MODELING HOUSEHOLDS ACTIVITY PARTICIPATION DECISIONS IN A RULE-BASED SYSTEM OF TRAVEL DEMAND

Renni ANGGRAINI
Lecturer
Department of Civil Engineering
Syiah Kuala University
Darussalam – Banda Aceh
Indonesia
Fax: +62-651-52222
E-mail: renni.anggraini@gmail.com

Theo A. ARENTZE
Associate Professor
Faculty of Architecture, Building & Planning
Eindhoven University of Technology
PO BOX 513, 5600MB, Eindhoven
The Netherlands
Fax: +31-40-243-8888
E-mail: t.a.arentze@tue.nl

Harry J.P. TIMMERMANS
Professor
Faculty of Architecture, Building & Planning
Eindhoven University of Technology
PO BOX 513, 5600MB, Eindhoven
The Netherlands
Fax: +31-40-243-8888
E-mail: h.j.p.timmermans@tue.nl

Tao Feng
PhD, Postdoctoral Researcher
Faculty of Architecture, Building & Planning
Eindhoven University of Technology
PO BOX 513, 5600MB, Eindhoven
The Netherlands
Fax: +31-40-243-8888
E-mail: t.feng@bwk.tue.nl

Abstract: This paper describes an empirical derivation of an activity participation choice model at the household level decisions taking into account the allocated activities and joint activity participation of household heads in discretionary activities. The households that we consider here are two-heads households; each is either a worker or non-worker. Attributes of households, such as, for example the presence of young children, attributes of the work activities and space-time settings are considered as explanatory variables. To deal with this large set of attributes and account for non-linear relationships between the variables, a decision tree induction method – CHAID – is used to derive a decision tree model. We show how the decision tree model can be used as a component in an activity-scheduling model, ALBATROSS, to predict travel demand in an activity-based-micro-simulation system. The model shows a satisfactory performance as indicated by its goodness-of-fit on validation data.

Key Words: *joint activity participation, activity-based model, rule-based model, household travel behavior*

1. INTRODUCTION

It has become progressively more evident that travel choices depend on choices to participate in activities. Recognizing that travel is a demand derived from individuals' needs to perform out-of-home activities, researchers in travel demand modeling have become increasingly interested in analyzing and predicting individuals' decisions about activity participation. The social context is considered important; individuals may share a household and decisions to participate in activities occasionally need to be synchronized or adapted to the needs of other household members. Therefore, inter-dependency in travel choices and activity-agenda choices between individuals within households constitutes a relevant research topic.

Operational models of individual's activity-scheduling behavior have begun to emerge recently. Activity-scheduling models share the objective to predict the sequence of decisions that leads to an observed activity pattern of households/individuals. Activity-based models

aim at predicting on a daily basis and for a household which activities are conducted, by whom, for how long, at what time, the location, and which transport mode is used when traveling is involved (Arentze and Timmermans, 2000, 2004, 2005; Miller and Roorda 2003). There has been some analytical studies on the interactions of individuals within households (Gliebe and Koppelman, 2002, 2005; Scott and Kanaroglou, 2002; Srinivasan and Bhat, 2004), but fewer attempts to integrate these interactions in comprehensive, multi-faceted activity-scheduling models.

ALBATROSS is one of the few operational activity-based models that incorporates household-level decision making (Arentze and Timmermans, 2000, 2004, 2005). It is a rule-based computational process model developed for The Dutch Ministry of Transportation, Public Works and Water Management. ALBATROSS differs from other models, which use utility maximization as a framework for modeling activity-travel patterns. In contrast, ALBATROSS uses IF-THEN rules as a formalism to represent and predict activity-travel choices of individuals and households. The decision rules are extracted from activity diary data in the form of a decision tree by using a CHAID-based decision tree induction method. The rules predict actions/choices in a probabilistic manner to reproduce non-systematic variance in choice behavior. In this paper, we use the same rule-based approach to model activity participation decisions at the household level. The present study is part of a larger project aimed at refining ALBATROSS by improving the representation of some aspects of household-level decision in activity-travel decisions.

The study focuses on two-heads households and considers the joint decision making of individuals related to household task allocation and joint participation in activities. Specifically, the objective of this paper is to identify and model household-level decision making regarding activity participation. We propose an activity classification and identify the activity types that likely relate to needs at the household level and can be allocated to household members. We use the term allocated activities to refer to these activities that are allocated to a household member. In addition, the proposed model predicts the activities that are conducted jointly by the household heads on a given day.

The remainder of this paper is arranged into several sections. First, the next section describes the proposed model of activity-travel scheduling that accounts for household-level decisions. Subsequently, the method used to derive rules that account for joint activity participation is described. The CHAID-based algorithm that is used to induce decision trees is briefly explained. The subsequent sections describe the activity-travel data set used to derive the decision-tree model and the results. The paper concludes with some conclusions and a brief discussion of forthcoming work.

2. THE MODEL

This paper is part of a comprehensive project aimed at refining ALBATROSS. This model predicts for each household in a studied population the schedule of activities and trips of each household head on a particular day. The activity schedules of children are not predicted by the model.

In the existing ALBATROSS model, the activity scheduling process consists of four major components: (1) work activity generation (including timing, duration, location and transport mode choice for each work trip), (2) secondary fixed activity generation (including timing, duration and location), (3) flexible activity generation (including timing, duration and location), and (4) trip-chaining decisions and transport mode choice for each tour.

The activity types distinguished are grouped into *fixed activities* and *flexible activities*. A fixed activity can be considered as an activity that has to be conducted within a particular time horizon on a regular basis, due to longer term commitments made by the individual. In contrast, a flexible activity is an activity that can be conducted freely at any time. Examples of fixed activities are work and escorting a child to school, while most non-work activities are flexible activities.

