Representing and estimating interactions between activities in a need-based model of activity generation

Representing and Estimating Interactions between Activities in a Need-Based Model of Activity Generation

Linda Nijland*
Faculty of Architecture, Building and Planning
Urban Planning Group
Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven
The Netherlands
Tel: +31 40 247 3315
Fax: +31 40 243 8488
E-mail: e.w.l.nijland@bwk.tue.nl

Theo Arentze
Faculty of Architecture, Building and Planning
Urban Planning Group
Eindhoven University of Technology
The Netherlands
E-mail: t.a.arentze@bwk.tue.nl

Harry Timmermans
Faculty of Architecture, Building and Planning
Urban Planning Group
Eindhoven University of Technology
The Netherlands
E-mail: h.j.p.timmermans@bwk.tue.nl

* Corresponding author

Submission date: 10th November, 2010

Word count: 6487 + 4*250 = 7487

Submitted for presentation to
Transportation Research Board
90th Annual Meeting
January 2011
Washington, D.C.

ABSTRACT

Although several activity-based models made the transition to practice in recent years, modelling dynamic activity generation and especially, the mechanisms underlying activity generation are not well incorporated in the current activity-based models. For example, current models assume that activities are independent, but to the extent that different activities fulfil the same underlying needs and act as partial substitutes, their interactions/dependencies should be taken into account. This paper describes the parameter estimation of a need-based activity generation model, which includes the representation of possible interaction effects between activities. A survey was carried out to collect activity data for a typical week and a specific day among an adequate sample of individuals. The diary data contain detailed information on activity history and future planning. Estimation of the model involves a range of shopping, social, leisure and sports activities, as dependent variables, and socioeconomic, day preference, and interaction variables, as explanatory variables. The results show that several person, household, and dwelling attributes influence activity-episode timing decisions in a longitudinal time frame and, thus, the frequency and day choice of conducting the social, leisure and sports activities. Furthermore, interactions where found in the sense that several activities influence the need for other activities and some activities affect the utility of conducting another activity on the same day.

INTRODUCTION

Considerable progress has been made in development and application of activity-based models over the last decade. Examples of fully operational models are CEMDAP (*1*), Albatross (*2*), Famos (*3*), and TASHA (*4*). Currently, the models are making the transition to practice where they find application as instruments for planning support and policy evaluation. However, there is still ample room for improvement. High on the research agenda are the generation of activities based on the needs they satisfy or induce, interactions between activities, scheduling at the household level, and activity scheduling for a multi-day period.

In the existing activity-based models mechanisms underlying activity generation are still poorly understood and not-well represented as argued by Roorda et al. (*4*) and Habib and Miller (*5*). Chapin (*6*) was the first to argue that daily activities of individuals are driven by basic needs and that this concept should be the basis of activity-based approaches. About three decades later, this notion is further emphasized by Miller (*7*) and Axhausen (*8*). Maslow's hierarchy of needs was suggested by Miller (*7*) as a framework for modeling short- and long-term household-based decision making, but this was to the best of our knowledge not implemented in the models of his group. Meister et al. (*9*) to some extent also implemented needs into their operational model of activity scheduling.

The influence of history and interactions between activities on activity patterns has thus far been largely ignored in activity generation models. Basically, existing activity-based models have assumed a set of activities, which were then related to travel decisions. Generally, activities have been assumed implicitly or explicitly to be independent (an exception is e.g., Bradley et al. (*10*)). However, different activities may (partially) satisfy the same underlying needs (*11*). That is, activities may be (partial) substitutes in satisfying underlying needs. Especially social, recreational and leisure activities tend to be partly substitutable because they satisfy a common need of relaxation (*12*). For example, both cycling and walking (as leisure activities, not as a way of travelling) will satisfy similar needs (e.g., physical exercise, fresh air). Those activities, when conducted together with others, will also contain an element of meeting other people and therefore will partly satisfy some general social needs as well. Technically, this reasoning implies that interactions between activities should be taken into account in dynamic activity-generating models.

Empirical studies based on multi-week activity diary data support the hypothesis that activity generation should be considered in a dynamic framework. Bayarma et al. (*13*) made use of six-week travel diary data in order to analyze multi-day travel behavior. They observed that the daily travel patterns of individuals are heterogeneous. Schlich and Axhausen (*14*) used the same longitudinal dataset to measure the repetitiousness of travel behavior. Bhat et al. (*15*) used a multivariate hazard model to analyze the length between successive participations of shopping, social, recreation, and personal business activities. They found different weekly rhythms for participating in those activities, except that the rhythm for shopping activities is less distinctive. Furthermore, the results showed that strong day-of-the-week effects on inter-episode durations within these activities exist.

Arentze and Timmermans (*11*) developed a theoretical framework based on the assumption that activities are driven by a limited and universal set of subjective needs at person

and household level. Within this framework the needs grow autonomously over time according to a logistic curve with parameters depending on the nature of the need and characteristics of the individual and the household. The model predicts the timing and duration of activities in a longitudinal time frame taking into account time budget constraints, possible interactions between activities, and both household-level and person-level needs. The face validity of the suggested framework and modeling approach is supported by the results of numerical simulations, demonstrating the possibility of incorporating positive or negative substitution effects between activities and complex dynamic interactions between activities in general. In a follow-up study, based on this framework, the authors developed a RUM model and explored the extent to which the model can be estimated using existing one-day datasets (*16*). Until now, however, their approach lacks a full empirical validation based on data specifically collected for that purpose.

