EJTIR

Issue **8**(4)
December 2008
pp. 301-319
ISSN: 1567-7141
www.ejtir.tbm.tudelft.nl

Car Allocation between Household Heads in Car Deficient Households: A Decision Model

Renni Anggraini
Faculty of Architecture, Building, and Planning, Eindhoven University of Technology[1]

Theo A. Arentze
Faculty of Architecture, Building, and Planning, Eindhoven University of Technology[2]

Harry J.P. Timmermans
Faculty of Architecture, Building, and Planning, Eindhoven University of Technology[3]

Received January 2008; accepted July 2008

This paper considers car allocation choice behaviour in car-deficient households explicitly in the context of an activity-scheduling process, focusing on work activities. A decision tree induction method is applied to derive a decision tree for the car allocation decision in automobile deficient households using a large travel-and-activity diary data set recently collected in the Netherlands. The results show a satisfactory improvement in goodness of fit of the decision tree model compared to a null model. Overall, the probability of males getting the car for work is considerably higher than that of female in many condition settings. However, activity schedule, spatial and socio-economic variables appear to have an influence as well. An analysis of impacts of condition variables on car allocation decisions reveals that socio-economic variables have only a limited impact, whereas attributes of the transportation and land-use system have a relatively big impact. The propensity of men driving a car to the work place is higher than that of women. However, the relative accessibility of the work location by bike compared to car appears to have a relatively large influence on who gets the car for work. Household income and presence of children also appear to have significant effects.

Keywords: travel demand modelling; activity-based modelling; decision tree induction; within-household interactions; car allocation

---

1. Introduction

One of the major indirect factors contributing to increasing traffic congestion in urban areas and highways is the increase of household automobile ownership. The vast majority of households own at least one car and an increasing number of households own more than one car. It is of no

[1] P.O. Box 513, 5600 MB, Eindhoven, The Netherlands, T: +31402473145, F: +31402475882, E: r.anggraini@bwk.tue.nl
[2] P.O. Box 513, 5600 MB, Eindhoven, The Netherlands, T: +31402472283, F: +31402475882, E: t.a.arentze@bwk.tue.nl
[3] P.O. Box 513, 5600 MB, Eindhoven, The Netherlands, T: +31402473960, F: +31402475882, E: h.j.p.timmermans@bwk.tue.nl

surprise therefore that car ownership and vehicle fleet choice is one of the areas in transportation research that has received much attention. A complementary active area of research focuses on transport mode choice analysis and modeling to shed light on preferences of individuals in choosing one option among several modes available for the trips they make (Xie, et al., 2003; Miller et al., 2005).

Despite the substantial amount of research on car ownership in general, the specific question of who is getting the car for which activities in car deficient households has received much less attention. In this context, car deficient households are households where the number of drivers exceeds the number of cars. Consequently, we know relatively little about the factors that play a role in this decision and about the decision process by which household members arrive at a choice on who should use a car (Hunt and Petersen, 2004; Vovsha and Petersen, 2007). A model of binary car-allocation choice (to use car or not) made by the household members for each tour in an integrated framework of intra-household car-use preferences has been proposed and estimated by Petersen and Vovsha (2005). They clearly showed that car-allocation decisions are inter-related with mode choice, joint travel arrangements, and schedule adjustments.

Yet, the outcome of this decision does not only have a direct impact on transport mode choice, but also has potentially important ramifications for activity-travel schedules of individual household members. Action spaces allowed by different transport modes vary substantially and therefore the generation, location and timing of activities and the organization of trips into tours depends strongly on the transport mode. Critical questions in better understanding this decision process include: how do households make trade-offs between mobility needs of drivers and are there differences between households related to socio-economic and situational variables? Current travel demand models have paid little attention to address these car allocation decisions.

The decision which person will use the car is a complex decision in car deficient households in the sense that many factors may influence this decision. For example, gender roles may imply that males are more likzely to use the car than women are. However, it may also be that in case the male is going to work for a long period of time in a day, while the female has many errands to complete, the flexibility of scheduling and rescheduling activities made possible by the car, may lead the household to decide that the female will use the car. As argued by Bianco and Lawson (1996), women are more dependent on the car than men because of their traditional responsibilities related to childcare and household maintenance as well as their concern for safety. On the other hand, due to a good provision of public transport and more dense cities, in Western European countries, we often see that women who do not participate in the labor force tend to use public transport or use slow modes. Apart from socio-demographic variables, the relative accessibility of locations for activities by car will have an influence.

Surprisingly, car allocation decisions have also not received much interest in the activity-based (micro-simulation) modeling literature. To date, fully operational activity-based micro-simulation systems include ALBATROSS (Arentze and Timmermans, 2000; 2004; 2005), TASHA (Miller and Roorda, 2003), Florida's Activity Mobility Simulator (FAMOS) (Pendyala, 2004), based on the Activity-Mobility Simulator (AMOS) (Kitamura et al., 1996), and the Prism Constrained Activity-Travel Simulator (PCATS) (Kitamura and Fujii, 1998), and CEMDAP (Bhat et al., 2004), and some projects that have been implemented in the US (Bowman, 2008; Vovsha, 2008). One of the reasons for developing activity-based models was that typical response patterns to transport demand management involved household decisions. Such responses could not be captured by trip-based models, at least not explicitly, as they were founded on individual as opposed to household behavior. In general, only few of the existing operational activity-based models are based on household decisions, and this statement also applies to the car allocation decision.

In this paper, we examine this emerging issue. The study focuses on households which have fewer cars than drivers. Car allocation decisions are considered as an element of a more encompassing activity scheduling process. A large number of factors that potentially influence car allocation decisions in car deficient households are considered. These factors relate to variables of the activity schedule and space-time setting as well as individual and household characteristics. In this study, we use ALBATROSS as a framework to investigate the car allocation decisions as part of an activity scheduling process. ALBATROSS is an operational activity-based model developed for the Dutch Ministry of Transportation, Public Works and Water Management for travel demand analysis. More specifically, the paper will report the conceptualization of the problem and present empirical results of a car allocation model for two household heads. It is assumed that before the car is allocated, participation in activities both at the household and person level is known. If one car is available in the household and both household heads are drivers, then the decision which person is going to use the car involves a household-level decision. For instance, if the two persons undertake a work activity during the same time slot, a decision needs to be made who can use the car for the trip to work. Note that the outcome could also be that both will use another transport mode for the work commute.

The paper is structured as follows. First, the next section briefly explains the ALBATROSS scheduling process model that provides the framework for the car allocation model. The sections that follow describe the data used for the analysis and the proposed car allocation model. After this section, the results of empirical analyses will be considered focusing on some descriptive statistics and the empirical derivation of the model. The paper is concluded by drawing conclusions and discussing some possibilities for future research.

2. ALBATROSS Process Model

ALBATROSS stands for A Learning Based Transportation Oriented Simulation System. The model considers household and personal activities and travel performed on a particular day and generates a schedule for each household head. The model takes into account the presence of children as an independent variable, but their activities are not explicitly represented. Work activities are presumably *primary fixed activities*, whereas several household activities and work-related activities, such as bring/get person, business, and others are assumed as *secondary fixed activities*. Shopping, social and leisure activities are called *flexible activities*. It should be noted that fixed activities are also predicted.