In order to identify household-level decision making in activity scheduling and taking into account available activity data, we cluster activities into 10 activity categories as displayed in Table 1. These activities are similar to the classification used in the current ALBATROSS model. However, to account for household decision making, we differentiate between personal (P) and household (HH) activities. Personal activities involve only one household head and are conducted due to personal commitments or to satisfy personal needs. Household activities are allocated activities to satisfy household needs and discretionary activities that involve joint activity participation.

Table 1 Activity Classification

No	Activity	Personal (P) or Household (HH) Level	Scope of Activities
1	Work	P	Full-time and part-time
2	Business	P	Work-related
3	Other	P	Other mandatory activity (school, etc)
4	Bring/get person	HH	Drop-off/pick-up children/spouse to a certain location
5	Shop-1-store	HH	Shopping 1 store
6	Shop-n-store	HH	Shopping n stores
7	Service-related	HH	Renting movie, getting (fast) food, institutional purposes (bank, post office, etc)
8	Social-independent	P	Visiting friends, relatives , etc
	Social-joint	HH	Same joint
9	Leisure-independent	P	Sports, café/bar, eating out, movie, museum, library
	Leisure-joint	HH	Same joint
10	Touring-independent	P	Making a tour by car, bike, or foot (e.g., letting out the dog, etc)
	Touring-joint	HH	

Although the current ALBATROSS model does consider interactions between persons in activity-travel of household heads, the decision mechanisms are defined at the individual level. Scheduling steps are made alternately between the household heads whereby at each decision step the schedule of one household head is used as a condition influencing the decision of the spouse. Some aspects, such as task allocation, car allocation, and joint participation in activity and travel, however, require household decision making (Anggraini, *et al.*, 2007).

In order to better capture within-household interactions, we identify the facets of activity-travel behavior that require household-level decisions and expand the household activity-travel scheduling process regarding each of seven major components, as shown in Figure 1. The left-hand side of the scheme shows the major steps in the scheduling process and the right-hand side whether the activity is treated as a personal or as a household activity. Thus, by explicitly modeling the interactions between the two heads in activity-travel decisions for out-of-home activities, we expect that the performance and the sensitivity of ALBATROSS will be improved.

In the first step, a work activity pattern consisting of at most two work episodes is generated for each household head. At the person-level, it involves the choice of number of work episodes, and the start time, duration and location of each work episode. The decision who is using the car for the work commute, in particular in car-deficient households, is a household decision and modeled as such.

Next, the part of the schedule that is related to secondary fixed activities, such as business and other mandatory activities, is modeled. Different from other facets, these secondary fixed activities are treated as individual/person-level decisions because these activities do not involve household tasks or discretionary activities. The model predicts which types of secondary fixed activities are conducted that day, and for each of these activities the start time, duration and location. Additionally, it identifies the potential trip-linkages of each scheduled activity with the work activity/activities. The next step deals with the allocated and discretionary activities. First, at the household-level, participation choices are predicted for allocated activities (bring/get, shop-1-store, shop-n-store, service), joint discretionary activities (social-joint, leisure-joint, and touring-joint) and independent discretionary activities (social, leisure, touring). Then, in case of allocated activities, the model predicts who will be involved in these activities (single household head (task allocation) or both (joint activity participation). The selection of joint and household-tasks activities and the allocation of household task activities are household level decisions. The selection decisions at household level will be the focus of this paper.

Timing of allocated and discretionary activities takes place at the next stage. The start time, duration, and time use of activity categories both at the household level and person-level are predicted. Having defined the timing and duration, trip-chaining, location and transport mode choice for each activity or tour are predicted. The timing, duration and location choices are household or person choices, depending on whether it involves a joint activity or not. It is essential to note that in all decision-tree models, the results of earlier decisions are used as conditions for each next decision and that the process results in a complete schedule for each person.

Thus, embedded in this larger scheduling model, the present paper focuses on activity participation decisions that involve both household heads either because the activity concerns a household task (an allocated activity) or joint participation in a discretionary activity.

3. HOUSEHOLD ACTIVITY PARTICIPATION MODEL SPECIFICATION

The need to participate in particular activities will vary between households dependent on household attributes. For example, households with school-going children may have to escort their children to school and pick them up again. Thus, we need to model the conditions under which different households will participate in each of the classified activity types.

The model acknowledges that an allocated activity can also be conducted jointly. However, since we are concerned here with the question whether an allocated activity is included in a household schedule or not, allocation of an allocated activity to the male, female or both persons jointly is considered a next household decision to be handled by another model. In this study, we focus on decisions to yes or no select activities at household level, which are related to allocated activities and joint activities of a discretionary nature.

