	0.93	0.99	2.25
Getting petrol	3	0.95	2.78	0	2	1.02	0.14	0.94	2.76

* Engagement not split across episodes

Exercise and *Drop-off/pick-up* exhibit unusually high average daily time allocations in the Leeds dataset, as compared to the PSR dataset. This is probably due to limitations in the data collection, as several participants recorded leisure activities such as hiking and cycling (quite popular around Leeds) as *exercise*. On the other hand, *drop-off/pick-up* episodes often include the time of the following activity because if the time taken to drop-off or pick-up was short, the tracking app may have confused it with a short stop on a longer trip.

To compare the different approaches tested in this study, we set aside 20% of the sample and estimate the models with the remaining 80%. This led to 4,429 days of data used for estimation, and 1,101 days used for forecasting comparison (i.e. validation). We randomly split the full dataset at the individual level, meaning that all observations from a single individual belong to either the estimation or validation sets, but are never spread across both.

3.2 Puget Sound Region dataset

The Puget Sound Region (PSR) dataset was collected through a household travel survey from four counties (King, Kitsap, Pierce and Snohomish) located in the Puget Sound Region (PSR), Washington State. The survey collected information in a trip diary format, and was collected using two modes: a proprietary smartphone *app* (Resource Systems Group, 2017) that automatically recorded participants' trips; or a more traditional travel diary filled by participants. Assignment to each group depended on participants owning a smartphone capable of running the tracking *app*. Additionally, the households completed a telephone or online survey recording their socio-demographic characteristics. The survey collected travel patterns (for example trip start and end time, origin and destination purpose, transport mode) of the household members on a randomly selected weekday (Tuesday, Wednesday and Thursday) in Spring 2014. A total of 4,786 participants from 2,419 households participated in the survey. The current study uses information from 3,618 participants after filtering based on age ($>$18).

Table 3 provides a summary of the socio-economic characteristics of the 3,618 survey respondents. The sample is slightly skewed in terms of gender – with 54% being male. The sample over-represents highly educated people with 65% of the sample having a bachelor or higher degree. One third of the sample belong to the 35 to 54 age group, while 43% of the respondents are older than 55. In terms of household income, 37% of the households have an income of more than \$100,000 per year, while 46% of households' income falls between \$25,000 and \$100,000 per year.

The next task in terms of data preparation was to create an activity diary from the trip data collected in the survey. The 16 trip purposes were first re-coded into 13 broad categories - *home* (i.e. being at home), *work* (either at main location or elsewhere), *shopping* (both maintenance shopping, such as grocery shopping, and non-maintenance, such as leisure shopping), *education* (e.g. day-care, school or university classes), *medical* (e.g. doctor's appointment), *personal-business* (e.g. bank, post office), *drop-off/pick-up* (i.e. driving someone else to their activity location, e.g. taking children to school), *exercise* (e.g. gym, walk, jog, bike ride), *eat-out* (e.g. go to restaurant to eat/get take-out), *leisure* (e.g. attend social event such as visit with friends, family, co-workers, attend recreational event such as movies, sporting event), *religious* (go to religious/community/volunteer activity), *travel* (e.g. transfer to another mode of transport such as changing from ferry to bus) and *other*. Like in the case of the Leeds dataset, we only disaggregated the first 11 activity types into episodes. All the travel and *other* activities undertaken during the day were aggregated into a single episode. Table 4 presents a summary of time engagement from the PSR sample.

As in the case of the Leeds dataset, we limited the number of episodes per activity to five, this time by dropping observations with more than five episodes of a single activity, as these cases

Table 3: Summary of Puget Sound Region (PSR) database: sample socio-demographics

		Male	Female	Total
Gender		1951	1667	3618
Bachelor degree		1249	1118	2367
Age	18-24	91	78	169
	25-34	359	316	675
	35-44	314	308	622
	45-54	316	264	580
	55-64	415	343	758
	65-74	309	242	551
	75-84	121	95	216
	>85	26	21	47
Household income (thousands of $ per year)	Missing			231
	< 25			375
	25-50			593
	50-75			543
	75-100			541
	>100			1335

constituted less than 5% of the sample.

Like in the Leeds sample, almost 100% of respondents participate in *home* and *travel* on the survey day, while 53% and 4% of the sample participate in *work* and in *education* activity, meaning increases by 7 and 1 percentage points, respectively, compared to the Leeds sample. Another considerable difference noted in the PSR sample is that while in Leeds only 33% of respondents engage in *shopping* on the survey day, 40% do so in the PSR sample.

In the PSR sample, respondents spend around 16 hours at *home* per day on average. Average aggregate duration for *work* is around 8 hours, which is 1.3 hours higher than in the Leeds sample. Similarly, the average aggregate duration for *education* is around 1.3 hours higher in the PSR sample (5 hour) compared to the Leeds sample (3.7 hours). Discretionary activities are generally shorter compared to the Leeds sample. For example, the average (aggregated across all episode) *shopping* activity duration in the PSR sample is only about 40 minutes, while the average duration in the Leeds sample is almost 50 minutes higher. Similarly, the average aggregate *travel* duration in the PSR sample is around 1.7 hours, which is almost 40 minutes shorter than the

Table 4: Summary of activity engagement and time consumption in the PSR sample

	Fraction of sample who Engage (%)	Time spent when engaged (Hr)		Number of episodes (#)				Length of episodes when engaged (Hr)	
				Min.	Max.	when engaged			
		Average	s.d.			Average	s.d.	Average	s.d.
Home	99	15.79	4.65	0	5	2.48	0.82	6.36	3.92
Work	54	8.04	2.55	0	5	1.43	0.76	5.63	3.36
Other*	6	5.68	5.37	0	1				
Education	3	4.89	3.18	0	3	1.12	0.36	4.36	2.97
Leisure	17	3.36	3.58	0	5	1.25	0.60	2.69	2.54
Religion	4	2.58	2.01	0	5	1.25	0.71	2.06	1.62
Travel*	100	1.70	1.40	0	1				
Medical	10	1.47	1.72	0	3	1.08	0.29	1.36	1.54
Exercise	16	1.20	1.02	0	4	1.12	0.36	1.08	0.96
Eat out	23	0.91	0.85	0	4	1.22	0.50	0.75	0.70
Personal business	22	0.77	1.32	0	5	1.32	0.68	0.58	1.02
Shopping	39	0.71	0.67	0	5	1.49	0.82	0.48	0.46
Drop-off/pick-up	13	0.41	0.56	0	5	1.51	0.71	0.27	0.43

* Engagement not split across episodes

average in the Leeds sample.

In terms of episodes, people tend to engage in an average 2.5 episodes of the *home* activity during the day, which is consistent with the out-of-home activities splitting the *home* activity into at least two episodes. Other than *home*, an average of around 1.5 episodes per day is noted for *work*, *shopping*, *pick-up/drop-off* and *personal-business* activity – for the rest of the activities an average closer to 1 episode is more probable.