ALBATROSS consists of four major components that together define a schedule for each household head for a certain day as displayed in Figure 1. It should be noted that this describes the computational process model underlying the system merely in main lines. The first component generates a work activity pattern consisting of one or two work episodes, if any, and the start time, duration and location. It also predicts the transport mode(s) used to travel to the work location(s). The second component determines the part of the schedule related to secondary fixed activities (bring/get person, business, and others). It determines which types of these activities are conducted that day and how many episodes and for each episode the start time, duration and location of each episode. Furthermore, it also determines whether particular trip-linkages are made with the work activity, if any.

The following component considers the scheduling of flexible activities. Almost similar to the previous component, it predicts activity types, number of episodes of each activity type and the start time, duration and location of each episode. The sequence of activities and possible trip-chaining links between activities are also determined in this stage. The latter decisions relate to all activities in the schedule, not just the flexible activities. Finally, the last component predicts the transport mode used for each tour (except for tours that include a work activity; for the latter tours the transport mode is known as the outcome of a higher-level decision).

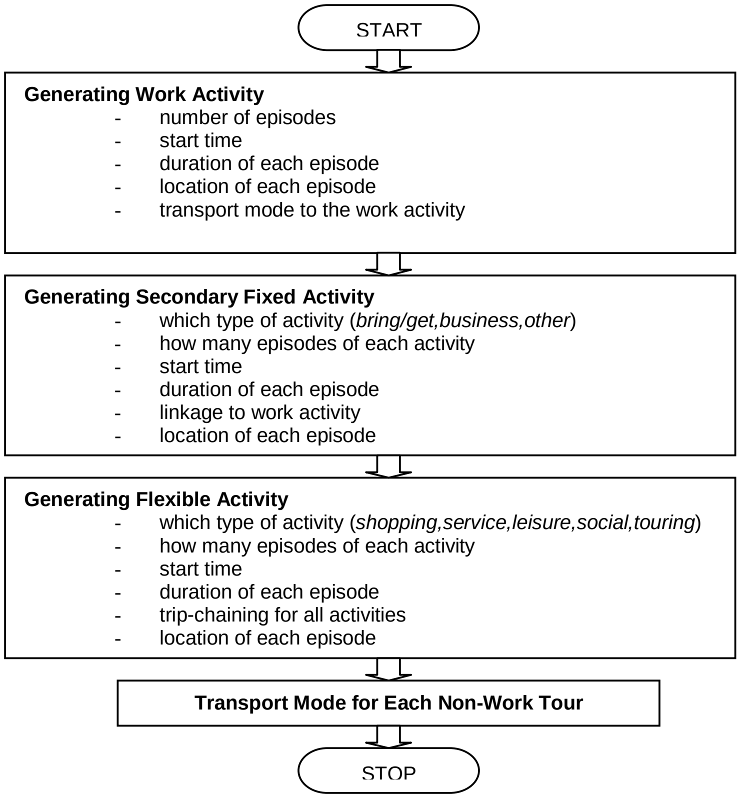

Figure 1. Schematic Representation of Main Steps of the ALBATROSS Process Model

The car allocation model developed in this study predicts who of the two household heads in car deficient households uses the car for a particular activity. As a case, we focus here on the work activity given that this activity usually is mandatory, conducted by one spouse individually (as opposed to jointly), tends to occupy a large part of the day and may serve as a second base location for other activities besides the home location. We emphasize, however, that car-allocation decisions are not confined to the work activity. In the last step of the Albatross scheduling process (Figure 1), the trips required for non-work activities and the way they are organized into tours are known. In that stage, a mode choice is made for each non-work tour (chain of trips including one or more activities). These choices are preceded by a car allocation decision as well. Although we focus here on the work activity, the same methodology developed here is used to model car-allocation decisions involved for non-work tours. A car-allocation decision restricts a subsequent mode choice: if the car is allocated to an activity or tour no further decision is needed and if the car is not allocated, then a choice is confined to other modes then the car. Note that car sharing is still open as an option if the car has not been allocated to an activity or a tour. In ALBATROSS, car sharing is represented as a car-passenger option. In other words, the car allocation decision has implications for the possibility of choosing the car-driver mode only, but leaves open the car-passenger mode.

Because ALBATROSS uses a sequential decision process, to generate a schedule for each household head, the information available for the car allocation model is limited. At the moment in the process when the car allocation model generates decisions, the schedules of the household heads regarding the work activity are known; the schedules regarding other activities then are still unknown. This does not mean, however, that the decisions cannot take requirements of other activities (which are scheduled in a later stage) into account. An outcome of the decision may well be that the car is not used for a work activity considering the household's needs for other activities. For example, presence of children is a condition variable the system can use to anticipate a possible escorting activity for which a car is needed and, hence, may inform the system not to allocate the car to a work activity (of one of the two partners). Due to the complexity of the scheduling problem it is inevitable that the decisions are made in a particular sequence.

3. Data

The data used in this study originates from the so-called MON survey (*Mobiliteit Onderzoek Nederlands* – Mobility Research Netherlands) held in 2004. The MON survey is conducted on a regular basis to obtain travel and activity information of residents in the Netherlands, and although it primarily uses a trip-diary it includes detailed data on activities (at destinations) as well. More specifically, it is a one-day travel diary of a sample of households that contains information about each household member. In addition, individual and household socio-demographics such as age, household composition, education level, income level, vehicle availability, residential location, and information about all trips made within 24 hours as well as out-of-home activities at destinations of trips are collected. For each trip, respondents are asked to report information about several attributes including type and duration of the activity at the destination, departure time and arrival time, trip purpose, transport mode, and origin and destination location. Furthermore, trip-chains can be identified. All in all, this information provides a suitable source to analyze activity-travel behaviour of Dutch residents because activity and travel information are both revealed. In this data collection, 29221 households filled out a one-day travel/activity diary and 28600 of these households fit the criteria for being considered here (forms of group housing, such as for example student housing, are excluded). The data were transformed to an activity-diary data format for the present estimation purpose.

4. Car Allocation Model Specification

As said, the car allocation model focuses on car deficient households (i.e., more drivers than cars present) and a joint decision between the two heads (mostly, a female and male). The total sample extracted from the MON data includes 28600 households. Given the purpose of this study, only the following households and days are relevant: (1) there are two heads in the household; (2) there is one car in the household; (3) both heads are drivers and (4) at least one of the heads has a work activity on the day considered. As it appears, 3523 households (and days) fit these criteria.

The car allocation decision model is schematically shown in Figure 2. A car-allocation decision is needed not only if the two heads in a household both have a work activity. Also, if only one of them performs a work activity, it is still necessary to identify whether the worker uses the car or not. Furthermore, the model includes the option that none of the household heads uses the car, but some other means of transport. Hence, the decision options are *male*, *female*, or *none*.