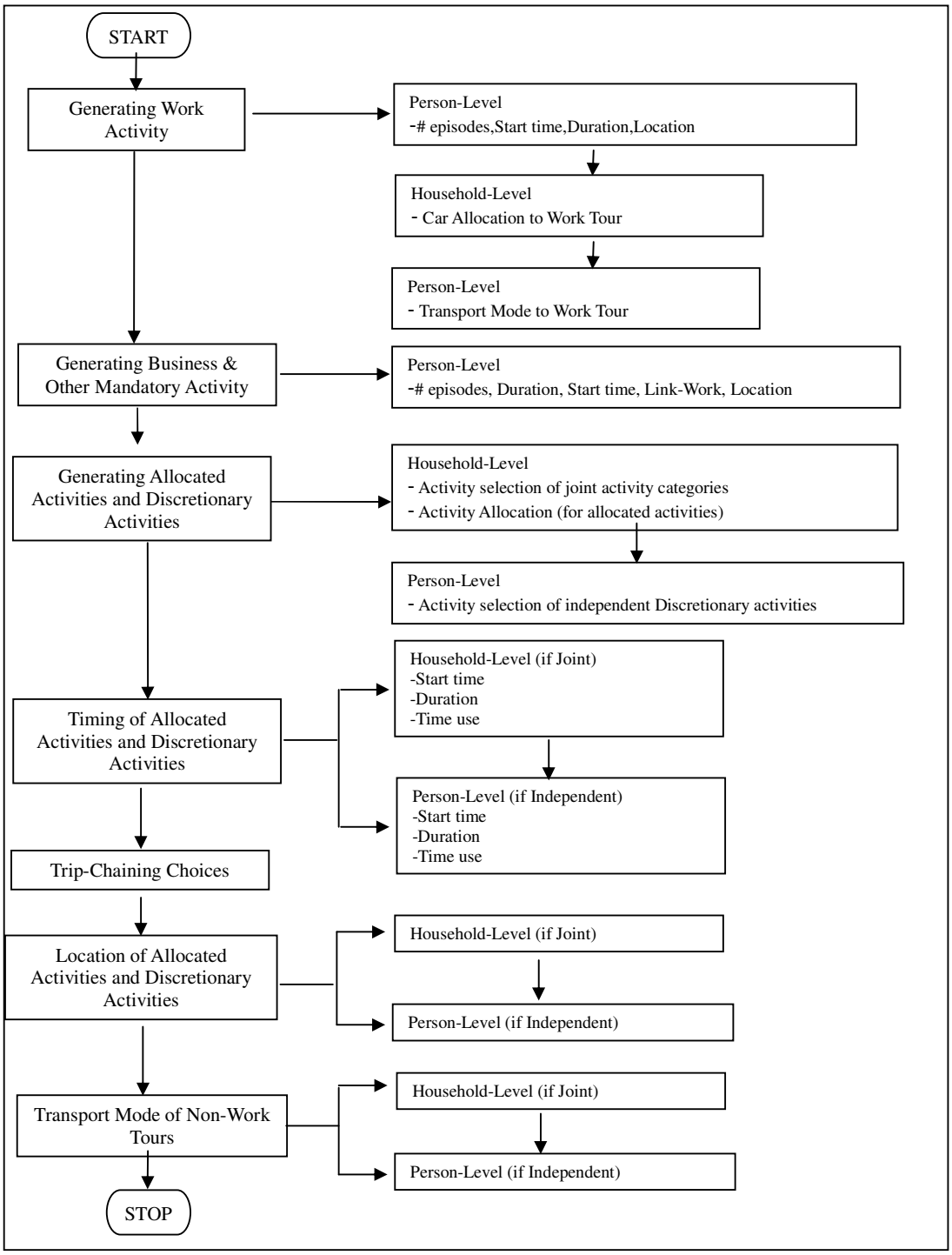
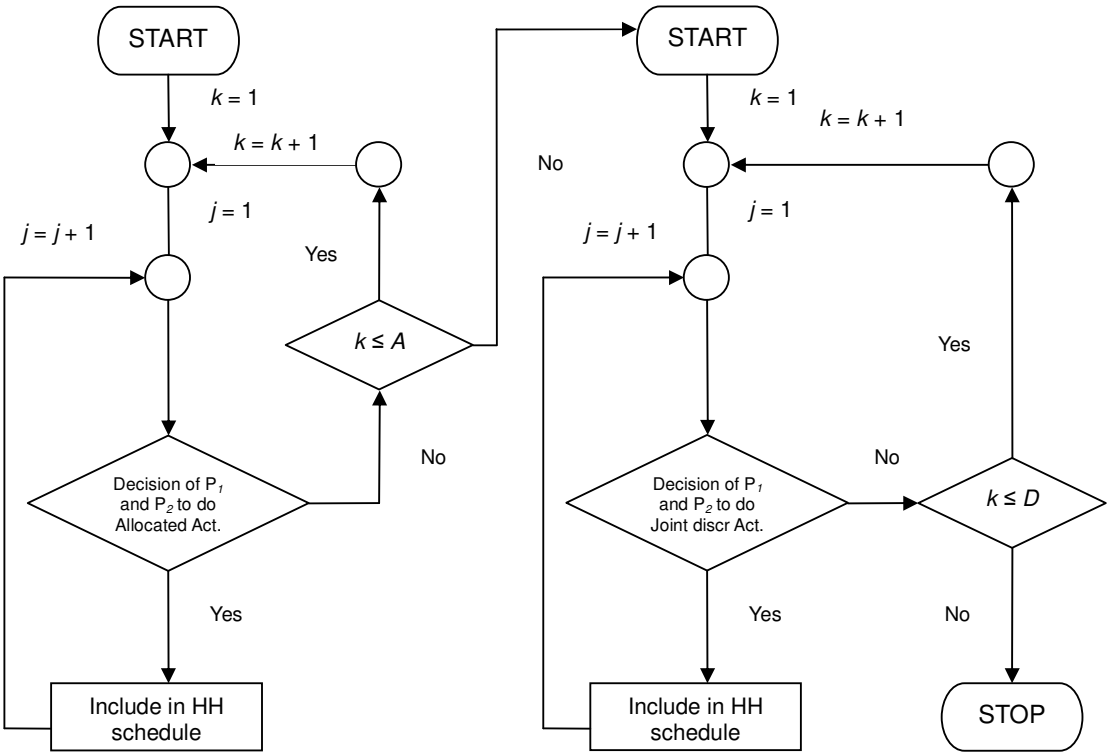


Figure 1 Household Activity-Travel Scheduling Process of a Refined ALBATROSS



Note:
k = index of activity category
j = index of activity of the k-th category
A = number of allocated activity categories (A=4)
D = number of joint discretionary activity categories (D=3)
P_1 = person 1 (male head)
P_2 = person 2 (female head)

Figure 2 The process of including a particular activity category to household schedule

We propose the following process model for the activity participation decisions (see Figure 2). First, activities are considered sequentially based on a pre-defined priority ordering of activity categories. Allocated activities are scheduled before discretionary activities. Furthermore, a particular priority order is assumed within each of these two main categories as well, which corresponds to the order in which the activities are listed in Table 1. For each activity category in order of priority, a decision tree decides whether an activity of that category will be conducted under a set of relevant conditions. If the answer is *yes*, then the next decision by the same decision tree involves whether or not a second activity of the same category is to be selected. This is repeated until the answer is negative. Then, the next activity category is considered repeating the same process. The process continues until a negative decision is generated for the last activity category. To model activity selection decisions for both main categories we develop and test a (single) decision tree as explained in the sections that follow.