As with the Leeds data, 80% of the PSR sample (3,000 observations) is used for model estimation and 20% (750 observations) was set aside for the validation of the model estimation and forecasting routines.

Figure 2 shows the average length of each episode (when conducted) in both datasets. The average length of episodes decreases monotonically only among a few activities: *exercise*, *private business* and *work* in the Leeds dataset, and *exercise*, *religious* and *work* in the PSR dataset. The Leeds dataset exhibits peaks in the fifth episode for several activities, due to the aggregation of later episodes into the fifth one.

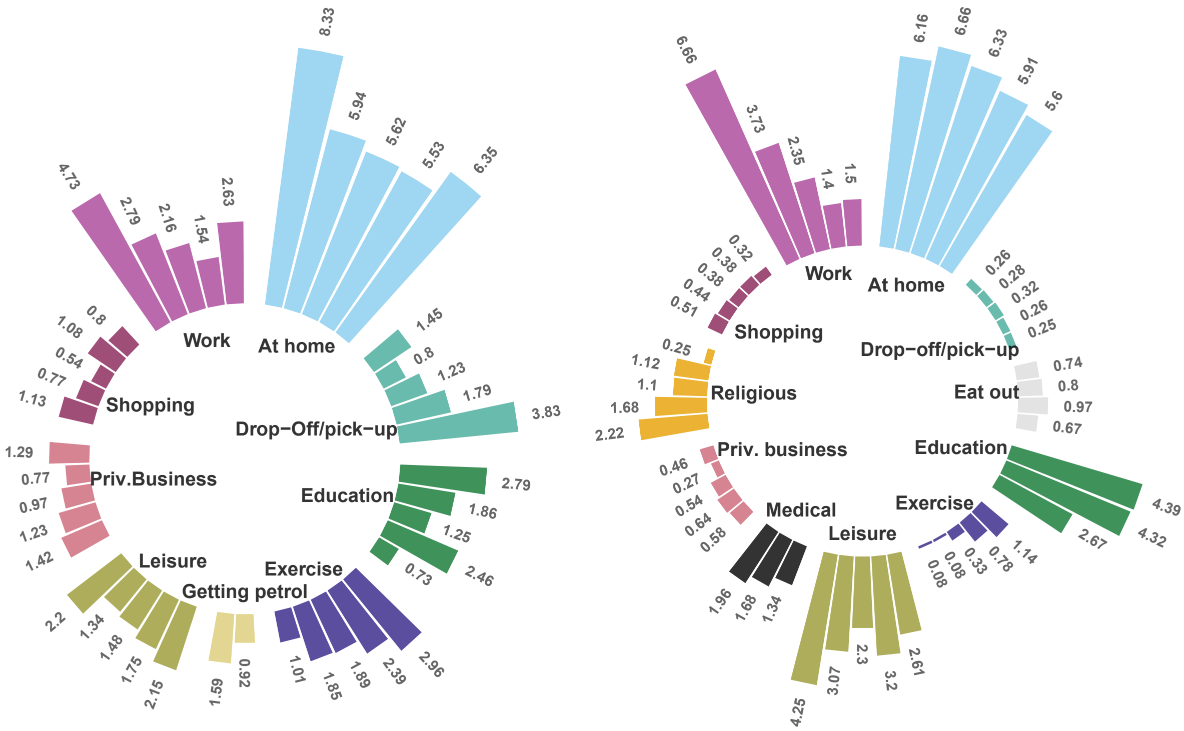

Figure 2: Average duration of episodes when engaged in the Leeds (left) and PSR (right) datasets, in hours. Episodes are ordered in a clockwise fashion.

4 Results

In this section, we present results from the proposed episode-based approach for both the Leeds and PSR datasets. We compare them against MDCEV models using the traditional aggregate approach. We begin by comparing the model parameters, followed by model fit (using the aggregate Root Mean Squared Error, RMSE), and finish with an analysis of the predicted episode numbers as compared to the observed ones.

4.1 Model parameters

The detailed parameter estimates for the Leeds models are shown in Table 5, while those for the PR models are shown in Table 6.

As Table 5 shows, coefficients signs and magnitudes are consistent across the aggregate and episode-based approach for the Leeds data. *Travelling* is the most popular activity (*ceteris* *paribus*) according to both the aggregate and episode-based model, followed by *education* and *work*. This reflects in these activities having the highest constants in their base utilities. In both models, participants are less likely to engage in *work*, *education* and *other* activities during the weekends, and instead are more likely to engage in *shopping*, *private business*, *getting petrol*, *leisure* and *exercise* activities during this period. Being *at home* is not significantly influenced by the weekend according to the aggregate approach, but it is instead positively influenced by it according to the episode-based approach. Older participants are less likely to engage in *education* activities according to both models, while there are other consistent effects of sex and income across both approaches. Concerning satiation, *work* exhibits the highest intercept in both models, meaning that -*ceteris paribus*- people spend more time working, compared to other activities. The overall effect, however, is also influenced by the base utility of the alternative, explaining why *home* is the activity consumed for the longest time (see Table 2). Satiation parameters are less influenced by the participants' characteristics, with just *drop-off/pick-up*, *exercise*, *home* and *travel* showing significant effects of covariates, all of which are consistent (or not significant) across the aggregate and episode-based approaches.

We included three penalty terms in the base utility of each alternative (i.e. $P_{\psi k} = 3$, $\forall k$), but progressively removed all non-significant parameters. We chose a third degree polynomial as a compromise between flexibility and parsimony. As Figure 3 shows, all remaining penalties have a net negative effect on the base utility of alternatives. As these only influence the base utility (ψ) from the second episode onwards (see Eq. 5), we can conclude that the objective of making later episodes less likely to take place is achieved by our functional form. As expected, *getting petrol* is the activity whose penalty becomes negative most quickly, because the vast majority of participants perform at most one episode of this activity a day. Instead, *at home* grows much slower, to make multiple episodes of the activity more likely.

We initially also included three penalty terms in the satiation effect of each alternative, but few of them reached significance, leading us to only retain linear penalties in the model. We observe that *work* and *exercise* have negative penalty parameters, meaning that later episodes of these activities tend to be shorter. *Leisure*, on the other hand, has a positive penalty, meaning that later episodes tend to be longer than previous ones, implying that later episodes are usually performed during the evening, when individuals have more time to spend in recreational activities. These results are consistent with the average duration of episodes described in Figure 2.