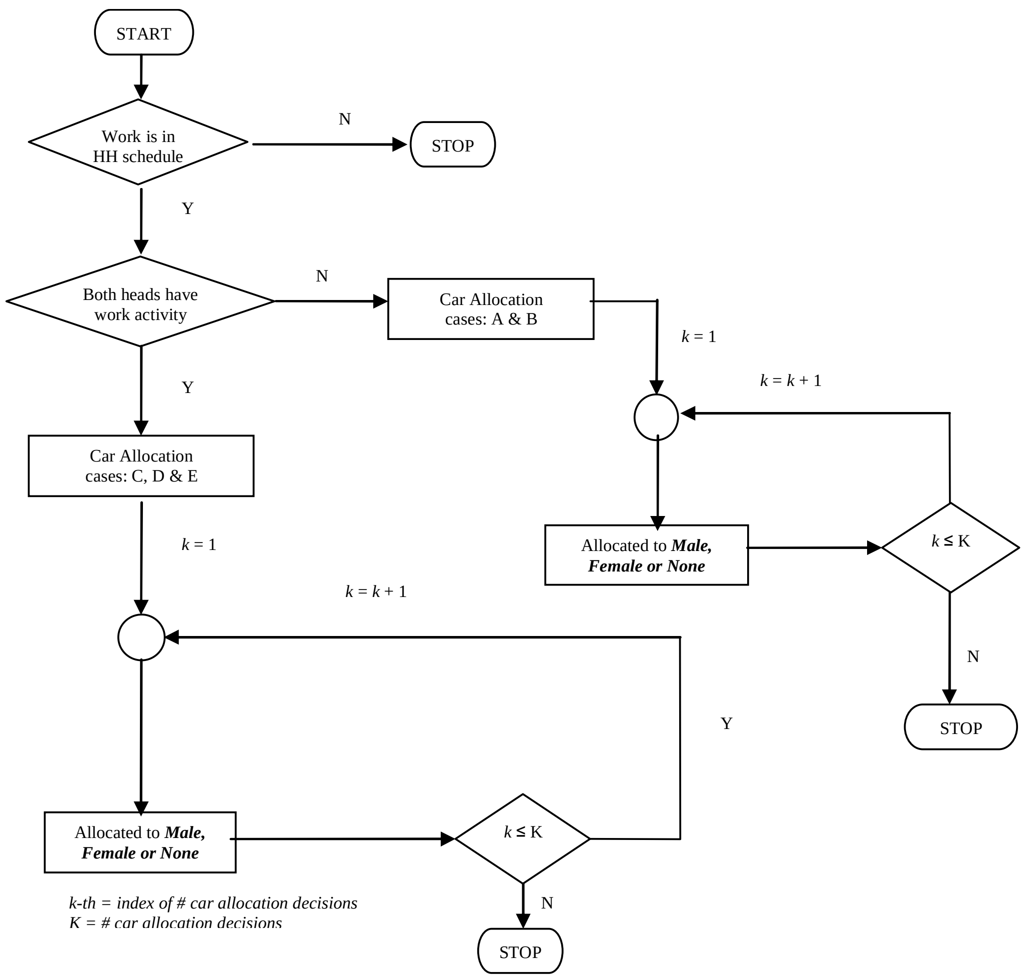

Figure 2. The process of car allocation model

In order to determine how many times such car allocation decisions should be made in a household on the day considered, we need to identify the number of work episodes performed by male and female heads. Table 1 shows the car-allocation cases that can be distinguished in that respect. Case A represents the situation that only one of the heads conducts one work episode, leading to only one car allocation decision in the household. In this case, one head may use the car, but also there is an option that he/she may not use the car. In Case B, two work episodes are included for only one of the household heads (for example, he/she returns home for lunch). This situation thus involves two car allocation decisions when the break is long enough to allow for traveling back home and back to work again.

In case C, both heads have one work episode, implying that one or two car allocation decisions have to be made by the two persons. One car allocation decision is to be made if the work episodes of the two heads overlap in time (taking travel times into account). On the other hand, when there is no overlap in time, 2 car allocation decisions have to be made.

The same principle of overlapping episodes also applies to Case D and Case E, leading to maximally 3 and 4 car allocation decisions respectively. For example, in Case D, when the male worker has 2 episodes and female worker has 1 episode, there are 1, 2, or 3 car allocation decisions involved. In case both the first and second work episode of the male are overlapped with the work episode of the female, then there is only 1 car allocation decision required. If the first episode of the male worker and the episode of the female worker are overlapped while the second episode of the male worker is not overlapped with the episode of the female worker, this would imply 2 car allocation decisions. Furthermore, if none of the two work episodes of the male are overlapped with the female's one, then 3 car allocation decisions are needed. The similar reasoning applies to Case E. In the stage of the activity-scheduling processes where the work-related car allocation decisions are made, other activities have not yet been scheduled. Therefore, other activities that, in the end, are possibly attached to the work activity are not considered in this model.

Table 1. Defining Car Allocation Decisions in Households

No.	Number of male's work episodes	Number of female's work episodes	Cases	Number of Cases	Number of car allocation decisions
1	0	1	A	520	1
2	1	0		1437	
3	0	2	B	132	2
4	2	0		520	
5	1	1	C	1047	1 or 2
6	1	2	D	144	1, 2, or 3
7	2	1		228	
8	2	2	E	68	1, 2, 3, or 4
Total Sample				4096	

In determining whether or not there is an overlap in time, the travel time has to be taken into account as well. The travel time by car mode (across the road network) is relevant here. First, the timing and duration of work episodes of the household heads are derived and then the type of overlap is determined. Note that, travel time by car is used because that is relevant for car allocation decisions. Further details are provided in Section 5.

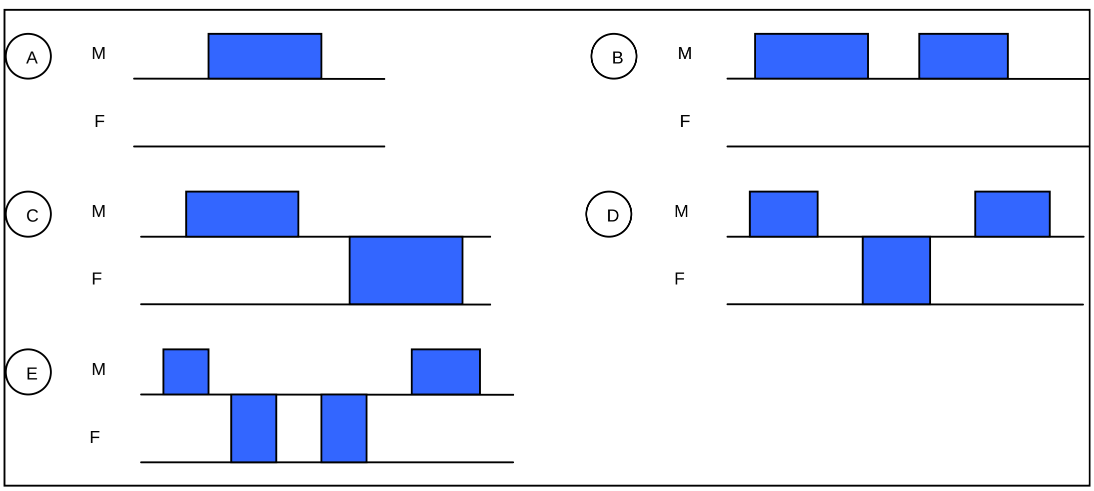

Figure 3. Examples of Distinguished Cases

5. Empirical Analyses

In this section we describe the results of deriving a decision tree model for car allocation choice. Before discussing these results, we will first consider some descriptive analyses carried out to get a better understanding of the characteristics of the sample after selecting car deficient households. Next, we briefly discuss CHAID, which is the decision tree induction method we use to derive decision rules from the MON data. To facilitate interpretation of decision tree results, we use a post-processing technique called impact tables. The impact table technique will be briefly discussed in the section that follows. Finally, in the last section, we discuss the results of the induction of the car allocation decision tree model and the corresponding impact table.