The aim of the research project underlying this current paper is to test the suggested approach empirically and to estimate parameters of the supposed relationships using data specifically collected for that purpose. The present paper describes the modeling approach and estimation results based on a survey, designed to model and predict the timing of activities with respect to underlying needs. The model focuses on social, leisure and sports activities (as those activities are most likely to be substitutable), the school/work hours on a typical week and the activity history and agenda of a specific sampled day. Factors included in the survey and the model consist of socioeconomic and demographic variables, activity history (e.g., time elapsed since last performance), and available time for discretionary activities (i.e., the amount of time spent on work or education on a day). The survey was held among a sample of approximately 300 individuals through a web-based questionnaire.

The organization of the paper is straightforward. First, we will briefly summarize the RUM specification of the need-based concepts and model. This is followed by a description of the survey and the sample. Section 6 describes the results of the parameter estimations. The paper closes with a discussion of the main findings of the study and remaining problems for future research.

NEED-BASED MODEL

Basic model

In this section we will briefly outline a model for predicting the timing of activities in a multi-day time frame that is proposed in Arentze et al. (*16*, *17*). The model is based on concepts from a more theoretical need-based model of activity generation, which we cited above, and has parameters that should be identifiable based on observed temporal patterns of activities. The model predicts a multi-day activity pattern agenda for a given person for a period of arbitrary length. Rather than solving some resource allocation optimization problem, the model assumes that individuals make activity-selection decisions on a daily basis. Although the model is able to take into account interactions between activities and between persons (in a household context), we will consider here a more limited situation where an individual is faced with a decision to conduct an activity i on a current day d given that the last time the activity was conducted was on day $s < d$ (this means that the time elapsed equals $d - s$ days). The utility of conducting an activity of type i on a given day d is defined as:

$$U_{nid}(s) = V_{1ni,d-s} + V_{2,nid} + \varepsilon_{1nis} + \varepsilon_{2nid} \qquad (1)$$

where n is an index of individual, d is the current day, s is the day activity i was conducted the last time before d, $V_{1ni,d-s}$ is the utility of satisfying the need for activity i built-up between s and d, $V_{2,nid}$ is a (positive or negative) preference for conducting activity i on day d and ε_{1nis} and ε_{2nid} are error terms related to need build-up (ε_1) and day (ε_2).

The utility components can be interpreted as follows. The first term (V_1) represents the amount of the need that has been built up across the elapsed time and that will be satisfied if the activity is implemented. The second term (V_2) represents a base utility dependent on preferences for day d. Note that events that are not driven by needs, but rather take place on a certain fixed day, can be modeled as activities with zero need growth ($V_1 = 0$) and a relatively high utility for the day ($V_2 >> 0$) when the event is to take place.

Implied by the first term is that a need for an activity grows over elapsed time since day s. There are several functional forms conceivable for a need's growth curve. The original model assumed a logistic growth function, but also suggests that under normal conditions need growth only moves around the area around the inflection point where the curve is approximately linear. To reduce the number of parameters, the RUM model, therefore, assumes a simple linear function here:

$$V_{1nit} = \beta_{ni} t \qquad (2)$$

where β_{ni} is a growth rate and t is the length of the need growth period between s and d ($t = d - s$).

The above equations define the (history-dependent) utility of an activity. A decision heuristic that takes into account limited time-budgets states that an activity i should be conducted on day d if d is the earliest moment when the utility of the activity per unit time exceeds a threshold. The utility-of-time threshold imposes a constraint on activity generation and represents an individual's scarcity of time. The smaller a time budget for activities, the larger the threshold needs to be. When the threshold is well adjusted, the rule leads to fully use of available time (i.e., the budgets are exhausted). At the same time, the rule ensures that every activity generates approximately an equal utility per unit of time when it is conducted. In that sense, the heuristic, even though it is very simple, will lead, as a tendency, to patterns where the utility of activities across a multi-day period cannot be improved by a revision of activity timing decisions when thresholds are well-adjusted to existing time budgets.

As a first step in estimating the model, the existing RUM model leaves activity duration out of consideration. This means that the threshold is defined on the level of utility of the activity rather than utility per unit time. The decision rule then becomes: conduct the activity at the earliest moment when the following condition holds:

$$U_{nid}(s) > u^{o}_{nd} \qquad (3)$$

where u_{nd}^{o} represents a threshold for implementing activities on day d, given existing time demands on that day. Note that defined in this way, the need-growth parameter β for some activity will capture the time needed to overcome the threshold taking into account a (average) duration of that activity. For example, keeping everything else equal, the need-growth speed will be smaller, i.e. it takes longer to overcome the threshold, if the activity has a longer duration.

Following a general approach in mixed-logit modeling, the choice model is derived from the assumption that ε_2, is normally distributed ($\varepsilon_2 \sim N(0, \sigma)$) and simulated, whereas the first error term, ε_1, is Gümbel distributed. Given this assumption, an ordered-logit framework of the following form can be derived from decision rule (3) (*12*):

$$P_{ni}(d \mid s) = \frac{\exp[Z_{nid}(s)]}{1+\exp[Z_{nid}(s)]} - \frac{\exp[\max_{k=s+1}^{d-1}[Z_{nik}(s)]]}{1+\exp[\max_{k=s+1}^{d-1}[Z_{nik}(s)]]} \qquad (4)$$

where

$$Z_{nid}(s) \equiv V_{1ni,d-s} + V_{2nid} + \varepsilon_{2nid} - u_{nd}^{o} \qquad (5)$$

Note that the conditional probabilities sum up to one across days after s:

$$\sum_{d>s} P_{ni}(d \mid s) = 1 \qquad (6)$$

Thus, P defines a choice probability distribution across days after s. In other words, the model predicts for a given activity and individual the probability of an interval time ($t = d - s$), thereby taking into account possible day-varying conditions related to day preferences and time budgets, in addition to need build-up rates. Note that the model determines whether or not an activity of a given type is conducted on a certain day; it leaves out of consideration whether this involves a single or multiple episodes of the activity on the same day.