According to the aggregate PSR model, *home* is the most likely activity for people engage to in, followed by *shopping* and *personal business*. The episode model also identifies *home* as the most likely activity to perform, but places *work* as the second. This difference is probably due to the different covariates retained in each model, based on their significance. Men are less likely than women to participate in *work* and more likely to participate in *shopping*, *medical*, *personal business*, *pick-up/drop-off*, *leisure* and *religious* activities according to both the aggregate and episode model. As expected, people above 75 years of age are less likely to participate in

Table 5: Parameter estimates of aggregate and episode-based approach of the Leeds models (robust t-ratios)

		Aggregate approach				Episode-based approach			
		Base utility (ψ)		Satiation (γ)		Base utility (ψ)		Satiation (γ)	
		Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Drop-off/Pick-up	Intercept	-3.116	(-19.17)	0.282	(7.19)	-3.062	(-28.37)	0.136	(7.88)
	Weekend			0.242	(2.00)			0.116	(2.60)
	π_1							-0.955	(-16.40)
Work	Intercept	-1.773	(-11.93)	4.456	(20.01)	-1.766	(-22.38)	2.076	(13.04)
	Weekend	-2.175	(-13.65)			-2.057	(-11.61)		
	π_1					-1.108	(-20.39)	-0.947	(-5.51)
	π_2					0.070	(4.76)		
Education	Intercept	0.147	(0.20)	4.456	(20.01)	0.639	(0.87)	1.549	(6.86)
	Weekend	-2.254	(-5.09)			-2.536	(-5.69)		
	Age	-1.076	(-5.01)			-1.248	(-5.43)		
	Income					-1.015	(-5.04)		
	π_1					-1.274	(-7.18)		
Shopping	Intercept	-2.888	(-18.89)	0.364	(17.38)	-2.900	(-30.15)	0.230	(20.62)
	Weekend	0.672	(9.18)			0.786	(9.64)		
	Female	0.293	(3.19)			0.312	(2.93)		
	π_1					-1.311	(-24.29)		
	π_2					0.068	(3.69)		
Private business	Intercept	-3.154	(-19.96)	0.511	(12.04)	-3.102	(-30.45)	0.282	(13.02)
	Weekend	0.463	(5.82)			0.493	(6.08)		
	Female	0.246	(2.16)			0.229	(1.67)		
	π_1					-1.308	(-18.34)		
	π_2					0.074	(3.33)		
Get petrol	Intercept	-5.376	(-24.00)	0.091	(4.04)	-5.321	(-31.08)	0.093	(3.94)
	Weekend	0.527	(2.76)			0.518	(2.69)		
	π_2					-3.755	(-6.93)		
Leisure	Intercept	-2.618	(-17.04)	1.455	(20.10)	-2.611	(-30.06)	0.678	(16.29)
	Weekend	0.973	(16.14)			1.126	(17.16)		
	Female	0.143	(1.69)			0.109	(1.17)		
	π_1					-1.215	(-28.71)	0.116	(2.29)
	π_2					0.069	(5.17)		
Exercise	Intercept	-3.642	(-22.31)	1.736	(9.49)	-3.656	(-32.26)	1.332	(8.97)
	Weekend	0.712	(8.18)	0.464	(1.50)	0.889	(8.60)	-0.203	(-1.43)
	π_1					-1.495	(-14.20)	-0.696	(-5.57)
	π_2					0.123	(3.93)		
At home	Intercept	0.000	(fixed)	2.030	(6.40)	0.000	(fixed)	0.845	(15.99)
	Weekend	-0.146	(-1.04)	2.839	(3.87)	0.204	(3.94)	0.684	(8.04)
	Female	0.118	(1.68)			0.136	(1.90)		
	π_1					-1.367	(-38.91)	1.577	(15.53)
Travel	Intercept	0.666	(4.36)	0.115	(13.40)	0.869	(9.26)	0.098	(14.32)
	Full time			0.024	(3.56)			0.022	(4.02)
Other	Intercept	-4.840	(-25.98)	0.612	(2.83)	-4.792	(-34.54)	0.612	(2.83)
Loglikelihood		-41,499.86				-79,633.72			
Number of parameters		40				57			

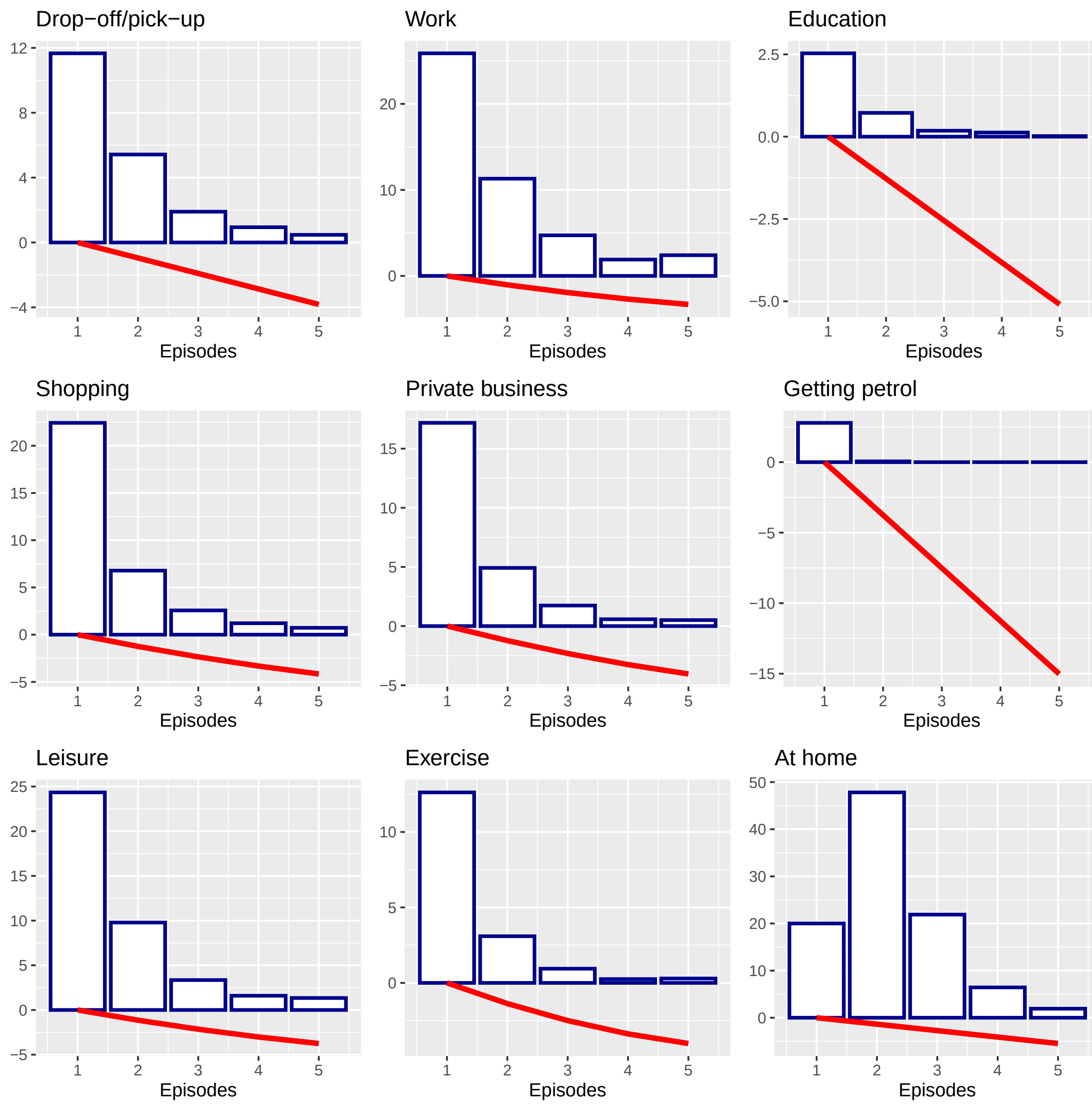

Figure 3: Frequency of episode engagement (%) and size of base utility penalty ($\sum_{p=1}^{p=3} \pi_{\psi kp}(i-1)^p$ where k, i index activity and episode, respectively) in the Leeds dataset