4. DATA

The data used for this purpose is derived from the Dutch National Travel Survey (MON=*Mobiliteit Onderzoek Netherlands*). The data used was collected in 2004 covering all

of the Netherlands. The survey is conducted on a regular basis to obtain travel and activity information of residents in the Netherlands. Although it is a one-day travel diary, the collected data is more complete regarding activities at trip destinations than its predecessor travel survey called OVG. It is a household survey where data is collected of all household members on the diary day as well as general information about household and individual attributes such as, gender, age, vehicle ownership and driving license ownership, home location, individual income, occupation, number of working hours per week, etc.

Additionally, respondents are invited to give information about all trips made on a designated day as well as out-of-home activities at trip destinations. Information for each trip includes start time, trip purpose, destination, activity type at the destination, and transport mode. Situational variables and spatial geography are also reported. All in all, this survey provides an exclusive data source to analyze activity-travel behavior of Dutch residents. In the data collection, 29221 households filled out a one-day travel/activity diary and 28600 of these households fit the criteria for being considered in ALBATROSS. The data were transformed to an activity-diary data format for the current estimation purpose. In this study, given the focus on two-heads households, households consisting of a single head are not included in the analysis. Then, there are 18037 households used for deriving the envisioned decision tree.

5. ANALYSES

5.1 Decision Tree Induction

We applied a CHAID-based tree induction method to identify the rules that describe which choices (i.e., actions) are made under which conditions. CHAID (*Chi-square Automatic Interaction Detector*) generates non-binary trees, i.e., trees where more than two branches can be attached to a single root or node, based on a relatively simple algorithm that is particularly well suited for the analysis of larger datasets. Other decision tree induction systems are C4.5 (Quinlan, 1993) and CART (Breiman et al., 1984). CHAID relies on the *Chi-square* test to determine the best next split at each step. CHAID generates a decision tree by splitting subsets of the space into two or more nodes repeatedly, beginning with the entire data set (Kass, 1980). In the present context, this means that it finds the conditions that allow us to differentiate between activity participation or not for each activity type. To determine the best split at any node, it evaluates each predictor variable and merges any allowable pair of categories of that predictor variable if there is no statistically significant difference within the pair with respect to the target variable. The split that maximizes a significance value of a Chi-square test, after adjustment for multiple tests (Bonferroni adjustment), across predictor variables is used for splitting if the split is significant. The process is repeated for each newly created group until no more significant splits are found. This process of extracting the rules is the same as the one used in the original ALBATROSS model. In order to develop the decision tree, 75% of the cases were used for training, i.e. extracting the decision tree (set of rules) from the data, while the remaining cases were used for validation. Generally, in deriving ALBATROSS decision models, the attributes of the household and person, space-time settings and the attributes of the completed part of the schedule at the previous stage of the scheduling process are used as predictors, while the choice outcome represents the action variable.

The CHAID decision tree induction method allows one to define a threshold (alpha) for splitting in terms of a significance level for the Chi-square (χ^2) measure and a minimum number of cases at the leaf nodes. Alpha was set to 5% and the minimum number of cases at

leaf nodes to 75. As a measure of prediction accuracy, the expected hit ratio is calculated. ALBATROSS uses a probabilistic action-assignment rule and therefore the hit-ratio measure used here represents the expected proportion of cases predicted correctly when a probabilistic action-assignment rule is used. It is calculated as: $\frac{1}{N}\sum_{kq}\frac{(f_{kq})^2}{N_k}$ where f_{kq} is the frequency of the qth action at the kth leaf node, N is the total number of cases and N_k is the number of cases at the k-th leaf node. Note that the expected hit ratio is comparable to a likelihood measure and, generally, yields lower scores than the deterministic counterpart of the measure.

5.2 Deriving Impact Tables

Decision trees derived from data may become very large and complex and, consequently, be difficult to interpret. This holds true particularly for the present application where the number of choice observations is very large. Arentze and Timmermans (2003) therefore developed a method to derive elasticity information from rule-based models to facilitate interpretation, which we will use here to describe the results of the tree induction. The principle of the proposed method is straightforward. After having derived a rule-based model from the training data, the model is used to predict for each condition variable a frequency table with the levels of the condition variables in rows and the frequency distribution across the levels of the target variable (i.e., the action variable) in columns. The frequency table is generated by applying the model as many times as there are levels of the condition variable. In each run, each training case is assumed to take on the level considered in the condition variable. Thus, the results of a run indicate the predicted frequency distribution for the action variable assuming that each training case has the selected level of the condition variable. Then, the impact of the condition variable is measured as the Chi-square for the frequency table. Formally:

$$IS_s = D(\mathbf{F}_s) \tag{1}$$

where D is a Chi-square measure of the frequency table generated ($\mathbf{F}_s$) for condition variable s. This measure can be decomposed into a measure of impact of each level of the action variable, as follows:

$$IS_{si} = D(\mathbf{F}_{si}) \tag{2}$$

where again D is a chi-square measure and $\mathbf{F}_{si}$ is the vector of predicted frequencies of the i-th action under the levels of the s-th condition variable. Apart from impact size, we also use a measure of the direction of impact proposed by Arentze and Timmermans (2003), which is defined as:

$$MS_{si} = \frac{\sum_{j=2}\left(f_{ij} - f_{i,j-1}\right)}{\sum_{j=2}\left|f_{ij} - f_{i,j-1}\right|} \tag{3}$$

where f_{ij} is the predicted frequency of action i under the j-th level of condition variable s. This measure can be interpreted as a measure of *monotonicity*. If the condition variable has a monotonically increasing impact on the frequency of action i across the levels of the condition variable, then MS_{si} equals 1 and if it has a monotonically decreasing impact it equals -1. Any

value in between these extremes indicates that the impact is non-monotonous in the direction indicated by the sign across the range of the condition variable.