The model represents dynamics of activity-generation decisions that follow from the fact that needs take time to re-build, and preferences and time budgets for conducting the activity may differ from day to day. A preference or size of the time-budget for a certain day of the week generates secondary effects on probabilities for other days. Secondary effects emerge because a need for the activity needs time to rebuild after the activity has been conducted. A static model which does not incorporate need build-up time, is not able to represent secondary effects of day-preferences and time-budgets, and, hence, would make wrong inferences about intrinsic day preferences.

Interactions between activities

In this section we propose a way to specify the above framework such that it can take into account interactions between activities. Interactions between activities can run through the needs component, V_1, and through the day component, V_2. As for needs, interactions occur if one activity increases or decreases the need of another activity. An example is a shopping activity that partially satisfies a need for a social activity and partially satisfies a need for being out in the

open air, etc. Interactions on the level of the day-utility component concern possible benefits of combining two activities on a same day. An example is that combining a social and shopping activity on a same tour saves travel time. In this paper, we propose the following way to incorporate these notions:

$$V_{1nid}(s) = \sum_{t=s+1}^{d}(\beta_{ni} + \sum_j \delta_{ij} I_{ntj}) - \sum_j \delta_{ji} \qquad (7)$$

$$V_{2nid} = \alpha_{nid} + \sum_j \varphi_{ij} I_{ndj} \qquad (8)$$

where n is an index of individual, β_i is the size of daily increase of a need for activity i, as before, and δ_{ij} is an increase in need of activity i caused by activity j, $I_{tj} = 1$ if activity j is conducted on day t and $I_{tj} = 0$, otherwise, α_{id} is a preference for conducting activity i on day d and φ_{ij} is the increase in utility of activity i when conducted on the same day as activity j. The last term on the RHS of Equation (7) represents the notion that a need increase caused by an activity i for some other activity j reduces the utility of i with the same amount. Thus, the first term on the RHS of Equation (7) represents the total need on a day d for the activity, depending on the history. Given the assumption that the existing need of an activity is fully satisfied when the activity is conducted, the total need for the activity is equal to a utility. The last term represents the total increase of needs for other activities caused by the activity. The need increase for other activities must be discounted, as it is a disutility.

The parameters β_i, α_{id}, u^0, δ_{ij}, and φ_{ij} are to be estimated on data. Parameters δ_{ij}, and φ_{ij} are new and represent the supposed two forms of interactions. We impose no restrictions on the ranges for these two parameters. As for the need-based interaction terms, a value of $\delta_{ij} > 0$ would represent a negative substitution effect (activity j increases the need for i) and $\delta_{ij} < 0$ a positive substitution effect (activity j decreases the need for i). Furthermore the substitution effects may be a-symmetric in the sense that $\delta_{ij} \neq \delta_{ji}$. Although we do expect that the two parameters have the same sign, we do not restrict the search range for the parameter in a loglikelihood estimation. Similarly, for the day-based interaction terms, φ_{ij} can take on a positive as well as a negative value and need not be symmetric for any pair ij. A negative value indicates that a negative preference exists to combine the activities on the same day and a positive value indicates that there is a positive preference for doing so. Note that the difference in effects of the δ and φ parameters is identifiable as that δ represents a longer lasting effect across days and φ represents a short-term effect within a day on utilities of activities.

We use the following decomposition of parameters:

$$\beta_{ni} = \beta_i^0 + \sum_k \beta_{ik} X_{1nik} \qquad (9)$$

$$u_{nd}^{o} = \mu_d^0 + \sum_m \mu_m X_{2dm} \qquad (10)$$

where X_1, X_2 are sets of explanatory variables of activity needs (Eq. 9) and time budgets (Eq. 10), and β^0 and μ^0 are base parameters and β and μ are effect parameters, to be estimated. On the other hand, for α, δ and φ parameters we do not estimate effect parameters for reasons of parsimony

(considering the degrees of freedom of the model). Finally, we use a mixed logit framework to estimate the scale σ_i of the day-based error term ε_{2i} ($\varepsilon_{2i} \sim N(0, \sigma_i)$) for each activity *i*.Thus, the model takes into account that variance in utility caused by unobserved daily circumstances can differ between activities.