work according to both models. Yet only the episode model points to them being more likely to participate in *medical* and *shopping* activities. Income has no impact on the episode model, while it does significantly influence *shopping* (base utility), *work*, *exercise*, *eating out* and *leisure* (satiation) in the aggregate model. The constants of the satiation parameters exhibit the same sign for most activities in both the *aggregate* and in the *episode-based* model. *Work* has the highest positive value of the satiation constant followed by *education*, indicating people's propensity to spend longer in these activities when they perform them. *Home* and *leisure* have very similar magnitudes for their satiation parameter θ_k.

As Table 6 shows, the significance of covariates varies strongly between the *aggregate* and *episode-based* approaches. By removing non-significant parameters, we end up with very different explanatory variables in the *aggregate* and *episode-based* approaches. Just by examining the model estimates, it is not possible to establish which of the two models reflects reality in a more truthful way. We can only judge them by the accuracy of their forecasts, which we measure in Section 4.2. Both approaches have similar precision at the aggregate level, but the *episodes-based* approach provides increased detail. This leads us to believe the *episode-based* approach to be a more reliable representation of individuals' behaviour.

As Figure 4 shows, all activities have negative values for their respective penalty terms except for *home*. The positive linear penalty for *home* indicates that this activity is more likely to be split into two episodes than in one episode as is the case for the other activity types (recall that the penalty term is applied starting from the second episode of an activity type). This is in line with the observed statistics and indicates the polynomial penalty terms were able to replicate the episode participation propensity of the individuals. *Work* and *education* have very negative penalties indicating these activities are more likely to be participated in one than in more episodes. On the other hand, the much lower magnitude of the negative penalty terms in the *shopping* and *personal business* activity indicate that many people are likely to participate in multiple episodes of these activities during the day, compared to participating in multiple episodes of *work* and *education* activities.

The log-likelihood of the *aggregate* and *episode-based* models is not comparable. While the aggregate approach in the Leeds sample has a final log-likelihood of -41,500, the episode-based approach peaks at -79,634. Similarly, for the PSR sample, the likelihood peaks at -27,665 for the aggregate model and at -51,957 for the episode model. The difference is due to the episode-based approach having to explain the allocation to more alternatives, meaning that the log-likelihood becomes more negative. McFadden's $\rho^2 = 1 - \frac{LL^*}{LL_b}$ is not comparable across approaches either, as the base model LL_b used for comparison is not well defined in the case of the MDCEV model. For logit models, LL_b is usually the "null" or "equiprobable" model with all parameters set to zero, but setting all parameters to zero in the MDCEV model is not possible, as all satiation parameters must be positive. And while LL_b could be defined as a model with constants only, it is not clear how many constants should be included in the episode-based formulation. Therefore, we assess

Table 6: Parameter estimates of aggregate and episode-based approach of the PSR models (robust t-ratio)

		Aggregate approach				Episode-based approach			
		Base utility (ψ)		Satiation (γ)		Base utility (ψ)		Satiation (γ)	
		Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Home	Intercept	-5.040	(-4.92)	0.530	(3.21)	-13.090	(-116.26)	0.607	(33.7)
	Age 18-34	-0.62	(-6.75)						
	Age 35-54			-0.72	(-7.93)				
	Age 55-74	-0.500	(-5.34)						
	Age >74					0.294	(3.07)		
	Income 50-100 k$			-0.05	(-1.18)				
	Income >100 k$								
	π_1					0.118	(1.89)		
	π_2					-0.716	(-12.64)		
	π_3					0.099	(9.08)		
Work	Intercept	-10.70	(-9.97)	2.04	(55.31)	-14.630	(-118.77)	1.39	(39.98)
	Age 18-34	3.060	(8.39)						
	Age 35-54	3.050	(8.4)			0.503	(8.78)		
	Age 55-74	2.220	(6.09)						
	Age >74					-2.272	(-8.4)		
	Male	-0.27	(-5.25)			-0.233	(-4.11)		
	Income 50-100 k$	0.110	(1.44)	-0.140	(-2.24)				
	π_1					-1.586	(-27.55)		
	π_2					0.090	(3.86)		
Shopping	Intercept	-7.710	(-7.57)	-1.110	(-21.29)	-15.120	(-106.27)	-1.307	(-53.65)
	Age 18-34	-1.110	(-7.93)						
	Age 35-54	-0.970	(-7.25)	-0.09	(-1.46)	-0.336	(-4.52)		
	Age 55-74	-0.440	(-3.47)						
	Age >74					0.647	(3.71)		
	Male	0.160	(2.33)	0.15	(2.55)	0.258	(3.82)		
	Income 50-100 k$	-0.20	(-2.57)	0.09	(1.41)				
	π_1					-1.300	(-21.35)		
	π_2					0.038	(1.61)		
Education	Intercept	-11.80	(-11.41)	1.44	(7.71)	-17.600	(-91.7)	1.52	(9.96)
	Age 18-34	1.74	(6.93)	0.560	(2.79)				
	Age 35-54	0.00				-0.914	(-3.18)		
	Age 55-74	-0.86	(-2.06)						
	Income 50-100 k$			-0.340	(-1.71)				
	π_1					-2.26	(-5.18)		
Medical	Intercept	-9.11	(-8.87)	-0.05	(-0.77)	-16.91	(-97.18)	-0.03	(-0.5)
	Age 18-34	-1.77	(-7.59)						
	Age 35-54	-1.33	(-6.76)	0.20	(1.71)				
	Age 55-74	-0.83	(-4.52)						
	Age >74					0.93	(4.24)		
	Male	0.32	(2.58)			0.28	(1.95)		
	Income 50-100 k$	-0.23	(-1.66)						
	π_1					-2.66	(-12.1)		
Personal Business	Intercept	-8.48	(-8.31)	-1.23	(-23.12)	-15.97	(-117.25)	-1.47	(32.21)
	Age 18-34	-1.43	(-8.42)						
	Age 35-54	-1.05	(-6.88)			-0.23	(-2.26)		
	Age 55-74	-0.55	(-3.76)						
	Age >74					0.74	(4.29)		
	Male	0.23	(2.74)			0.23	(2.44)		
	π_1					-2.66	(-12.1)		