5.1 *Descriptive Analyses*

As discussed above, only a subset of households is relevant for the car allocation model, because the problem concerns car allocation to work activities in car deficient households. A total of 3,523 households were selected from the MON data, yielding 4,096 relevant cases of car allocation decisions. To describe the final sample, some further descriptive analyses were conducted.

Table 2. Distribution of households across household composition and SEC (%)

Household Composition	SEC				Total
	Low	Mid-Low	Mid-High	High	
Double, One Worker	3.2	13.3	12.3	11.6	40.3
Double, Two Worker	0.8	11.7	19.7	23.9	56.1
Double, No Worker	0.9	1.2	1.0	0.5	3.6
Total Sample (4096)	4.9	26.2	32.9	35.9	100

Table 2 displays the frequency distribution of households across household composition and socio-economic class combinations after selection. High-level income households are in the majority (35.9%) and consist most frequently of double-two-worker households. Double means two adults (male-female adult) household. This is followed by mid-high income (32.9%), mid-low income (26.2%) and low income households (4.9%).

Table 3. Distribution of household heads across household composition and work status of household heads by gender (%)

Household Composition	Work Status, Male			Total	Work Status, Female			Total
	Non-worker	Part-time	Full-time		Non-worker	Part-time	Full-time	
Double, One Worker	10.8	1.9	27.6	40.3	29.5	3.2	7.6	40.3
Double, Two Worker	0	8.4	47.7	56.1	0	31.2	24.9	56.1
Double, No Worker	3.6	0	0	3.6	3.6	0	0	3.6
Total Sample (4096)	14.4	10.3	75.3	100	33.1	34.4	32.5	100

The distribution of household heads across household composition and work status by gender is presented in Table 3. Over 75% of males are full-time worker. Females are approximately equally distributed across the work-status categories (33.1%, 34.4% and 32.5% for no, part-time and full-time worker respectively). This suggests that gender still plays an important role in work commitments and task allocation.

Table 4. Work duration statistics by work status and gender

Working Status	Male			Female		
	Average duration of work activity (min)	Standard Deviation (min)	Freq.	Average duration of work activity (min)	Standard Deviation (min)	Freq.
Part-time	293.78	245.20	422	207.09	219.70	1412
Full-time	373.63	235.71	3085	257.88	245.47	1331
Total	364.02	238.26	3507	231.73	233.90	2743

Table 5. Work duration statistics by day of the week and gender

	Male			Female		
Day of the Week	**Average duration of work activity (min)**	**Standard Deviation (min)**	**Freq.**	**Average duration of work activity (min)**	**Standard Deviation (min)**	**Freq.**
Monday	373.84	237.49	708	239.32	235.01	576
Tuesday	381.85	228.92	663	249.10	238.62	513
Wednesday	384.33	233.22	623	230.87	229.94	485
Thursday	367.19	235.38	683	226.62	231.05	541
Friday	356.01	241.31	635	223.60	236.36	461
Saturday	237.96	242.12	135	211.68	228.38	114
Sunday	172.83	230.95	60	155.15	219.24	53
Total	364.02	238.26	3507	231.73	233.90	2743

Table 4 shows the distribution of duration across work activities for male and female heads by work status. Note that persons may conduct more than one work activity a day; the figures presented refer to durations on a per-activity basis (as opposed to a per-episode basis). As we can see, males on average work approximately one and a half times as long hours than females per work activity. Furthermore, in each work status group, the average duration of males' work activity is higher than that of female. The frequency of work activities conducted by full-time male worker is leading among its class, as a result of the fact that 75% of the males work full-time. This also suggests that gender still plays a significant role in household task allocation.

Finally, Table 5 describes the household heads work activity duration split up by day of the week. As can be seen, on average, working hours of males is similar from Monday through Friday, about 6 hours. Meanwhile, working hours of females is on average about 3-4 hours during working days. Again, this result shows that on average males work longer hours than females per work activity.

5.2 Decision Tree Induction

We applied a CHAID-based tree induction method to identify the decision rules that can describe car allocation choice behavior. CHAID (Kass, 1980) 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 large datasets and probabilistic action assignment. Other commonly used decision tree induction systems are C4.5 (Quinlan, 1993) and CART (Breiman et al, 1984). All these methods use a recursive process of splitting the sample based on condition variables into partitions that are as homogeneous as possible regarding the action variable (i.e., the car allocation choice in this case). CHAID relies on the *Chi-square* test to determine the best next split at each step. To determine the best split at any node, it merges any allowable pair of categories of the condition variable if there is no statistically significant difference within the pair with respect to the action variable. This is done for each candidate condition variable. The split having the highest significance value (after Bonferroni correction for multiple tests) across condition variables is selected and implemented. The process is repeated until no more significant splits are found also taking into account a pre-defined minimum number of cases requirement at leave and parent nodes. This process of extracting rules is the same as the one used in the ALBATROSS model. In order to develop the decision tree, 75% of the cases were used for training and the remaining cases were used for validation. Generally, in deriving ALBATROSS decision models, attributes of the household, person, space-time setting and schedule as far as known in the stage considered of the assumed decision process are used as condition variables. Observations of condition variables and action variables (car allocation choice) in each case are extracted from the diary data.

The CHAID decision tree induction method allows one to define the threshold for splitting in terms of a significance level for the Chi-square (χ^2) measure and a minimum number of cases at leaf nodes. Alpha was set to 5% and the minimum number of cases to 50. The number of leaf nodes gives an indication of the complexity of the resulting tree. As a measure of prediction accuracy, the expected hit ratio is used. The expected hit-ratio represents the expected proportion of cases predicted correctly when a probabilistic action assignment rule is used. It is calculated as: $\frac{1}{N}\sum_{ki}\frac{(f_{ki})^2}{N_k}$ where f_{ki} is the frequency of the *i*th action at the *k*th 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.