MODEL ESTIMATION

As expressed in Equation (6), Equation (4) defines a probability distribution across days *d* after *s*. Whether or not this form can be used to determine likelihoods of observations depends on the nature of observations. In the survey conducted to estimate the model (see below) individuals recorded their activity agenda for a given day (*d*) and in addition for an exhaustive list of activities the day the activity was performed the last time (*s*). In case of such observations, we know that the activity has not been conducted in the time between *s* and *d*. According to the model, the probability that the activity has not been conducted in the period from *s*+1 and *d*–1 is defined as:

$$Q_{ni}(d \mid s) = 1 - \frac{\exp[\max_{k=s+1}^{d-1}[Z_{nik}(s)]]}{1 + \exp[\max_{k=s+1}^{d-1}[Z_{nik}(s)]]} \qquad (11)$$

Therefore, the probability of observing *i* in the agenda for day *d* knowing that the activity has not been conducted until that day is given by:

$$L_{ni}(1 \mid d, s) = P_{ni}(d \mid s) / Q_{ni}(d \mid s) \qquad (12)$$

$L_{ni}(1 \mid d, s)$ is the likelihood of observing activity *i* given observation day *d* and recalled last day *s*. This likelihood has the following property:

$$L_{ni}(1 \mid d, s) + L_{ni}(0 \mid d, s) = 1 \qquad (13)$$

where $L_{ni}(0 \mid d, s)$ is the likelihood of not observing activity *i* given observation day *d* and recalled last day *s*. The likelihood for a sample of observations can be defined as a function of the model's parameters as follows:

$$L(Y \mid \theta) = \prod_n \prod_i L(y_{ni} \mid \theta) \qquad (14)$$

where *Y* is a sample of individuals, θ is the set of parameters included in the model, y_{ni} is a binary variable of observing activity *i* in case of individual *n*.

$L(y_{ni})$ is a *simulated* likelihood to estimate for each activity the scale, σ_i, of the day-based error term, ε_{2i}, simultaneously with the other parameters, θ, as follows:

$$L(y_{ni} \mid \theta, \sigma_i) \approx (1/K) \sum_k^K L(y_{ni} \mid \theta, \sigma_i, \mathrm{E}_{2nik}) \qquad (15)$$

where $L(y_{ni} \mid \theta, \sigma_i, \mathrm{E}_{2nik})$ is a likelihood as defined by Equation (12), E_{2nik} is a vector of drawn error terms across days in the observed interval ($\mathrm{E}_{2nik} = (..., \varepsilon_{2nikd}, ...)$), *K* is a pre-defined number of draws of this vector and $L(y_{ni} \mid \theta, \sigma_i)$ is the simulated likelihood of the observation for individual *n* regarding activity *i* conditional on given settings of the parameters.

The likelihood function (or loglikelihood function) appears to be non-smooth in the area of the optimum values of β parameters in particular. Furthermore, due to the dependency relationship between activity probabilities across days, i.e. the secondary effects, convergence of search processes for optimal parameter values in standard loglikelihood methods is very slow. To circumvent these problems, we used a Bayesian method of estimating parameters. Bayesian methods are known to be more robust, as they do not use a function maximization process (*18*).

The Bayesian method we used for the present estimation task is based on the following equation:

$$K(\theta_i \mid \bar{\theta}_{i-}^{n}, \bar{\theta}_{i+}^{n-1}, Y_n) = \frac{L(\mathrm{y}_n \mid \bar{\theta}_{i-}^{n}, \theta_i, \bar{\theta}_{i+}^{n-1}) K(\theta_i \mid \bar{\theta}_{i-}^{n}, \bar{\theta}_{i+}^{n-1}, Y_{n-1})}{\sum_\theta L(\mathrm{y}_n \mid \bar{\theta}_{i-}^{n}, \theta, \bar{\theta}_{i+}^{n-1}) K(\theta \mid \bar{\theta}_{i-}^{n}, \bar{\theta}_{i+}^{n-1}, Y_{n-1})} \qquad (16)$$

where θ_i is the *i*-th parameter of the model, $K(\theta_i)$ is either a posterior (LHS) or prior (RHS) probability distribution across values of parameter θ_i, y_n is the *n*-th observation in the sample, Y_n is the set of observations up to *n* ($Y_n = y_1 \ldots y_n$), $\bar{\theta}_{i-}$ is a vector of expected values for parameters θ_1, θ_2, …, θ_{i-1}, and $\bar{\theta}_{i+}$ is a vector of expected values for parameters θ_{i+1}, θ_{i+2}, …, θ_m (*m* = number of parameters of the model). Equation (16) describes an incremental Bayesian learning process. Initially, a uniform distribution across some predefined wide-enough range is assumed for each parameter of the model, reflecting the assumption that no prior knowledge about parameter values exists. Observations are processed one at a time in sequence y_1, y_2, …. . For each observation the posterior distribution is determined one parameter at a time in sequence θ_1, θ_2, …., θ_m using Equation (16), whereby all other parameters are set to their current expected values (denoted as $\bar{\theta}$). The priors in each next case are set to the posteriors obtained from the last case. After all cases have been processed, the posterior distributions represent final estimates. Note that in this method each observation is used only once to update beliefs about the parameters.

DESIGN OF THE SURVEY

Data had to be collected in order to estimate the parameters of the above model. The questionnaire developed for this purpose was administered through the internet to reduce respondent burden and shorten the data entry time. In total, 37 social, sports, leisure and service-related activities were included in the survey. The activities chosen for this questionnaire were based on the activities used in earlier activity diary surveys (e.g., Amadeus (*19*)). The

questionnaire consisted of six different parts. For estimating the parameters we focus on four of them, namely:

- Socio-economic and demographic variables; as person, household, and dwelling attributes, questions concerning e.g. gender, age, household composition, income, dwelling type, education level, number of children, age youngest child, living area, car availability, and driver's license were added.