		Aggregate approach				Episode-based approach			
		Base utility (ψ)		Satiation (γ)		Base utility (ψ)		Satiation (γ)	
		Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Pick-up/	Intercept	-10.31	(-10.14)	-1.49	(-12.21)	-17.23	(-98.49)	-1.85	(-47.29)
Drop-off	Age 35-54	0.77	(7.19)			0.89	(7.92)		
	Male	0.42	(3.83)	0.12	(1.14)	0.51	(4.11)		
	Income 50-100 k$			-0.15	(-1.11)				
	Income >100 k$			-0.20	(-1.68)				
	π_1					-0.71	(-10.78)		
	π_2					0.11	(6.42)		
Exercise	Intercept	-9.25	(-9)	-0.58	(-4.73)	-16.16	(-124.79)	-0.31	(-5.77)
	Age 18-34	-0.39	(-1.94)						
	Age 35-54	-0.49	(-2.49)	0.16	(1.32)	-0.11	(-1.02)		
	Age 55-74	-0.25	(-1.31)	0.33	(2.9)				
	Male	0.10	(1.07)						
	Income 50-100 k$			0.26	(2.07)				
	Income >100 k$			0.41	(3.52)				
	π_1					-1.97	(-15.59)		
Eat out	Intercept	-9.11	(-9.01)	-0.83	(-7.97)	-15.62	(-127.89)	-0.85	(-19.3)
	Age 35-54			-0.23	(-2.03)	-0.22	(-2.36)		
	Income 50-100 k$			0.25	(2.16)				
	Income >100 k$			0.36	(3.36)				
	π_1					-1.79	(-23.34)		
Leisure	Intercept	-9.05	(-8.86)	0.63	(7.37)	-16.12	(-111.29)	0.66	(12.65)
	Age 18-34	-0.69	(-4.02)						
	Age 35-54	-0.91	(-5.3)			-0.40	(-3.57)		
	Age 55-74	-0.41	(-2.53)	0.20	(1.96)				
	Male	0.33	(3.5)			0.24	(2.25)		
	Income 50-100 k$			0.27	(2.4)				
	Income >100 k$			0.24	(1.98)				
	π_1					-1.59	(-17.15)		
Religious	Intercept	-10.04	(-9.64)	0.90	(11.49)	-17.97	(-86.28)	0.55	(5.88)
	Age 18-34	-1.66	(-4.88)						
	Age 35-54	-1.46	(-4.87)						
	Age 55-74	-0.76	(-2.84)						
	Age >74					1.13	(2.95)		
	Male	0.26	(1.36)			0.51	(2.24)		
	π_1					-1.41	(-6.36)		
Travel	Intercept	0.00	(fixed)	-7.56	(-7.48)	0.00	(fixed)	-14.17	(-128.4)
Other	Intercept	-10.54	(-10.32)	1.35	(6.47)	-17.18	(-120.67)	1.41	(7.21)
	Age 18-34	-0.30	(-1.54)						
	Age 35-54	-0.28	(-1.62)						
	Male	0.28	(1.8)						
	Income >100 k$			0.59	(1.74)				
Loglikelihood		-27,665.9				-51,957.0			
Number of parameters		89				62			

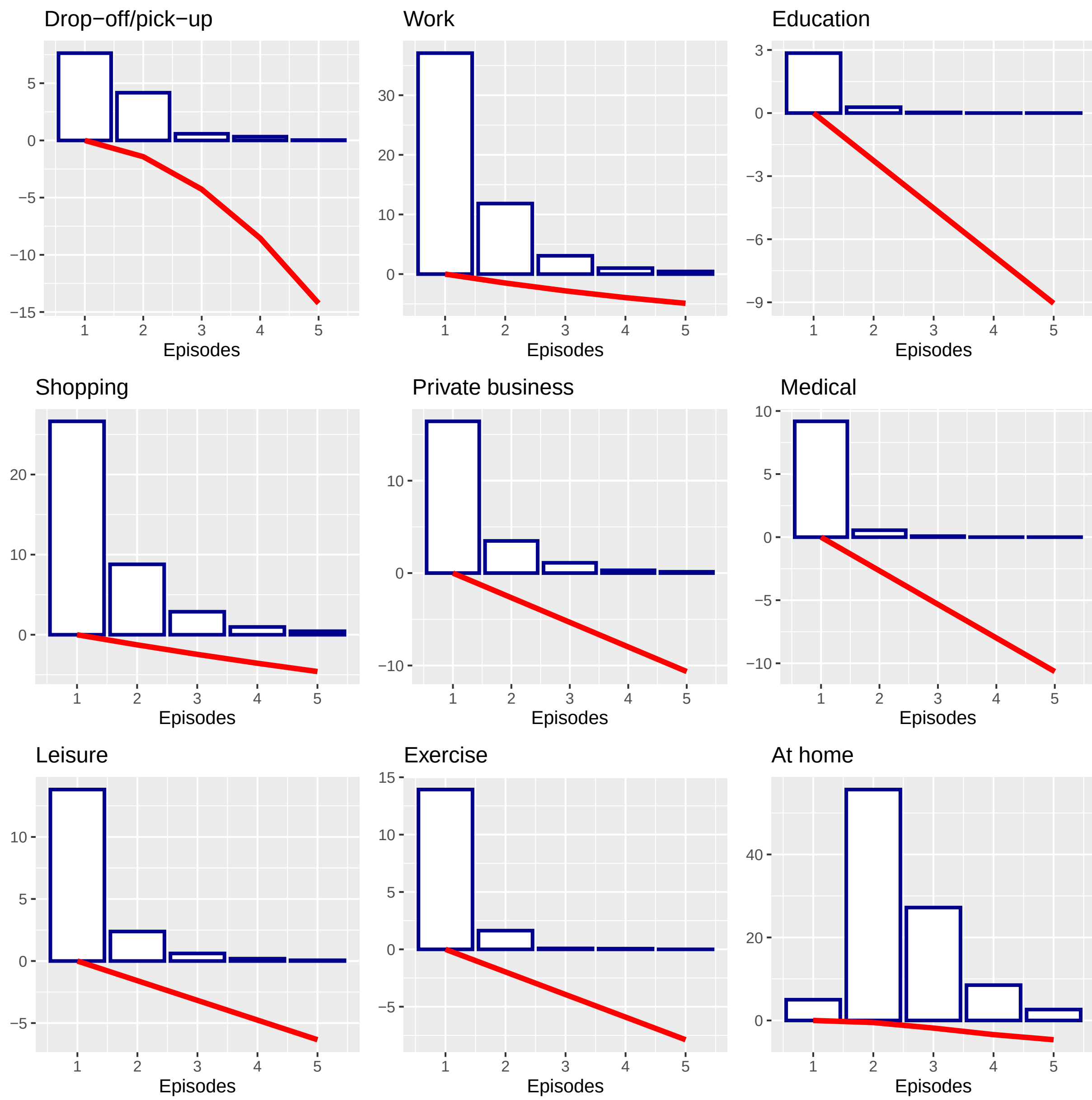

Figure 4: Frequency of episode engagement (%) and size of base utility penalty ($\sum_{p=1}^{p=3} \pi_{\psi kp}(i-1)^p$ where k, i index activity and episode, respectively) in the PSR dataset

the goodness of fit by comparing each model performance when forecasting out-of-sample, as described in the next section. This measure has the benefit of not being based on the model LL, not requiring the definition of a base model LL_b, and being a direct measure of a model's forecasting capabilities.