5.3 Condition and Action Variables

The model assumes that the choices or actions of individuals and households depend on a set of condition variables. Table 3 portrays the condition variables that were used as input to the algorithm. The condition variables concern household level (including accessibility indicators), individual level, and activity level variables. Note that in this stage of the scheduling process, work, business, and other mandatory activities are known. However, only information related to the work activity such as number of work activities, duration of each work activity, mode to work place, and total time engaged in work activity is fully used as condition variables. Continuous condition variables, such as for example duration, are discretisized by using an equal-frequency interval method which divides a continuous variable into *n* parts, in which each part contains approximately the same number of cases.

It is important to note that some of the variables are related to the *household level* or *person level*, while others are defined at the *schedule level* or *activity level*. The presence of the children in a household is taken as a condition variable, together with other household attributes, such as urban density (of the residence location), household composition, and socio-economic class (#1-5 in Table 3). Given the selection of double-head households, household composition refers to only three household types: double-one-worker, double-two-workers and double-no-workers. The number of cars in the household is also included as a household variable (# 21). A final set of household-level variables relates to measures of accessibility given the home location of the household. On this level, 8 variables (#10-17) are included:

1. Daily goods sector: number of employees within 3.1 km
2. Non-daily goods sector: number of employees within 4.4 km
3. All sectors: number of employees within 4.4 km
4. Size of population within 3.1 km
5. Daily goods sector: distance within which 160 employees work
6. Non-daily goods sector: distance within which 260 employees work
7. All sectors: distance within which 4500 employees work
8. Distance within which 5200 people live

Note that attributes related to individuals can be incorporated only if they are specified explicitly for the male and female head. Individual attributes such as work status and age are explained in #6-9 in Table 3. The work status attribute indicates for the male or the female householder, whether the person has no work, part-time work, or full-time work. Those who work more than 30 hours per week are considered full-time workers. Age and possession of driving license are also used as individual attributes (# 8 & 9 & 22 & 23).

Attributes or conditions defined at the schedule level include the number of work activities, if any. Since there are only very few diaries where more than two work activities are included, we limit the number of work activities to 2 work episodes per person (#18-20). Duration of work activity conducted by male or female and the total work duration across male and female in a household are also used as condition variables. Transport mode used to travel to the work place by each (male or female) worker is also used as condition variable (#27 & 28).

Table 3 Condition Variables for Household Activity Participation Model

No	Variable	Classification	Acronym	Category
1	Urban Density	0=most densely , 4= least densely	Urban	Ordinal
2	Household Composition	2=2 heads, 1 worker, 3=2 heads, 2 workers, 4=2 heads, no workers	Comp	Nominal
3	Youngest child in HH	0=no children, 1=<6, 2=6-11, 3=12-17 yr	Child	Nominal
4	Day of the week	0=Monday to 6=Sunday	Day	Nominal
5	Socio-economic class	0=low, 1=low-mid, 2=mid-high, 3=high	SEC	Ordinal
6	Working status – M	0= non-worker, 1= part-time, 2= full-time	WstatM	Nominal
7	Working status – F	0= non-worker, 1= part-time, 2= full-time	WstatF	Nominal
8	Age of person – M	0=<35, 1=35-<55, 2= 55-<65, 3= 65-<75, 4= 75+ years	AgeM	Ordinal
9	Age of person – F	0=<35, 1=35-<55, 2= 55-<65, 3= 65-<75, 4= 75+ years	AgeF	Ordinal
10	Accessibility – 1	0=<=115, 1=<=253, 2=<=307, 3=<=507, 4=<=675, 5=>675	nEmp1	Ordinal
11	Accessibility – 2	0=<=395, 1=<=635, 2=<=762, 3=<=938, 4=<=2525, 5=>2525	nEmp2	Ordinal
12	Accessibility – 3	0=<=8785, 1=<=12995, 2=<=16120, 3=<=20199, 4=<=70314, 5=>70314	nEmp3	Ordinal
13	Accessibility – 4	0=<=5050, 1=<=8845, 2=<=13217, 3=<=16833, 4=<=22884, 5=>22884	SizePop	Ordinal
14	Accessibility – 5	0=<=71, 1=<=127, 2=<=165, 3=<=202, 4=<=346, 5=>346	Dist1	Ordinal
15	Accessibility – 6	0=<=92, 1=<=145, 2=<=176, 3=<=258, 4=<=334, 5=>334	Dist2	Ordinal
16	Accessibility – 7	0=<=92, 1=<=128, 2=<=201, 3=<=274, 4=<=360, 5=>360	Dist3	Ordinal
17	Accessibility - 8	0=<=0, 1=<=105, 2=<=126, 3=<=163, 4=<=278, 5=>278	Dist4	Ordinal
18	Number of work episodes – M	0=no work, 1=1 ep, 2=2 ep	NworkM	Ordinal
19	Number of work episodes – F	0=no work, 1=1 ep, 2=2 ep	NworkF	Ordinal
20	Number of work episodes – HH	0=no work, 1=1 ep, 2=2 ep, 3=3 ep, 4=4 ep	NworkHH	Ordinal
21	Number of cars in HH	0, 1, 2+	Ncar	Ordinal
22	Driving license possession – M	0= no, 1= yes	DrivM	Nominal
23	Driving license possession – F	0= no, 1= yes	DrivF	Nominal
24	Duration of work act – M (min)	0=0; 1=≤435; 2=436-544; 3=545-575; 4=>575	DurM	Ordinal
25	Duration of work act – F (min)	0=0; 1=≤297.25; 2=297.26-480; 3=481-555; 4=>555	DurF	Ordinal
26	Duration of work act in HH (min)	0=0; 1=≤475; 2=476-566; 3=567-820; 4=>820	DurHH	Ordinal
27	Mode to work place – M	0= non-work act, 1=PT, 2=CP, 3=CD, 4=S	ModeM	Nominal
28	Mode to work place – F	0= non-work act, 1=PT, 2=CP, 3=CD, 4=S	ModeF	Nominal
29	Mode to work place – HH	0= non-work act, 1=PT, 2=CP, 3=CD, 4=S	ModeHH	Nominal
30 – 35	Time available for non-work act (1-6) – M (min)	0=0, 1=<=30, 2=30-60, 3=60-90, 4=90-120	Time1M-Time6M	Ordinal
36 – 41	Time available for non-work act (1-6) – F (min)	0=0, 1=<=30, 2=30-60, 3=60-90, 4=90-120	Time1F-Time6F	Ordinal
42 – 47	Time available to use car for non-work act in HH (1-6)	0=0, 1=<=30, 2=30-60, 3=60-90, 4=90-120	Time1C-Time6C	Ordinal
48	Schedule includes work act – M	0= no, 1= yes	yWorkM	Nominal
49	Schedule includes business act – M	0= no, 1= yes	yBusiM	Nominal
50	Schedule includes other act – M	0= no, 1= yes	yOthM	Nominal
51	Schedule includes work act – F	0= no, 1= yes	yWorkF	Nominal
52	Schedule includes business act – F	0= no, 1= yes	yBusiF	Nominal
53	Schedule includes other act – F	0= no, 1= yes	yOthF	Nominal
54	# activities 1 currently done in HH	0=0, 1=1, 2=2, 3=3, 4=4, 5=5+	nbr	Ordinal
55	# activities 2 currently done in HH	0=0, 1=1, 2=2, 3=3, 4=4, 5=5+	nsh1	Ordinal
56	# activities 3 currently done in HH	0=0, 1=1, 2=2, 3=3+	nshn	Ordinal
57	# activities 4 currently done in HH	0=0, 1=1, 2=2, 3=3+	nser	Ordinal
58	# activities 5 currently done in HH	0=0, 1=1, 2=2+	nsoc	Ordinal
59	# activities 6 currently done in HH	0=0, 1=1, 2=2+	nlei	Ordinal
60	# activities 7 currently done in HH	0=0, 1=1, 2=2+	ntou	Ordinal
61	HH activity type	1=bring/get, 2=shop-1-store, 3=shop-n-store, 4=service, 5=social-joint, 6=leisure-joint, 7=touring-joint	HHact	Nominal