- The activity pattern of the day before; subjects were asked to indicate which activities they conducted the day before they filled out the questionnaire including some characteristics of those activities (e.g., duration, travel time, planning time horizon, and accompanying persons)

- History: The last time the activities were conducted; respondents had two ways to indicate this. First, they could specify the date, which could be selected with the help of a calendar. Second, they could indicate how many days, weeks or months ago they last performed the activity. A third option was n/a (not applicable) which could be marked if it was longer than 6 months ago or if they never do the activity. The history information was requested for the exhaustive list of 37 activities (not just the activities conducted on the day before).

- Time budgets: the standard week pattern in terms of school and work hours of the respondent. This data was obtained from a part of the questionnaire where respondents had to indicate, for every day of the week, which of the given activities they normally (phrased as 'almost always') conduct on that day. For each selected activity the subjects had to specify the usual duration and travel time. Eighteen activities were included in this part, like work, education, bring/collect child(ren), grocery shopping and some sports, leisure and social activities. In the current analysis we use the time spent on work or education on the days of the week as an explanatory variable.

TABLE 1 Composition of the sample

		Sample (%)	Population (%)
Gender	Female	53	50.5
	Male	47	49.5
Age	15 -< 25 yr	7	15
	25 -< 45 yr	48	37
	45 -< 65 yr	34	33
	65 -< 85 yr	10	16
Education	Below average	14	35
	Average	25	41
	Above average	61	24
Household composition	Single, no children	23	35
	Single, children	3	6
	Double, no children	38	29
	Double, children	33	29
	Multiple persons	1	1

SAMPLE

Subjects were selected from a sample of neighborhoods in the city of Eindhoven and seven surrounding towns. About 4000 invitation cards were distributed to households in the chosen neighborhoods in June and July 2009. Additionally, we approached approximately 400 individuals, who in an earlier survey (*20*) had indicated their willingness to participate again in an Internet survey, by e-mail. As an incentive, twenty vouchers of 50 Euros were allocated to respondents through a lottery. Altogether, 438 individuals started and 290 of them completed the web-based questionnaire.

Table 1 compares composition of the sample to the national population of the Netherlands with regard to some relevant socio-economic variables. The sample is reasonably representative except that above-average educated groups are overrepresented. This bias is typical for surveys in general (*21*). Households consisting of two persons (married or living together) are a little overrepresented and the elderly (65+ years) and young persons (< 25 years) are somewhat underrepresented.

TABLE 2 Activity groups and their activities included in the estimations

Activity group	Activities included
Daily shopping	Daily shopping
Non-daily/ Fun shopping	Non-daily shopping
	Fun shopping
Social visits	Visiting relatives/friends
	Receiving visitors
	Visiting (e.g., birthday) party
Leisure	Going out for dinner
	Visiting a theatre
	Attending a concert
	Visiting a café, bar or discotheque
	Going to the cinema
	Visiting a museum
	A day out (visit a city, recreation park)
Sports	Sports outdoors, club/association context
	Sports outdoors, flexible
	Sports indoors, club/association context
	Sports indoors, flexible

The activity data used for the analyses in the current paper consists of the cases where the respondent indicated the date of (or the time passed since) the last performance of the activity. The variable ‘time passed since last performance’ showed the amount of days between the last performance and the day before the respondents filled out the questionnaire. The activity could either be conducted or not be conducted on the latter day. Both of these options were included in the model estimation. Altogether about 4200 cases could be used for the analyses. By taking some of the most frequently conducted activities together; five activity groups were created, namely: daily shopping, non-daily/fun shopping, social visits, leisure, and sports. Note that the

activity groups are only used at the level of the parameter estimation, in the model the activities are used individually (i.e., the time elapsed since last performance of the activity is calculated for the activities separately, not for the activity group in general). Table 2 shows which activities were put together. In total those activities contain 2620 cases that can be used for the estimation of the parameters of the need-based model.

TABLE 3 Explanatory variables considered for the need-based model (base level in italics)

Variable	Code	Description / range
Day of the week	mon	Monday
	tue	Tuesday
	wed	Wednesday
	thu	Thursday
	fri	Friday
	sat	Saturday
	sun	Sunday
Gender	male	Male
	female	Female
Age group	age30-	< 30 years old
	age3040	30 – 39 years old
	age4050	40 – 49 years old
	age5060	50 – 59 years old
	age60+	60 years and older
Household composition	hh_s_no	Single, no children
	hh_sd_c	Single or Double, with child(ren)
	hh_rst	Double, no children, living in at (grand)parents/relatives, student accommodation, group accommodation
Household income	ibav	below average
	i1av	average
	iaav	above average
Age youngest child	aych06	0 – 5 years old
	aych6+	6 years and older
Hours spent work a day	tswork	Continuous
Education level	edul	Low
	edu1av	
	eduh	High
Living area	city	City
	village	Village, countryside
Dwelling type	*dwap*	Flat, apartment
	dwgarden	House
Car availability	carA	Yes, always
	carO	Yes, to be agreed with others
	carN	No

RESULTS

The selection and categorization of explanatory variables on individual and household levels to be included in the analysis was based on number of cases available for each (dummy) variable. This number may not be too low in order to get a reliable result. A threshold of 400 cases was used. The need-based model and the Bayesian estimation method to estimate the model (using the Bayesian estimation method described above (Eq. (16))) were both developed in C.