4.2 Forecast fit comparison

To measure the forecasting accuracy of both the aggregate and episode-based approach, we estimated the model with 80% of the whole sample, and then used that model to forecast for the remaining 20% of the data i.e. the holdout sample. All fit measurements presented in this and the following subsections are based on the holdout sample only. We measured the fit using the Root Mean Squared Error (RMSE) at the sample level, which we defined as follows:

$$RMSE = \sqrt{\frac{1}{K}\sum_k \left(\sum_n \sum_t \sum_i x_{ntki} - \sum_n \sum_t \sum_i \hat{x}_{ntki}\right)^2} \tag{7}$$

where $\hat{x}_{ntki}$ is the forecasted time allocation to episode i of activity k for observation t from individual n, with the observed values given by x_{ntki}, and K is the total number of activity-types.

Table 7 presents forecast and fit indices for the Leeds sample. Under the "Time (hours)" heading, we present the observed and predicted aggregated consumption of each activity. The forecasts are similar for the aggregate and episode-based approach, but with the second achieving a 15% smaller RMSE. This is an important improvement over the *aggregate* approach, achieved mostly due to better fit on the most popular activities (*work*, *home*, *leisure*). Such large improvement, however, is not observed in the PSR data set, so it may be dataset-specific. Under the "Activities (obs)" heading, we present the observed and predicted number of participants during a day (observations) that engage in each activity, i.e. the number of observations that perform at least one episode of the corresponding activity. Once again, we see that the forecast is very similar between the aggregated and episode-based approaches, with the episode-based approach having a slightly (4%) smaller RMSE. Under the "Episodes (epi.)" heading, we present the observed and predicted total number of episodes in the whole sample. As the *aggregate* approach cannot predict more than one episode, this heading does not apply to it. The *episode-based* approach achieves a RMSE of 128, with an average 17% error in its prediction.

We observe a similar pattern in the PSR sample, as presented in Table 8. The aggregate and episode-based approach achieve very similar fit in terms of aggregated time consumption (column "Time") and activity engagement (column "Activities") with the exception that the episode-based approach is performing better while predicting the number of individuals engaging in different activities (unlike the Leeds data). The *episode-based* reaches an RMSE of 152 when predicting the number of episodes of each activity in the whole sample, with an average 28% error per activity.

Table 7: Forecast fit comparison in the Leeds sample

	Time (hours)			Activities (obs)			Episodes (epi.)	
	Obs	Forecast		Obs	Forecast		Obs	Forecast
		Agg.	Epi.		Agg.	Epi.		Epi.
Drop-off/Pick-up	372	250	231	210	189	277	350	309
Work	3304	3609	3263	488	450	548	814	733
Education	194	193	199	52	39	50	77	54
Shopping	642	481	485	398	311	396	656	459
Private Business	418	434	400	264	234	309	368	344
Petrol	3	11	12	31	26	26	31	26
Leisure	1357	1636	1542	424	393	501	722	618
Exercise	1016	603	554	249	148	191	335	202
Home	16808	15742	16247	1088	1051	1073	2377	2061
Travel	2205	3395	3421	996	1023	1045	996	1045
Other	105	69	69	46	38	37	46	37
TOTAL	26424	26424	26424	4246	3901	4452	6772	5887
RMSE (sample)		**517**	**436**		**47**	**45**		**128**

Table 8: Forecast fit comparison in the PSR sample

	Time (hours)			Activities (obs)			Episodes (epi.)	
	Obs	Forecast		Obs	Forecast		Obs	Forecast
		Agg.	Epi.		Agg.	Epi.		Epi.
Home	11361	10489	10481	714	705	693	1117	1323
Work	2928	2837	2836	356	305	353	492	398
Shopping	207	359	387	296	224	268	459	310
Education	145	114	117	27	17	19	34	19
Medical	105	145	141	83	55	55	88	55
Personal business	134	144	148	154	119	138	223	148
Drop-off/pick-up	42	65	70	83	71	94	129	100
Exercise	135	209	201	122	86	91	135	93
Eat Out	151	229	233	163	125	139	193	145
Leisure	523	433	444	139	95	108	167	112
Religious	92	100	106	36	23	27	45	27
Travel	1274	2019	1987	724	725	725	724	725
Other	257	210	201	40	34	32	40	32
TOTAL/ Budget	17354	17354	17354	2937	2583	2742	4506	3487
RMSE (sample)		**353**	**350**		**37**	**21**		**152**

4.3 Episodes forecast analysis

In this subsection, we analyse the results from the episode-based approach in more detail, in particular its prediction of the number and duration of episodes. As the aggregate approach can only forecast a single episode per activity, we ignore it in this section. We begin by analysing the results from the Leeds dataset.

Table 9 and 10 present, under the "Total time (hours) per episode" column, the observed and predicted total time spent in each episode for each activity, from the first to the fifth episode. We observe that the total amount of time spent in each activity across the whole sample is decreasing with the order of the episodes, a phenomenon reproduced by our modelling.

While the RMSE of the total time expenditure is higher for the first episode in both samples, this is only a scale effect. If we look at the RMSE as a percentage of the average duration of each episode across activities, we obtain 26, 5, 49, 35, and 51% for the first, second, third, fourth, and fifth episodes in the Leeds sample, and 53, 49, 140, 160, and 210% in the PSR sample. This points to larger mean errors for sparsely consumed episodes or, in other words, the model predicts less accurately for those activity-episodes that are less common in the sample.

The effect of the penalty is perhaps clearer when the predicted number of episodes is analysed. In the "Observations per episode" columns in Table 9 and 10, we present the observed and predicted number of individuals during a day (observations) performing one, two, three, four or five episodes for each activity. In this case, we did not consider the order in which the episodes were performed in the forecast, but only the total number of episodes. This is due to our forecasting algorithm not enforcing the order in which the episodes should be engaged with, as discussed in section 2.4. To calculate these numbers, we register for each set of draws ϵ_{ki} the number of episodes an individual performs. We then calculate the frequency of engaging in one, two, three, four or five episodes across all draws, which is our estimate for the probabilities of an individual engaging in each possible number of episodes. Finally, we obtain the expected number of individuals performing each number of episodes by summing these probabilities across individuals.