Note: M =Male; F=Female; ep=episode; PT=Public Transport; CP=Car Passenger; CD=Car Driver; S=Slow (walking, bicycling)

The combination of conditions 27 & 28 defines transport modes used in the household (#29).

In addition, available time in the schedule is captured. In order to identify how much time is available for conducting activities other than the work activity at different times of the day for male and female, we segmented the time period from 8 am to 8 pm into 6 time spans of 2 hours. For each 2 hour-period, we calculated the available time in the schedule and classified this time into 5 categories, where zero means no time left for conducting a non-work activity and the remaining categories identify how much time is left for participating in a non-work activity in multitudes of half an hour (#30-41).

Whether the car is available for a non-work activity is another condition variable. More specifically, the number of cars available and the transport mode(s) used for work activities, if any, of the two household heads are taken into account. Time is segmented in the same way as above (Note: zero means that either the car is not available due to work activities or because no car is available in the household).

Variables #48-53 indicate whether work, business and other mandatory activities (such as go to school) are conducted on the given day or not. Note that these are also schedule-level variables.

The remaining variables (#54-60) are also schedule level variables. These variables indicate for each allocated and joint-discretionary activity category the number of activities that, at the moment of the decision, are included in the schedule as a consequence of previous activity selection decisions. Thus, these variables are dynamic and depend on the assumed priority order of activities. At the time of the first decision, no activity is included in the schedule. Therefore, all these variables (#54-60) are zero in the beginning. For a second decision, the result of the first decision is known and the corresponding variable receives a value of one, and so on. Thus, this set of variables indicates for each current decision the state of the schedule as a result of previous activity selection decisions.

The last variable encodes the activity type that is considered in the current selection decision. This variable has seven levels corresponding to the seven activity categories (allocated and discretionary joint activities) considered in the process model.

As a result, a total of 61 condition variables were defined. The action variable, i.e. the output of the household activity participation decision, indicates whether or not an activity (type) is added to the (household) schedule.

5.4 Results

In total 153,856 observations can be derived from the data set. In order to develop the decision tree, 75% of these cases (115,458) were used for training and the remaining cases were used for validation. Of 153,856 cases, the frequency of observing yes decisions for each activity category can be seen in Table 4. Shop-1-store gets the highest frequency among the activity categories, of about 13,333 decisions.

Table 4 Frequency of observing YES decisions

Activity	Frequency
Bring/get	5,779
Shop-1-store	13,333
Shop-n-store	2,596
Service	2,970
Social-joint	1,825
Leisure-joint	1,281
Touring-joint	865
Total	**28,649**

Given a minimum group size of 75 cases at child nodes and a 5% alpha level, the tree generated by CHAID consists of 386 leaf nodes (decision rules) as can be seen in Table 5. The hit ratio (based on a probabilistic assignment rule) of the model, compared to a null-model (a root-only decision tree) indicates a modest but significant improvement: the hit-ratio of a root-only tree (i.e., null-model) equals 0.697 and the hit ratio of the tree after splitting equals 0.777. A Chi-square-based contingency coefficient of 0.455 confirms that there is a moderately strong impact of the decision tree structure on the action variable. The overall accuracy on the validation set is almost the same; it has dropped only slightly from 0.777 to 0.772.