Equations 7 – 10 were used for the estimation of the parameters. As explanatory variables of activity needs (X_1) we included the person, household, and dwelling attributes shown in Table 3. As said, parameters indicating possible interaction effects between activities consist of δ estimates, which show whether the need for an activity is influenced by another activity, and φ parameters, which represent whether the utility of an activity is affected by undertaking another activity on the same day. In the current analysis, work hours (as a continuous variable) and car availability (dummy coded) were used as explanatory variables (X_2) that influence the threshold value, as those variables are likely to affect time budgets on a day. In the current formulation of the model, temporal constraints such as limited opening hours are not represented separately from other, individual-related constraints. All constraints are represented by a single threshold function. It is possible to extend the model and represent the latter constraints as an all-or-nothing availability variable for days. We leave this for future research. The threshold parameters are estimated over all activities as time-budget is day related rather than activity related. The β (need growth), α (day preferences) and day-error-scale parameters are estimated for each activity group separately. The number of draws was set as $K = 100$.

The results of the parameter estimation are shown in Table 4. The outcome should be interpreted in the following way. In terms of need-build up, the β_0 parameter represents the intercept when all other β variables are zero. Person, household and dwelling attributes influence the value of β. For example, if the respondent lives in a house with garden (Dwgarden) it decreases the value of β for Grocery shopping with 0.093. Keeping everything else equal, a decrease of β means an increase of the interval time, which is defined as the amount of time between conducting an activity and conducting the same activity again. Thus, we find that individuals living in a house with garden go less often to the supermarket or other store for daily shopping than subjects living in a flat/apartment after having corrected for possible differences in available time (given work hours), car availability, specific day preferences, and interactions between activities. On the other hand, keeping everything else equal (in particular thresholds), below average and above average educated persons do grocery shopping more often than respondents with a moderate education level. Furthermore, individuals that live in a city have faster build-up times for needs for daily shopping. The results of Non-daily and Fun shopping indicate that being single (hh_s_no) and/or being between 40 and 60 years old decreases the need to go shopping. Conversely, subjects whose youngest child is between 0 and 6 years old have shorter interval times for shopping, keeping everything else equal. In case of Social visits higher educated respondents have longer build-up times for needs for visiting relatives/friends. On the contrary, keeping everything else equal, elderly people (50+) show a higher need-recover rate for social visits than younger persons. The activity group Leisure shows negative effects for β values when the household income is lower than average, the subject lives in a city, and the age lies

below 30 or between 40 and 60 years. The age group 60+, higher income households, and living in a house with garden, on the other hand, increase the level of the needs which lowers the interval time. Finally, the results of Sports show that respondents between 40 and 50 years old, individuals living in a house with garden, and higher educated subjects have a longer need rebuild time for sports activities.

If we look at day preferences, we see that individuals tend to have an intrinsic preference for doing grocery shopping on Saturdays, Social visits on Saturdays and Sundays, and Sports on Tuesdays. On the other hand, individuals do not prefer Daily shopping on Sundays or stores are closed on that day (in the Eindhoven region, by the time of the data collection, supermarkets were closed on Sundays except that some of them could be open about once a month on a fixed date). Furthermore, they display a decreased preferences for Social visits on Thursdays, and Leisure activities on Tuesdays and Sundays. The day-error-scale (DaySTDEV) values show that in case of Grocery shopping the random circumstances on the day (e.g., weather conditions) are less influential than in the case of the other activity groups. In other words, grocery shopping will be done (almost) regardless of the circumstances of the day.

The δ parameters and φ parameters indicate in two different ways the possible interactions between activities. The δ estimates represent whether activities within the activity groups affect the need for an activity from the column activity group. The φ parameters, on the other hand, show whether the utility of the column activity group is influenced by conducting an activity of the activity group considered on the same day. The results of the δ estimations show some significant parameters. E.g., a social-visit activity increases a need for Non-daily shopping (or people who often undertake social visits also frequently conduct non-daily shopping activities). Conversely, Non-daily shopping raises the need for a Social visit as well. Furthermore, Social visits and Sports activities increase the need for Leisure and Leisure increases the need for Sports (or people who often engage in social/sports activities also tend to do leisure activities more often). The only significant parameter with a negative sign is Social visits in case of Sports activities: Social visits decrease the need for Sports (or people who often undertake social visits tend to undertake sport activities less frequently than others). This might be caused by the fact that sports activities done together with others also satisfy the need for social contact. The results of the φ estimates show several interaction effects between activities: the utility of Grocery shopping decreases when a social visit is conducted on the same day; the utility of Non-daily shopping reduces when Grocery shopping is done on the same day, and the utility of a leisure activity diminishes when another leisure activity is performed on the same day. In other words, interaction effects among leisure activities exist: if a leisure activity is performed the probability of conducting another leisure activity on the same day decreases. This counts for shopping activities as well. On the contrary, the utility of Leisure rises when a non-daily shopping activity is conducted on the same day.

Some variables can also have an impact on the threshold value. For this study we only included the number of work hours by day of the week and car availability as an explanatory variable. The results show that the amount of time spent on paid work on a day increases the threshold value and, hence, decreases the probability of conducting the activity on that day. In this study, car availability does not have a significant impact on the threshold value.

The Rho square of the estimation was calculated by using the log-likelihood of the estimated model and the log-likelihood of a null-model. A complete null model, where all parameters are set to zero is not a good indicator of the reference goodness-of-fit in that the need-growth and threshold value cannot be equal to zero. In order to find an appropriate reference goodness-of-fit we used 'mean' values of the intercepts of β and a value close to the threshold intercept parameter to calculate the Log-likelihood of a null-model. For all intercept β parameters we chose 0.5 and for the threshold intercept a value 2. The Rho-square calculated on that basis is 0.557. This indicates a good performance of the model. However, the adjusted Rho-square is noticeably lower with a value of 0.455, which suggests a lack of data compared to the number of variables incorporated in the model.