The (expected) number of individuals conducting each number of episodes confirms that the penalty parameterisation works as expected. In both the Leeds and PSR samples we observe that most individuals engage in two episodes of the *home* activity. On the other hand, no individuals engage in more than one episode of *getting petrol* in the Leeds sample, just as in the observed data. Similarly, *education*, *medical* and *religious* activities are only performed once a day in the PSR sample.

Once again, we observe that the RMSE of the "Observations per episode" forecast decreases with the number of episodes, but again this is just a scale effect. If we calculate the ratio between these RMSE values and the average number of people engaging in each number of episodes, we obtain 33, 17, 50, 120, and 131% for the first, second, third, fourth, and fifth episodes in the

Leeds sample, and 46, 64, 77, 163 and 263% for the PSR sample. In other words, the earlier episodes are predicted more accurately than the later ones. This is because our data contains many observations with a low number of episodes being performed, and limited observations with many episodes.

Table 9: Detailed episode forecasting in the Leeds sample

	Time (hours) per episode										Observations per episode									
	Observed					Forecasted					Observed					Forecasted				
Episode:	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th	1	2	3	4	5	1	2	3	4	5
Drop-off/Pick-up	248	54	40	5	26	142	56	22	8	3	120	60	17	6	7	246	29	1	0	0
Work	2476	538	192	57	42	2272	578	239	114	61	297	109	45	21	16	388	136	22	1	0
Education	142	37	5	9	0	144	39	12	3	1	35	12	2	3	0	46	4	0	0	0
Shopping	456	102	43	22	19	335	99	33	13	5	253	80	33	16	16	337	55	4	0	0
Private Business	308	65	29	7	10	274	81	28	11	5	201	37	15	7	4	277	31	1	0	0
Petrol	3	0	0	0	0	12	0	0	0	0	31	0	0	0	0	26	0	0	0	0
Leisure	863	303	113	37	41	964	361	134	56	27	247	106	38	16	17	395	94	10	1	0
Exercise	794	143	67	11	0	452	66	22	9	5	191	36	17	4	1	180	11	0	0	0
Home	9256	5218	1819	392	123	9069	5247	1450	383	97	226	536	245	61	20	323	531	202	17	0
Travel	2205					3421					996					1045				
Other	105					69					46					37				
RMSE (sample)						**394**	**37**	**125**	**21**	**15**						**80**	**19**	**23**	**18**	**12**

Table 10: Detailed episode forecasting in the PSR sample

	Time (hours) per episode										Observations per episode									
	Observed					Forecasted					Observed					Forecasted				
Episode:	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th	1	2	3	4	5	1	2	3	4	5
Home	4485	4314	1912	471	179	5537	3601	1057	222	65	39	391	203	58	23	248	279	149	17	0
Work	2464	370	76	18	0	2148	497	132	42	16	257	71	20	7	1	309	42	1	0	0
Shopping	143	39	17	6	2	273	79	24	8	3	194	60	27	11	4	229	37	2	0	0
Education	124	18	3	0	0	107	9	1	0	0	21	5	1	0	0	19	0	0	0	0
Medical	99	4	1	0	0	132	9	1	0	0	79	3	1	0	0	54	0	0	0	0
Personal business	91	29	9	6	0	108	30	8	2	1	108	28	14	3	1	129	9	0	0	0
Drop-off/pick-up	27	12	2	1	0	39	23	7	1	1	48	27	5	3	0	87	6	0	0	0
Exercise	121	14	1	0	0	169	28	4	1	0	111	10	0	1	0	89	2	0	0	0
Eat out	122	24	2	2	0	194	33	5	1	0	140	18	3	2	0	133	6	0	0	0
Leisure	419	80	11	12	0	350	74	16	3	1	118	16	3	2	0	105	3	0	0	0
Religious	78	13	0	1	0	86	16	3	1	0	29	6	0	1	0	27	0	0	0	0
Travel	1274					1987					724	0	0	0	0	725				
Other	257					201					40	0	0	0	0	32				
RMSE (sample)						**399**	**219**	**258**	**75**	**35**						**68**	**37**	**19**	**13**	**7**

5 Discussion

In this paper, we propose a framework to enrich the MDCEV family of models with a new tool to model time use data. In particular, instead of modelling the total amount of time allocated to each activity across a whole day (or any other unit of time), we propose to model the duration of each instance or *episode* of the performed activities. In this framework, an *episode* is a continuous amount of time during which an individual engages in a given activity. There can be several episodes of the same activity within a single day, e.g. working in the morning, then performing another activity, then working again in the afternoon.

Describing and predicting time use at the episode level can provide valuable information on the number of trips performed during a day, as different activities need to be performed at different locations. This is a result that could not be inferred from activity-level time allocation alone. Furthermore, information about the episode duration is relevant as it informs the level of satiation for different activities. This could be important when planning the provision of services or understanding preferences for time use.

Our approach consists of creating multiple alternatives per activity, representing unique episodes. In terms of parameterisation, all alternatives belonging to the same activity share the same parameters measuring the impact of individual and activity characteristics on time use. At the same time, polynomial penalties are used to differentiate between the utilities of different episodes of the same activity type. When forecasting, the efficient algorithm proposed by Pinjari and Bhat (2011) can be applied, just as with the MDCEV model.

Our results indicate that the proposed episode-based approach to time use modelling is an improvement over current practice using the MDCEV model. While it does not improve the fit of the aggregate consumption as compared to a traditional MDCEV model, it does provide additional information in the form of the number of episodes each individual is likely to engage in. Such new insights do not impose additional burden in data collection, as most time use datasets are constructed from individuals' diaries recording their schedule. As a result, coding the information into aggregate time consumption per activity, or disaggregated time consumption across several episodes does not imply additional costs, other than additional data management. In other words, our approach provides new key information at marginally higher cost.

Nevertheless, we acknowledge two main limitations of the present framework. The first and most relevant one is that the current formulation does not enforce the orderly performance of episodes when forecasting. The *episode-based* formulation proposed in this paper only modifies the deterministic part of the base utility of each alternative, while keeping its stochastic part the same as in a traditional MDCEV model. This has the benefit of estimation and forecasting being the same as in the traditional MDCEV model, but it also assumes the error components (ε_{ki}) of each alternative to be independent, even across episodes of the same alternative. This can be problematic when simulating choices using an *episode-based* approach. When simulating, a single set of draws of the error terms is generated. It is possible that those draws lead to the base utility of later episodes being larger than that of earlier episodes of an alternative ($\psi_{ki} < \psi_{k(i+1)}$), and thus to later episodes being consumed while earlier episodes are not (e.g. episode 2 is consumed while episode 1 is not). While this is mostly a theoretical issue (and a practical one in simulation), it is not a problem for common forecasting, because the forecast is obtained as the average over multiple sets of draws. When averaging, the deterministic penalty terms dominate over the stochastic error terms, effectively enforcing the ordered consumption of episodes. In cases where the model is used for simulation, the labelled order of the consumed episodes should be ignored, instead focusing only on the number of episodes consumed.