5.3 Deriving Impact Tables

Decision trees derived from data may become very large and complex and, consequently, difficult to interpret. This holds true particularly for the present application where the number of choice observations is very large. Arentze and Timmermans (2003) developed a method to derive elasticity information from rule-based models to facilitate interpretation, which we will use here to describe the results of tree induction. The principle of the proposed method is straightforward. After having derived a rule-based model from training data, the model is used to predict for each condition variable a frequency cross 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 for a given condition variable 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 on the condition variable. The frequency distribution across actions of the action variable predicted under that setting is recorded. Repeating this process for each level of the condition variable yields a frequency cross table of the condition variable against the action variable. The impact of the condition variable is then measured as the Chi-square for this frequency table. Formally:

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

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 on each level of the action variable, as follows:

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

where again D is a chi-square measure and $\mathbf{F}_{si}$ is the vector of predicted frequencies of the *i*-the 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) defined as:

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

where f_{ij} is the predicted frequency of action *i* under the *j*-th level of condition variable *s* and *J* is the number of levels. 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. We emphasize that the monotonicity measure is meaningful only for variables that are naturally enumerated; it is not informative for variables that are purely nominal.

5.4 Condition and Action Variables

Table 6 portrays the condition variables that were used as input to the tree-induction algorithm. The condition variables concern household level (including accessibilities), individual level, and activity level variables (note that in this stage of the scheduling process only work activities are known). Continuous condition variables, such as travel time, duration, and parking price, 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.

The presence of young children in a household is taken as a condition variable as well as other household and individual attributes, such as work status, socio-economic class (in Euro), urban density (number of home addresses per area unit in the zone where the household lives classified on a 5-point scale) and the day of the week (no. 1-8 in Table 10). The number of work activity episodes that is performed by male or female is 0, 1 or 2 episodes (no. 9-11). Accessibility variables, such as travel time, train and bus connections, parking price and free-paid parking place ratio were also used (no. 12-29, except no. 18-19). They are calculated based on national datasets of the transport system (car, bike/walk and public transport). They all relate to the trip to the work location. If there is no work activity conducted by the person on a particular day, the variables are set to zero for that person. If a work activity is conducted in the same postcode area as where the person lives, then travel time is set to zero too. Travel time by car is included as a direct measure of accessibility. Travel time ratios between modes are used as indicators of relative accessibility by particular modes. Ratios are used to allow the algorithm to identify impacts of modes more easily.

Work duration is an attribute of the activity for which a car allocation decision is made (no.18-19). The definition of this variable takes the overlap pattern into account. To explain this, consider for example, a case where the male has a work activity of 9 hours and the female has two work episodes of 4 hours each with a one hour break in between. In this case, there are two allocation decisions if the overlap concerns only one of the female's work episodes. For both decisions the considered work duration for the male is 8 hours and for the female 4 hours. On the other hand, if the male's work activity overlaps with both female's work episodes, just one allocation decision needs to be made. For that decision the considered work duration for the male is 9 hours, as before, but for the female it becomes 8 hours.

Note that some of the variables relate to the schedule level (a day of the household), whereas others are defined at the level of the activity which involves a car-allocation decision (i.e., a work activity of one or both of the heads). The variables that correspond to the schedule level are number of work episodes of male and female respectively and number of car-allocation-decision cases in a household (no.31). The number of car allocation cases occurring in a household can be 1, 2, 3, or 4 cases (see Table 1 and Figure 3). The variables at activity level are the following. For each car allocation case, the timing of work activities of both persons have to be considered to determine whether or not there is an overlap in time (no. 30). Obviously, if only one person performs a work activity in a particular time period, then there is no overlap in time, and otherwise there might be. Variable no.32 indicates the type of overlap in terms of all possible combinations of number of work activities (none, one or two) by male and female. In Cases (1) and (2) only the male or female has a work activity in a particular time period. In contrast, in cases (3) to (6) there is a time overlap between their work activities.

Table 6. Condition Variables for Car Allocation Model

No	Variable	Classification	Acronym
1	Urban Density	0=most densely , 4= least densely	Urban
2	Household Composition	2=DONEWORK, 3=DTWOWORK, 4=DNOWORK	Comp
3	Presence of the youngest children	0=no children, 1=<6, 2=6-11, 3=12-17 yrs	Child
4	Day of the week	0=Monday to 6=Sunday	Day
5	Age of person	0=<35, 1=35-<55, 2= 55-<65, 3= 65-<75, 4= 75+ yrs	Age
6	Socio-economic class	0=0-16,250 (low), 1=16,251-23,750 (low-mid), 2=23,751-38,750 (mid-high), 3=38,750+ (high)	SEC
7	Working status – M	0= non-worker, 1= part-time, 2= full-time	WstatM
8	Working status – F	0= non-worker, 1= part-time, 2= full-time	WstatF
9	Number of work episodes – M	0, 1, 2	NworkM
10	Number of work episodes – F	0, 1, 2	NworkF
11	Number of work episodes in household	1,2,3,4	NworkHH
12	Travel time by car – M (in minute)	0=0; 1=≤8; 2=9-14; 3=15-22; 4=>22	TTcM
13	Travel time by car – F (in minute)	0=0; 1=≤6; 2=7-11; 3=12-18; 4=>18	TTcF
14	Travel time ratio between PT and car – M	0=0; 1=≤1.00; 2=1.01-1.98; 3=1.99-4.11; 4=>4.11	TTptM
15	Travel time ratio between PT and car – F	0=0; 1=≤1.00; 2=1.01-2.14; 3=2.15-4.49; 4=>4.49	TTptF
16	Travel time ratio between car and bike – M	0=0; 1=≤0.25; 2=0.26-0.37; 3=0.38-0.81; 4=>0.81	TTcbM
17	Travel time ratio between car and bike – F	0=0; 1=≤0.30; 2=0.31-0.42; 3=0.43-1.00; 4=>1.00	TTcbF
18	Duration of work episode – M (in minute)	0=0; 1=≤275; 2=276-520; 3=521-565; 4=>565	DurM
19	Duration of work episode – F (in minute)	0=0; 1=≤240; 2=241-380; 3=381-540; 4=>540	DurF
20	Train accessibility – M	0= no, 1= yes	TrAcM
21	Train accessibility – F	0= no, 1= yes	TrAcF
22	Bus accessibility – M	0= no, 1= yes	BusAcM
23	Bus accessibility – F	0= no, 1= yes	BusAcF
24	Work conducted by male	0= no, 1= yes	Mwork
25	Work conducted by female	0= no, 1= yes	Fwork
26	Ratio # paid parking places to total #	0=0; 1=≤0.09; 2=0.10-0.15; 3=0.16-0.28;	RParkM
27	parking places – M	4=>0.28	
	Ratio # paid parking places to total #	0=0; 1=≤0.07; 2=0.08-0.14; 3=0.15-0.24;	RParkF
28	parking places – F	4=>0.24	
	Average price of parking – M	0=0; 1=≤9; 2=10-25; 3=26-66; 4=>66	PParkM
29	Average price of parking – F	0=0; 1=≤8; 2=9-22; 3=23-44; 4=>44	PParkF
30	Overlapping between two persons' episodes	0=no, 1=yes	overlap
31	Number of car allocation cases in household	1=1, 2=2, 3=3, 4=4	NcarAl
32	Type of case for allocating the car	1=Male only, 2=Female only, 3=M&F (each 1 ep), 4=M (2 ep) & F (1 ep), 5=M (1 ep) & F (2 ep), 6=M&F (each 2 ep)	cases

Note: M = Male; F = Female; PT = Public Transport; ep = episode

As a result, a total of 32 condition variables were defined. The action variable, as the output of the car allocation model, involves assigning the car to *male*, *female*, or *none* of the two household heads.