DISCUSSION AND CONCLUSIONS

This paper described a first attempt of estimating a model of activity generation that is based on notions of dynamic needs with the aim to reveal (positive or negative) substitution relationships between activities. Data used were especially collected for this purpose. The survey included, for a list of 37 activities, the time elapsed since last conducting the activity, if the activity was conducted the day before and if and when the activity was already planned.

The purpose of the present study is 1) to show that it is possible to specify a model and collect data which can be used to estimate the parameters of a dynamic need-based activity generation model and 2) to identify interactions between activities such as to find out to what extent activities are substitutable in the framework of the need-based model. Although the size of the sample is somewhat limited for the number of variables included in the model, we demonstrated that the developed methodology is feasible. The results of the parameter estimations indicate that several socioeconomic and dwelling variables have an impact on episode interval timing and day choice decisions of the shopping, social, leisure and sports activities considered in the present study. Day preferences and interaction parameters show significant effects as well. Especially the fact that interaction effects are significant is highly relevant because it suggests that activities cannot be assumed independent when generating the dynamics of activity participation.

New data should be collected all year round, to capture seasonal influences, and in larger amounts. An interesting avenue is to validate the results with data from a national travel survey, such as for example the Dutch travel survey (called the MON). The Bayesian estimation method used in the present study supports pre- specification of a-priori distributions of parameters that could be set based on other data sources such as the MON. In that approach, data collected specifically for the model would be used for fine tuning rather than estimating parameters from scratch. There are also meaningful ways of extending the model, e.g., by incorporating the effects of travel time and cost on activity participation choice. Furthermore, we plan to carry out analyses on the data collected in order to identify to what extent planned activities affect activity scheduling decisions.

TABLE 4 Estimation results (significant estimates in bold)

	Grocery shopping		Non-daily shopping		Social visits		Leisure		Sports	
Variable	estimate	t-value	estimate	t-value	estimate	t-value	estimate	t-value	estimate	t-value
β 0	**0.632**	**22.734**	**0.138**	**6.316**	**0.283**	**4.597**	0.027	0.730	0.061	0.843
β Male	0.015	0.148	-0.013	-0.264	-0.052	-1.108	-0.138	-1.476	-0.048	-0.395
β Age30-	0.049	0.286	0.127	1.792	-0.025	-0.655	**-0.112**	**-2.130**	-0.089	-0.731
β Age4050	0.025	0.242	**-0.074**	**-3.512**	-0.019	-0.166	**-0.156**	**-2.196**	**-0.155**	**-2.044**
β Age5060	0.020	0.156	**-0.079**	**-6.021**	**0.128**	**2.160**	**-0.133**	**-3.585**	0.015	0.147
β Age60+	0.079	1.106	-0.005	-0.049	**0.220**	**6.840**	**0.058**	**5.075**	0.100	1.214
β Hh_sd_c	0.108	0.705	-0.048	-0.727	-0.019	-0.384	-0.013	-0.245	0.022	0.467
β Hh_s_no	0.070	1.157	**-0.032**	**-4.820**	0.037	0.335	-0.053	-0.505	-0.115	-1.409
β Dwgarden	**-0.093**	**-2.085**	0.053	0.911	-0.046	-1.005	**0.155**	**4.343**	**-0.214**	**-3.675**
β Ibav	0.114	1.919	-0.038	-0.440	0.172	1.721	**-0.102**	**-4.346**	-0.102	-0.975
β Iaav	0.066	0.571	-0.088	-1.429	-0.049	-0.575	**0.186**	**4.146**	-0.036	-0.972
β Edul	**0.180**	**3.826**	-0.128	-1.789	-0.151	-1.223	0.007	0.233	0.068	0.958
β Eduh	**0.159**	**2.367**	0.069	0.995	**-0.179**	**-6.163**	-0.078	-1.816	**-0.240**	**-5.771**
β Aych06	-0.004	-0.053	**0.197**	**2.670**	-0.120	-0.935	-0.112	-1.318	0.127	0.833
β City	**0.148**	**3.556**	0.017	0.171	-0.072	-1.524	**-0.210**	**-14.998**	0.168	1.028
α Mon	0.236	0.812	0.090	0.279	-0.143	-0.811	-0.125	-0.631	-0.009	-0.026
α Tue	0.144	1.068	-0.007	-0.024	0.354	1.397	**-0.372**	**-2.054**	**0.403**	**3.217**
α Thu	-0.089	-0.287	-0.389	-0.927	**-0.647**	**-4.943**	-0.140	-0.883	-0.300	-1.088
α Fri	0.311	0.730	0.083	0.389	-0.022	-0.111	0.167	0.468	-0.108	-0.920
α Sat	**0.496**	**3.396**	0.201	0.833	**0.178**	**2.317**	0.208	1.124	0.089	0.360
α Sun	**-0.378**	**-2.836**	-0.048	-0.171	**0.406**	**5.486**	**-0.459**	**-4.829**	0.087	0.422
DaySTDEV	**1.227**	**2.663**	**3.274**	**6.852**	**3.278**	**15.539**	**2.959**	**9.820**	**3.627**	**17.210**
δ										
δ GS	0.006	0.011	0.446	1.865	-0.018	-0.072	0.185	0.880	-0.371	-0.815
δ NDS	0.077	0.302	0.236	0.805	**0.245**	**2.503**	0.080	0.313	-0.292	-0.992
δ SV	0.095	1.396	**0.539**	**4.898**	0.222	1.163	**0.316**	**1.984**	**-0.257**	**-11.820**
δ Ls	0.012	0.198	0.106	0.930	0.063	1.456	0.210	1.256	**0.423**	**5.567**
δ Sp	-0.101	-0.513	0.066	0.249	0.026	0.104	**0.266**	**2.661**	-0.160	-1.664
φ										
φ GS	-0.018	-0.033	**-0.209**	**-3.086**	-0.136	-0.617	0.190	0.673	0.340	0.745
φ NDS	-0.030	-0.123	-0.301	-0.812	0.390	1.710	**0.502**	**5.761**	-0.478	-1.289
φ SV	**-0.424**	**-12.307**	0.079	0.218	0.338	0.807	-0.095	-0.299	0.323	0.854
φ Ls	-0.447	-1.091	-0.051	-0.133	0.346	0.798	**-0.496**	**-2.166**	-0.084	-0.205
φ Sp	0.548	1.811	-0.035	-0.094	-0.203	-0.386	-0.424	-1.263	-0.130	-0.341
All activities										
Thr0	**1.592**	**9.706**							LL	-711.166
ThrTswork	**0.201**	**9.320**							LL0	-1604.770
ThrcarA	-0.019	-0.198							rho square	0.557
ThrcarO	0.027	0.303			Nr. of obs.	2620			rho sq (adj.)	0.455