The episode-based approach does not consider individuals' overall schedule, instead looking at episodic consumption in a simultaneous way. It is reasonable to believe there might be scheduling effects across activities (see e.g. Allahviranloo et al., 2017; Timmermans et al., 2002; Wets et al., 2000). For example, if an individual has engaged in many episodes throughout the day, he or she might be more inclined to limit the number of episodes in the evening. Or, if a drop-off episode happened early in the day, another pick-up episode is likely to happen later in the day. However, including scheduling in the formulation of the problem would inevitably lead to an integer optimisation problem, and to a substantial increase in complexity. Instead, our approach seeks to be as efficient and as simple as possible. If scheduling is needed (e.g. for applications to activity based modelling), this can easily be achieved at a later stage with an additional algorithm.

While the present work represents an effort into improving the realism of our time use models, other elements could of course be incorporated to capture the full complexity of human behaviour. Activity engagement throughout the day is also known to be affected by social interactions. For example, it is likely that many activities are planned at the household level, and not independently by each individual (Arentze and Timmermans, 2009; Timmermans et al., 2002). These kinds of interactions are not included in the proposed modelling approach, though some of their effects could be captured by introducing correlations between the base utility of alternatives across individuals of the same household.

Further refinements to the episodes-based approach are possible. Especially in the presence of longer panels (such as the two-week Leeds data), a mixed MDCEV approach, i.e. one incorporating random heterogeneity across individuals, would be able to capture correlations across days for the same individual in terms of the frequency of conducting different activities.

In summary, the proposed *episode-based* modelling approach extends and furthers the methodology in time use research. This approach is capable of offering additional information at virtually no additional cost compared to the traditional time use modelling approach. This extra information can be key in understanding people's preferences and behaviour, and furthermore, it can more accurately predict the total number of trips during a day, in addition to the overall time expenditure. The approach can be applied to datasets with any number of activities and episodes, as its parametrisation does not lead to an explosion of parameters in situations with a high number of alternatives or episodes. Furthermore, the approach can be applied using any software capable of estimating MDCEV models, as it does not require any modification to the estimation or forecasting algorithm.

6 Acknowledgements

The Leeds authors acknowledge the financial support by the European Research Council through the consolidator Grant 615596-DECISIONS.

References

Allahviranloo, M., Regue, R., Recker, W., 2017. Modeling the activity profiles of a population. Transportmetrica B: Transport Dynamics 5, 426–449.

Arentze, T.A., Timmermans, H.J., 2009. A need-based model of multi-day, multi-person activity generation. Transportation Research Part B: Methodological 43, 251–265.

Auld, J., Mohammadian, A., 2009. Framework for the development of the agent-based dynamic activity planning and travel scheduling (adapts) model. Transportation Letters 1, 245–255.

Bhat, C.R., 2008. The multiple discrete-continuous extreme value (mdcev) model: role of utility function parameters, identification considerations, and model extensions. Transportation Research Part B: Methodological 42, 274–303.

Bhat, C.R., Castro, M., Khan, M., 2013. A new estimation approach for the multiple discrete–continuous probit (mdcp) choice model. Transportation Research Part B: Methodological 55, 1–22.

Bhat, C.R., Guo, J.Y., Srinivasan, S., Sivakumar, A., 2004. Comprehensive econometric microsimulator for daily activity-travel patterns. Transportation Research Record 1894, 57–66.

Bhat, C.R., Koppelman, F.S., 1999. A retrospective and prospective survey of time-use research. Transportation 26, 119–139.

Bhat, C.R., Misra, R., 1999. Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. Transportation 26, 193–229.

Calastri, C., Hess, S., Daly, A., Carrasco, J.A., 2017. Does the social context help with understanding and predicting the choice of activity type and duration? an application of the multiple discrete-continuous nested extreme value model to activity diary data. Transportation Research Part A: Policy and Practice 104, 1–20.

Calastri, C., dit Sourd, R.C., Hess, S., 2020. We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. Transportation 47, 175–201. doi:10.1007/s11116-018-9858-7.

Enam, A., Konduri, K.C., 2017. Day pattern generation system for jointly modeling tours and stops: Bi-level multiple discrete continuous probit model. Transportation Research Record 2665, 69–79.

Enam, A., Konduri, K.C., Eluru, N., Ravulaparthy, S., 2018. Relationship between well-being and daily time use of elderly: evidence from the disabilities and use of time survey. Transportation 45, 1783–1810.

Garikapati, V.M., You, D., Pendyala, R.M., Vovsha, P.S., Livshits, V., Jeon, K., 2014. Multiple discrete-continuous model of activity participation and time allocation for home-based work tours. Transportation Research Record 2429, 90–98.

Gärling, T., 1998. Behavioural assumptions overlooked in travel choice modelling. Travel Behaviour Research: Updating the state of play , 3–18.

Hanemann, W.M., 1984. Discrete/continuous models of consumer demand. Econometrica: Journal of the Econometric Society , 541–561.

Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. Journal of Choice Modelling , 100170.

Liu, C., Susilo, Y.O., Karlström, A., 2017. Jointly modelling individual's daily activity-travel time use and mode share by a nested multivariate tobit model system. Transportmetrica A: Transport Science 13, 491–518.

Lloyd-Smith, P., 2020a. rmdcev: Kuhn-Tucker and Multiple Discrete-Continuous Extreme Value Models. URL: https://github.com/plloydsmith/rmdcev. r package version 1.2.4.

Lloyd-Smith, P., 2020b. rmdcev: Kuhn-Tucker and Multiple Discrete-Continuous Extreme Value Models. URL: https://github.com/plloydsmith/rmdcev. r package version 1.2.4.

Pinjari, A.R., Bhat, C., 2010. 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 44, 562–583.

Pinjari, A.R., Bhat, C.R., 2011. Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: Application to residential energy consumption analysis. Technical Report.

Resource Systems Group, 2017. rMove. http://rmove.rsginc.com/index.html.

Saxena, S., Pinjari, A.R., Paleti, R., 2019. Multiple discrete continuous choice models with conditional constraints on budget allocations: An application to disaggregate time-use analysis, in: International Choice Modelling Conference 2019.

Timmermans, H., Arentze, T., Joh, C.H., 2002. Analysing space-time behaviour: new approaches to old problems. Progress in human geography 26, 175–190.

Wang, D., Li, J., 2011. A two-level multiple discrete-continuous model of time allocation to virtual and physical activities. Transportmetrica 7, 395–416.

Wets, G., Vanhoof, K., Arentze, T., Timmermans, H., 2000. Identifying decision structures underlying activity patterns: an exploration of data mining algorithms. Transportation Research Record 1718, 1–9.