5.5 Results

For deriving the car allocation model for work activities, a total of 4,096 observations could be derived from the data set. 75% of these cases (3,114) were used for training and the remaining cases were used for validation. Of 4,096 cases, the probabilities of the car being allocated to male and female are 37.28% and 17.77% respectively. In the remaining cases, 44.95%, the male and female heads use other modes to the work place.

Table 7 shows the frequency distribution of allocation decision outcomes over household types in terms of work status of the heads. In households where male is a non-worker, in about 50% of the cases household heads choose some other mode to travel to the work place. In households where male is a part-time worker and the female is a non-worker, the car is allocated to the male in about 43.59% of the cases. However, if both male and female are part-time workers, about 50.75% of the cases they use some other mode than car. In households where male is a full-time worker and female is a non-worker, the car is allocated to the male in 43.67 % of the cases. In sum, the figures show that the male gets the car more often than the female even in two-worker households.

Table 7. Frequency Distribution of Work Status across the Action Variables

No	Work status		% of getting the car			Total
	Male	Female	Male	Female	None	
1		Non-worker	32.88	15.75	51.37	146
2	Non-worker	Part-time	9.23	40.00	50.77	130
3		Full-time	36.10	13.74	50.16	313
4		Non-worker	43.59	16.67	39.74	78
5	Part-time	Part-time	38.81	10.45	50.75	67
6		Full-time	25.99	25.99	48.01	277
7		Non-worker	43.67	7.26	49.07	1129
8	Full-time	Part-time	35.88	20.91	43.21	1215
9		Full-time	36.98	27.13	35.90	741
	Total		1508	747	1841	4096

Given a minimum group size of n=50 cases at parent nodes and a 5% alpha level, the tree generated by CHAID consists of 29 leaf nodes (decision rules). The hit ratio (based on a probabilistic assignment rule and the training set) of the model, compared to a null-model (a root-only decision tree) indicates a significant improvement achieved by the tree: the hit-ratio of the null-model of 0.374 is significantly increased to 0.540.

Figure 4 shows the resulting car allocation tree model graphically by branch from the root note. The first split is implemented on travel time ratio between car and bike for the work activity performed by female ($TT_{cb}F$). Recall that the variable is set to zero if the person has no work activity or the work activity takes place in the same post code area as the home location. This results in five branches from the root.

Branch #1 represents the condition where the female has no work activity or a zero travel time to the work place. Within this node a split is implemented on travel time by car for the male work activity (TT_cM), and so on. The probability distribution across *male, female* and *none* options is shown in italic font at each leaf node. Each path from the root to a leaf node represents a decision rule. For example, the path printed in bold (Figure 4) represents the rule:

<u>IF</u>: $TT_{cb}F = 0 \wedge TT_cM = 1 \wedge TT_{cb}M = 0\text{-}2 \wedge DurM = 3\text{-}4$
<u>THEN</u>: Male = 35.1%, Female = 0%, and None = 64.9%

This rule denotes that IF female either does not have a work activity or the work and home location are in the same zone AND travel time (by car) of male is 8 minutes or less AND the travel time ratio between car and bike for the male is less than 0.37 (traveling by car is at most 2.7 times as fast as the bike) AND male's work duration is at least 521 minutes (8.68 hours), THEN the probability that the male gets the car is 35.1%. Thus, the propensity of not using the car to work by male is as high as 64.9% under these circumstances (where the male's work location is relatively well accessible by bike).