REFERENCES

[1] Bhat, C.R., and S.K. Singh. A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A*, 34, 2000, pp. 1-22.

[2] Arentze, T.A., and H.J.P. Timmermans. A learning-based transportation oriented simulation system. *Transportation Research Part B*, 38, 2004, pp. 613-633.

[3] Pendyala, R.M., R. Kitamura, A. Kikuchi, T. Yamamoto, and S. Fujii. Famos, The Florida activity mobility simulator. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

[4] Roorda, M.J., E.J. Miller, and K.M.N. Habib. Validation of TASHA: a 24-hour Activity Scheduling Microsimulation Model. *Transportation Research Part A: Policy and Practice*, 42 (2), 2008, pp. 360–375.

[5] Habib, K. M. N., and E. J. Miller. Modelling Daily Activity Program Generation Considering Within-Day and Day-to-Day Dynamics in Activity-Travel Behaviour. *Transportation*, 35 (4), 2008, pp. 467–484.

[6] Chapin, F.S. *Human activity patterns in the City*. John Wiley and Sons, New York, 1974.

[7] Miller, E. An Integrated Framework for Modelling Short- and Long-run Household Decision-making. Presented at the Progress in Activity-Based Analysis Conference, Maastricht, The Netherlands, 2004.

[8] Axhausen, K.W. (Ed.). *Moving Through Nets: The Physical and Social Dimensions of Travel*. Elsevier, Oxford, 2006.

[9] Meister, K., M. Krick, and K.W. Axhausen. A GA-based Household Scheduler. *Transportation*, 32 (5), 2005, pp. 473–494.

[10] Bradley, M., J.L. Bowman, and B. Griesenbeck. SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. *Journal of Choice Modelling*, 3(1), 2010, pp. 5-31.

[11] Arentze, T.A., and H.J.P. Timmermans. A need-based model of multi-day, multi-person activity generation. *Transportation Research B*, 43, 2009, pp. 251-265.

[12] Nijland, L., T. Arentze, and H. Timmermans. Eliciting the Needs that Underlie Activity-Travel Patterns and Their Covariance Structure: Results of Multimethod Analyses. *Transportation Research Record*, 2157, 2010, pp. 54-62.

[13] Bayarma, A., R. Kitamura, and Y.O. Susilo. Recurrence of daily travel patterns: stochastic process approach to multiday travel behavior. *Transportation Research Record*, 2021, 2007, pp. 55-63.

[14] Schlich, R. and K.W. Axhausen. Habitual travel behaviour: evidence from a six-week travel diary. *Transportation* 30 (1), 2003, pp. 13–36.

[15] Bhat, C.R., S. Srinivasan, and K.W. Axhausen. An analysis of multiple interepisode durations using a multivariate hazard model. *Transportation Research Part B* 39 (9), 2005, pp. 797–823.

[16] Arentze, T.A., D. Ettema, and H.J.P. Timmermans. Estimating a model of dynamic activity generation based on one-day observations: Method and results. *Transportation Research Part B*, 2010, in press.

[17] Arentze, T.A., D. Ettema, and H.J.P. Timmermans. A longitudinal household based model of activity generation: approach and estimation on one-day activity-travel data. Presented at the 12th Conference on Travel Behavior Research, Jaipur, India, 2009.

[18] Rossi, P.E., G.M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. John Wiley & Sons, Sussex, UK, 2005.

[19] Timmermans, H.J.P., T.A. Arentze, M. Dijst, E. Dugundi, C-H. Joh, L.L. Kapoen, C. Maat, K.J. Veldhuisen. Amadeus: a framework for developing a dynamic multi-agent, multi-period activity-based micro-simulation model of travel demand. Presented at 81st Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.

[20] Sun, Z., T.A. Arentze, and H.J.P. Timmermans. A Heterogeneous Latent Class Model of Activity Rescheduling, Route Choice and Information Acquisition Decisions under Multiple Uncertain Events. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.

[21] Bricka S., and J. Zmud. Impact of Internet Retrieval for Reducing Nonresponse in a Household Travel Survey. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.