Table 5 Results of the Household Activity Participation Model

Indicators	**Results**
N alts	2
N cases	115,458
N attr	61
N leafs	386
hit *r(0)*	0.697
hit *r(t)*	0.777
hit *r(v)*	0.772
χ^2	30,194
C	0.455

Note: *N* alts : number of choice alternatives
N cases : number of observations in training data set
N attr : number of attributes
N leafs : number of leaf nodes
hit *r(0)* : expected ratio of correctly predicted cases (null model)
hit *r(t)* : expected ratio of correctly predicted cases (training set)
hit *r(v)* : expected ratio of correctly predicted cases (validation set)
χ^2 : Chi-Square value
C : contingency coefficient

In the tree model, *activity type* is the most significant condition variable, as it is the variable on which the first split is implemented. Due to limited space and given the large number of decision rules, we cannot display the entire results of the decision tree. Instead, in order to give a summary view of the outcome, we will discuss the results of the impact analysis in terms of the IS and MS measures explained above.

Table 6 displays the so-called impact table for the activity participation model. In this case, the choice variable is a binary variable (yes/no decision), so that the MS measures are perfectly correlated ($MS_{yes} = -MS_{no}$). As it appears, activity type (*HHact*) is by far the most important variable for the activity selection decision. The *monotonicity* index MS in this case is close to zero (+/- 0.21) and negative for the yes decision indicating that the frequency of adding an activity decreases across the activity categories in the order they are put, but not monotonically. The second most important variable is day of the week (*Day*). There is a tendency (MS_{yes} = -0.23) of decreasing probability of adding an activity of the activity category concerned to the schedule with increasing values of this variable, which is what one might expect, ever since the first value in this variable is Monday.

Table 6 Impact of Condition Variables of HH Activity Participation Model

No	Variables	IS	IS_{yes}	IS_{no}	MS_{yes}	MS_{no}
1	HHact	107437	91962.17	15452.71	-0.21	0.21
2	Day	4168.91	3390.36	778.64	-0.23	0.23
3	nbr	2053.11	1616.28	436.82	0.23	-0.23
4	Child	1986.13	1601.17	384.98	0.05	-0.05
5	nsh1	1051.52	864.75	186.81	-0.49	0.49
6	nser	790.11	629.86	160.25	0.52	-0.52
7	nshn	501.63	405.30	96.39	0.29	-0.29
8	DurHH	499.02	406.40	92.64	-1.00	1.00
9	ModeF	273.66	221.14	52.53	-0.07	0.07
10	ModeHH	75.3	61.26	13.91	-0.25	0.24
11	Dist1	54.04	43.99	10.05	0.34	-0.33
12	SEC	48.87	39.61	9.23	1.00	-1.00
13	Time3M	26.23	21.22	5.01	0.81	-0.81
14	Time4F	23.57	19.14	4.38	1.00	-1.00
15	DurF	22.29	18.10	4.13	-1.00	1.00
16	ntou	20.8	16.79	4.00	1.00	-1.00
17	NworkHH	15.22	12.47	2.74	-1.00	1.00
18	AgeF	13.02	10.60	2.44	0.12	-0.12
19	yBusiM	12.69	10.37	2.32	-1.00	1.00
20	DrivF	11.34	9.30	2.06	1.00	-1.00
21	yWorkM	8.73	7.13	1.58	-1.00	1.00
22	NworkF	7.59	6.14	1.44	-0.15	0.14
23	AgeM	7.43	6.05	1.35	-1.00	1.00
24	Ncar	6.00	4.81	1.17	0.56	-0.57
25	yOthM	5.76	4.70	1.08	1.00	-1.00
26	Dist2	5.5	4.41	1.03	-0.04	0.02
27	DrivM	4.99	4.07	0.92	1.00	-1.00
28	ModeM	4.68	3.80	0.86	-1.00	1.00
29	Dist3	4.40	3.50	0.83	0.05	-0.06
30	Urban	4.31	3.54	0.81	0.06	-0.05
31	nlei	3.58	2.88	0.67	1.00	-1.00
32	nEmp2	3.52	2.81	0.64	-0.15	0.15
33	DurM	3.39	2.79	0.64	-1.00	1.00
34	Time1F	2.93	2.38	0.55	-0.25	0.25
35	SizePop	2.64	2.09	0.48	0.85	-0.85
36	Time4M	2.48	2.02	0.47	1.00	-1.00
37	NworkM	2.31	1.86	0.43	0.28	-0.29
38	WstatM	1.76	1.43	0.35	-1.00	1.00
39	Comp	1.65	1.40	0.29	0.38	-0.38
40	Time2F	1.56	1.24	0.27	1.00	-1.00
41	Time3F	1.47	1.20	0.28	0.63	-0.66
42	nEmp1	1.32	1.17	0.26	-0.07	0.07
43	Dist4	1.21	0.94	0.21	-1.00	1.00
44	nsoc	1.11	0.92	0.22	1.00	-1.00
45	Time3C	0.92	0.74	0.18	1.00	-1.00
46	Time1M	0.73	0.59	0.14	-0.02	0.03
47	Time4C	0.46	0.41	0.09	-1.00	1.00
48	Time5C	0.37	0.28	0.06	0.23	-0.23
49	nEmp3	0.33	0.28	0.07	1.00	-1.00
50	WstatF	0.17	0.17	0.04	-0.79	0.75
51	yWorkF	0.11	0.09	0.02	-1.00	1.00
52	Time5M	0.09	0.07	0.01	0.42	-0.39

The next most influential variable is the number of bring/get activities already included in the schedule at the moment of the decision (*nbr*). Given a positive sign of MS for the yes decision (MS_{yes} = 0.23), there is a tendency of bring/get activities to generate other activities. However, given that MS is smaller than 1, it does not monotonically increase, at some point the probability of adding activities decreases with increasing number of bring/get activities in the

current schedule.
With respect to the relative impacts of different types of variables, several observations are relevant. In terms of socio-demographic variables, in order of decreasing importance, we find that presence of young children in the household (*Child*), income (*SEC*), age of female head (*ageF*), female has a driving license (*DrivF*), age of male head (*ageM*), number of cars (*Ncar*), male has a driving license (*DrivM*), work status of male (*WstatM*), household composition (*Comp*), and work status of female (*WstatF*) all have an influence.