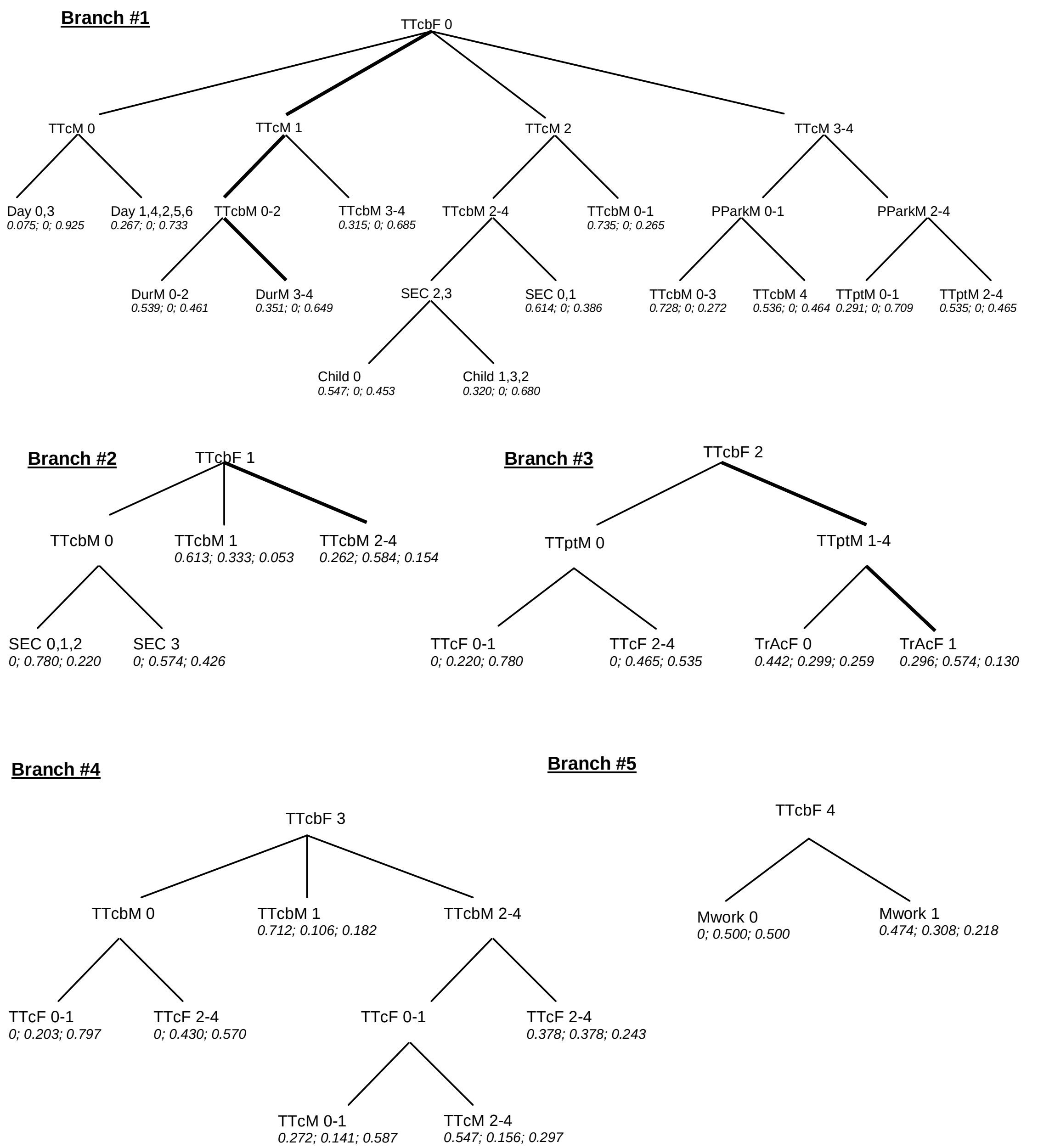

Figure 4. Car Allocation Tree Model with 5 Major Branches

As another example, in branch #2, the rule printed in bold indicates that IF travel time ratio between car and bike for female is less than 0.30 (relatively good accessibility by car) AND the travel time ratio between car and bike for male is at least 0.26 (relatively good accessibility by bike), THEN the probability of female getting the car (26.2%) is yet lower than that of male (58.4%). This rule indicates that even if the male's work place is well accessible by bike, the propensity of the male to use the car is considerably higher than that of female. Furthermore, in branch #3, as another example, the rule printed in bold indicates that IF the female and male both have a work activity and the travel time ratio between car and bike for the female is in between 0.31 and 0.42 (the car is between 3.2 and 2.4 times faster than bike) AND travel time ratio between public transport and car of male is greater than zero AND there is a train connection between home and the female's work location, THEN the female's probability of getting the car is substantially higher than the male's, namely 57.4% and 29.8% respectively.

The results of a performance analysis are shown in Table 8 in the form of a confusion matrix for the training and validation set. A confusion matrix describes the model performance in terms of a distribution of predicted choices for each observed choice category in the data set. The confusion matrix shown is based on probabilistic model predictions. The diagonal will have high numbers in case of good prediction. Off-diagonal elements of the matrix indicate the probabilities of predicting wrong actions for each observed choice category.

Table 8. Confusion matrix for the training and validation sets

Observed	Training set (N=3114)				Validation set (N=982)			
	Predicted				Predicted			
	Male'	Female'	None	Total'	Male'	Female'	None	Total'
Male	0.543	0.092	0.365	0.377	0.524	0.099	0.378	0.360
Female	0.197	0.471	0.332	0.176	0.166	0.478	0.356	0.184
None	0.307	0.130	0.563	0.448	0.321	0.113	0.566	0.455
Total	0.377	0.176	0.448	0.540	0.365	0.175	0.459	0.534

Table 8 shows that the model achieves a substantial improvement compared to a null-model as diagonal cells have higher percentages. For example, as it appears in the training set, in 37.7% of the cases we observe males using the car for the work activity. In 54.3% of these cases the model predicts car allocation correctly, while for the remaining cases the model predicts incorrectly that the female will use the car (9.2%) and none of the heads use the car (36.5%). Note also that, due to the probabilistic assignment rule used, the predicted distribution exactly matches the observed distribution overall cases. In that sense the predictions are bias free. Comparing the diagonals of the training and validation set suggests a small decrease in accuracy. As the bottom-right cell shows, the overall accuracy on the validation set is slightly decreased from 0.540 to 0.534. We consider the small decrease in accuracy as acceptable.

To evaluate the quantitative impacts of each condition variable on the action variable, Table 9 displays the impact table for the car allocation model. The condition variables are listed in order of decreasing impact on the action variable overall (the IS column). Note that ISmale, ISfemale, and ISnone show the size of the impact for each action separately.

Table 9. Impact Table of Condition Variables of Car Allocation Model

No.	Variables	IS	IS_{male}	IS_{female}	IS_{none}	MS_{male}	MS_{female}	MS_{none}
1	TTcbF	3719.77	120.71	2376.06	1223.00	-0.16	0.33	-0.38
2	TTcM	948.75	519.71	0.01	429.02	1.00	-1.00	-1.00
3	TTcbM	446.41	257.61	87.46	101.34	0.09	-0.20	-0.03
4	TTcF	58.58	0.14	43.07	15.37	-1.00	1.00	-1.00
5	PParkM	50.8	28.82	0.00	21.99	-1.00	-	1.00
6	TTptM	45.57	25.64	0.20	19.74	1.00	-1.00	-1.00
7	SEC	8.2	2.15	2.02	4.02	-1.00	-1.00	1.00
8	Day	5.69	3.10	0.00	2.59	0.33	-	-0.33
9	DurM	5.11	2.79	0.00	2.32	-1.00	-	1.00
10	TrAcF	4.66	0.54	3.79	0.33	-1.00	1.00	-1.00
11	Mwork	3.41	2.02	0.92	0.47	1.00	-1.00	-1.00
12	Child	2.42	1.33	0.00	1.09	-1.00	-	1.00

When we look at the differential impacts of types of condition variable, we see that socio-economic variables have only a limited impact, whereas attributes of the transportation system have a relatively big impact. Especially, travel time ratios and parking tariffs for the work location emerge with substantial impacts. The variable that gives by far the biggest impact is the travel time ratio between car and bike for female (TTcbF). The monotonicity measure (MSfemale = 0.33) clarifies that with increasing ratio of this variable, the probability that the female gets the car increases. At first sight, this seems implausible as the ratio indicates the relative accessibility by bike. However, note that a value of zero of this ratio means that the female has no work activity or a work activity in the home postcode area. Hence, an increase of the ratio from a zero value means a change in condition from no travel to positive travel for the female and, hence, an increase in the probability of getting the car. The fact that the impact is non monotonous indicates that the probability does not increase in the higher range, i.e. where an increase indicates an improvement of relative accessibility by bike. Logically, zero value should be taken out from the classification level. However, the software does not feasible to allow it. Then, this may be the shortage of this approach.

The *monotonicity* measure for the variable that gives the second biggest impact, *TTcM,* indicates that as travel time (by car) of male goes up, the frequency of allocating the car to the male increases monotonically (MS_{male} = 1), as expected. In sum, travel time and parking price variables have a big influence on car allocation decisions between the two household heads in a car deficient household, as indicated by the results of the first six variables.

In terms of socio-economic variables, we find that the most influential variable is socio-economic class (*SEC*). Interestingly, the probability of getting the car decreases monotonically for both male and female (MS_{male} dan MS_{female} = -1.00) as income rises. This result is somewhat counter-intuitive, given that car possession tends to be higher among high income groups. It should be noted, however, that since we consider car-deficient households we have corrected for number of cars available in the household (we consider only double-adult households having one car). Within this group, third variables such as education level and availability of public transport at the work place may exert an influence. Income is correlated with education level and possibly urban density at the location of employment (larger cities) and the latter variables are correlated with use of public transport. As a consequence, increasing income may lead to decreasing car allocation to work activities. The probability of male getting the car increases when the male has a work activity on the day concerned (*Mwork*). The presence of young children in the household (*Child*) is the last socio-economic variable that has an impact on car allocation decisions. Again interestingly, the tendency of not using the car by male increases monotonically (MS_{none} = 1.00) when the value of this variable increases, i.e. going from no children to presence of children with increasing age. Since there is at the same time no influence on the probability that the car is allocated to the female, it indicates that the car stays at home more often (possibly, for non-work activities of the female).