The first variable that gives the biggest impact, *Child*, clarifies that with increasing level of this variable, the frequency of household activities is not increasing monotonically (MS = 0.05). Interestingly, the MS measure for the variable that gives the second biggest impact, *SEC,* indicates that as income goes up, the frequency of household activities is monotonically increasing (MS = 1). Having a driving license increases the mobility of both heads and increases the number of household activities as indicated by the positive sign (MS = 1) for *DrivM* and *DrivF* variables.

In terms of the accessibility of locations, the number of employees within 3.1 km in the daily good sector (*Dist1*) turns out to be the most influential variable. With increasing number of employees the frequency of household activities increases, although not monotonically (MS = 0.34). In addition to that, urban density (*Urban*) has less influence to the household activities. The frequency of household activities is increasing non-monotonically (MS = 0.06) with the increasing level of urban density, meaning that lower urban density leads to lower frequency of household activities. Hence, household heads prefer to conduct household activity in a more urbanized area. The effect, however, is very small and far from monotonic.

In summary, almost all of the 61 condition variables used as input to the induction process recur in the decision tree. Only 9 variables do not affect predictions of activity participation, as indicated by a zero value of the chi-square-based impact measure.

6. CONCLUSIONS AND DISCUSSION

Incorporating household decision-making into activity-based models of transport demand has received increasing attention recently. In contributing to this emerging line of research, the present paper reported the development and test of a model of activity participation, based on household decisions, as part of the full-fledged activity-travel scheduling model ALBATROSS. Focusing on *allocated activities* and *joint discretionary activities*, a rule-based activity participation model was derived from activity-trip diary data. Given the large number of observations in the data, more than 300 condition-action rules (decision rules) were derived. The validity of the decision tree model is satisfactory in the sense that the derived rules are readily interpretable and the overall goodness-of-fit of the model is acceptable. The improvement in goodness-of-fit relative to a null model indicates that there is a moderately strong association between condition variables defined at the household, individual, activity and schedule level, and participation in household activities decisions (by type and frequency). Furthermore, the performance of the model on a validation dataset suggests that the derived decision rules are generalizable to unseen cases. These results suggest that the way of structuring the household decisions as we proposed in this study has merit.

Using impact tables, this study suggests that activity type is the most influential variable in this context, followed by day of the week and number of bring/get activities already included

in the schedule by the time of making the activity participation decision. The presence and age of children in the household and income are also relatively influential variables among the socio-demographic variables considered in the model. In terms of time-availability variables, the total duration of work activity across household heads appears to be the most influential variable. Accessibility variables only have modest impact. The accessibility of facilities in the daily goods sector appears to be the most important variable in this category.

Household-level activity-travel decisions are not limited to activity generation problems. Travel choices may involve household decision making, especially in cases where there are fewer cars than driving licenses in the household. The decision to travel jointly also requires a joint decision of the persons involved. In our future research, we intend to develop similar process and decision tree models related to these choice facets as well.

REFERENCES

Anggraini, R., Arentze, T.A., and Timmermans, H.J.P. (2007). Modeling Car Allocation Decisions in Automobile Deficient Households. **Proceedings of ETC 2007 Conference**, Noordwijk, The Netherlands.

Arentze, T.A., and Timmermans, H.J.P. (2000). **ALBATROSS: A Learning-based Transportation Oriented Simulation System.** EIRASS, Eindhoven University of Technology, The Netherlands.

Arentze, T.A., and Timmermans, H.J.P. (2003) Measuring Impacts of Condition Variables in Rule-Based Models of Space-Time Choice Behavior: Method and Empirical Illustration, **Geographical Analysis**, **35**, 24-45.

Arentze, T.A., and Timmermans, H.J.P. (2004). A Learning-Based Transportation Oriented Simulation System. **Transportation Research Part B**, **38,** 613-633.

Arentze, T.A., and Timmermans, H.J.P. (2005). **ALBATROSS 2.0: A Learning-based Transportation Oriented Simulation System.** EIRASS, Eindhoven University of Technology, The Netherlands.

Breiman L., Friedman, J.J., Olshen, R.A., and Stone, C.J. (1984). **Classification and Regression Trees.** Wadsworth, Belmont [CA].

Gliebe, J.P., and Koppelman, F.S. (2002). A Model of Joint Activity Participation Between Household Members. **Transportation**, **29,** 49-72.

Gliebe, J.P., and Koppelman, F.S. (2005). Modeling Household Activity-Travel Interactions as Parallel Constrained Choices. **Transportation**, **32,** 449-471.

Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. **Applied Statistics, 29,** 119-27.

Miller, E.J., and Roorda, M.J. (2003). A Prototype Model of Household Activity/Travel Scheduling. **Proceedings of Transportation Research Board (TRB) 2003**, Washington DC., USA.

Quinlan, J.R. (1993). **C4.5 Programs for Machine Learning**. San Mateo, Calif.: *Morgan Kaufmann Publishers.*

Scott, D., and Kanaroglou, P. (2002). An Activity-Episode Generation Model that Captures Interaction between Household Heads: Development and Empirical Analysis. **Transportation Research B**, **36B,** 875-896.

Srinivasan, S., and Bhat, C. (2004). Modeling the Generation and Allocation of Shopping Activities in a Household. **Proceedings of Transportation Research Board (TRB) 2004**, Washington, DC., USA.