As for the situational variables, day of the week (*Day*) is the most influential variable. There is a non-monotonous tendency (MS_{male} = 0.33) of increasing probability of allocating the car to the male as the week proceeds from Monday to Sunday (the lowest value on this variable is Monday). Day of the week has no influence on the probability of allocating the car to the female. Another variable that has no influence on the probability of allocating the car to the female is work duration of the male (*DurM*). The probability of the male getting the car decreases monotonically as his work duration goes up (MS_{male} = -1.00). The presence of a train connection between home and the female’s work location (*TrAcF*) increases the probability of the female getting the car and decreases the probability that the male gets the car. Probably, the existence of a train connection acts as a proxy for distance and urban density: a train connection generally exists only between locations with relatively high density and far enough apart.

6. Summary and Conclusions

This paper considered car allocation choice behavior in car-deficient households explicitly in the context of an activity-scheduling process. Focusing on work activities, a car allocation model based on rules derived from a large travel diary data set using a CHAID-based induction algorithm was presented. The face-validity of the decision tree model is good in the sense that the derived rules and impacts of condition variables are readily interpretable. The overall goodness-of-fit of the model is satisfactory. Although the performance on a validation set decreased slightly, the set of decision rules seems stable across training and validation set to a satisfactory extent.

The propensity of men driving a car to the work place is higher than that of women in car deficient households, particularly, when women have no work activity or women's work place is in the same zone as the home location. This finding is consistent with a common notion that women use a slow or public transport mode more often to travel to activity locations. Similar to that, women tend to use the car when men have no work activities or men's travel time to work place is zero. When the female's work location is relatively well accessible by car, women are prevalent in getting the car.

In terms of decision rules results, in 43.1% of the rules men get the highest probability to use the car while in only 20.7% of the rules women have the highest probability to use the car. In the remaining of the rules (36.2%) none of the heads using the car gets the highest probability. Note that, the percentage rules are unweighted.

As the impact table analysis showed, travel time variables and, in particular, the relative accessibility of the work place by car compared to bike by far plays the most important role in car-allocation decisions in two-driver, single-car households. Work duration, day of the week and the existence of a train connection between home and work location also has an impact on the decisions. Although socio-economic variables appear to have only small effects on the decisions, presence of young children and household income has an influence too.

As we showed, car allocation decisions can be modeled as an element of a more encompassing activity scheduling process. ALBATROSS proved to be a suitable framework for this. This focus of our approach meant at the same time that only a limited set of explanatory variables at the level of the individual and household was taken into account. From an analytical perspective, it is interesting to extend the set of explanatory variables and investigate what the effects are of additional attributes such as job characteristics and car characteristics on these decisions. Furthermore, given that attributes of transportation systems appear to be significant, it is worth while to include even more detailed descriptors of the transportation system, e.g. public transport services and parking facilities. Finally, the present study focused on car allocation decisions in relation to the work activity. Clearly, car allocation decisions may also occur at the level of non-work activities in a scheduling process. The same approach as developed in this study can be applied for that purpose.

Acknowledgement

The authors are grateful to two anonymous reviewers for their constructive comments and suggestions that helped improve the paper. The authors are responsible for all errors and omissions.

References

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 the Goodness-of-Fit of Decision-Tree Models of Discrete and Continuous Activity-Travel Choice: Methods and Empirical Illustrations. *Journal of Geographical Systems*, vol. 5, pp. 185-206.

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*, vol. 35, pp. 24-45.

Arentze, T.A. and Timmermans, H.J.P. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B,* vol. 38, pp. 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.

Bhat, C.R., Guo, J.Y., Srinivasan, S. and Sivakumar, A. (2004). A Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. *Transportation Research Record*, no. 1894, Travel Behavior and Values 2004, Transportation Research Board, National Research Council, Washington, D.C.

Bianco, M. and Lawson, C. (1996). Trip-Chaining, Childcare, and Personal Safety: Critical Issues in Women's Travel Behavior. Women's Travel Issues, Proceedings from *Second National Conference*, pp. 123-143.

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

Bowman, J.L. (2008). From Theory to Practice: What Can We Learn from Our U.S Experience? Paper presented at *Transportation Research Board 2008,* Washington D.C, USA.

Golledge, R.G., Kwan, M.-P. and Garling, T. (1994). Computational-Process Modeling of Household Travel Decisions Using a Geographical Information System. *Working Paper UCTC* No.218, UCLA.

Hunt, K.L. and Petersen, E. (2004). The Role of Gender, Work Status and Income in Auto Allocation Decisions. Paper presented at the *83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., USA.

Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, vol. 29, pp. 119-127.

Kitamura, R., Pas, E.I., Lula, C.V., Lawton, T.K. and Benson, P.E. (1996). The Sequenced Activity Mobility Simulator (SAMS): An Integrated Approach to Modeling Transportation, Land Use and Air Quality. *Transportation*, vol. 23, pp. 267-291.

Kitamura, R. and Fujii, S. (1998) *Two Computational Process Models of Activity-Travel Behavior*. In: Garling, T., Laitila, T. and Westin, K. (eds.) Theoretical Foundations of Travel Choice Modeling. Elsevier Science, Oxford, pp. 251-279.

Miller, E.J. and Roorda, M.J. (2003). A Prototype Model of Household Activity/Travel Scheduling. CD-ROM of the *82nd Annual Meeting of the Transportation Research Board, National Research Council*, Washington, D.C.

Miller, E.J., Roorda, M.J. and Carrasco, J.A. (2005). A Tour-based Model of Travel Mode Choice. *Journal of Transportation*, vol. 32, no. 4, pp. 399-422.

Pendyala, R.M. (2004). Phased Implementation of a Multimodal Activity Based Modeling System for Florida. FAMOS: The Florida Activity Mobility Simulator. Final Report Submitted to the Florida Department of Transportation, Research Center. Volume I: Technical Documentation. Available at: http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BA496rpt.pdf and Volume II: Users Guide. Available at: http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BA496_Manual.pdf (assessed December 2008).

Petersen, E. and Vovsha, P. (2005). Auto Allocation Modeling in Activity-Based Travel Models. Presented at the *10th TRB Transportation Planning Application Conference*, Portland.

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

Wets, G., Vanhoof, K. Arentze, T.A. and Timmermans, H.J.P. (2000). Identifying Decision Structures Underlying Activity Patterns: An Exploration of Data Mining Algorithms. *Transportation Research Record*, no. 1718.

Vovsha, P. and Petersen, E. (2007). Incorporating Car-Type Preferences and Intra-Household Car Allocation into Travel Demand Models. Paper presented at *Transportation Research Board 2007*, Washington DC, USA.

Vovsha, P. (2008). State of the Practice Metropolitan Activity Based Models in US. Paper presented at *Transportation Research Board 2008*, Washington DC, USA.