Competing Risk Hazard Model of Activity Choice, Timing, Sequencing, and Duration

DICK ETTEMA, ALOYS BORGERS, AND HARRY TIMMERMANS

Recently hazard models have become increasingly popular in transportation research for modeling duration processes of various kinds. The application of hazard models is extended to the field of activity scheduling to account for the continuous nature of the decision-making process underlying activity performance. A competing risk hazard model of the accelerated time type, which describes simultaneously the duration of the present activity and the choice of the next activity, is presented. Both a generic and an activity-specific version of the model were estimated. The covariates used in the model represent factors that affect activity scheduling, such as time of day, opening hours, travel times, priorities, and time budgets. An interactive computerized data collection procedure was used to obtain specific data needed to calculate the covariates. The estimated models performed satisfactorily, suggesting that competing risk models are a useful tool for describing activity scheduling as a continuous decision-making process. This is an important finding, especially because influencing the timing of activities and trips is a subject of increasing interest to policy makers.

In past decades, activity scheduling has been a topic of increasing interest in the transportation research community (*1*). The central assumption underlying this stream of research is that people travel to participate in various activities that satisfy their personal needs. Thus, the key question in understanding how travel decisions are made and how people will adapt their travel behavior to changes in their environment is how people decide about activity performance and related travel behavior. More specifically, it requires an understanding of the activity scheduling process, which encompasses decisions about which activities to perform, at which locations, at which times, in which sequence, and which travel modes and routes to use.

Modeling efforts in transportation have addressed several aspects of activity scheduling. For instance, discrete choice models of destination choice, mode choice, and route choice are well known and widely applied, whereas multidimensional models encompassing several of these choices, often using a nested logit approach, are becoming increasingly popular (*2*). More specific applications include models of combined activity and destination choice throughout the day (*3*) and trip chaining models, describing the sequencing of activities (*4,5*). Other approaches describe the choice of complete activity patterns explained by their scheduling convenience (*6*) or the planning phase that precedes activity execution (*7*).

Another approach in choice modeling with possibly relevant implications for activity scheduling is the development of dynamic discrete choice models (*8,9*). These models typically describe how choice behavior develops over time. By including state dependence and heterogeneity, the choice made at time t is explained partly by choices made previously so that changes in behavior are modeled rather than independent choices. Models of this type have been applied in the analysis of panel data to describe vehicle transactions and various kinds of travel behavior (*8,7*). Applications in the field of activity scheduling, however, are scarce.

Dynamic discrete choice models, however, do not go without severe computational difficulties, especially if the number of alternatives and waves is large, which is typically the case in activity scheduling analysis. Furthermore, activity performance is increasingly regarded as a continuous process, in which individuals can decide during activity performance to end an activity and start another one. The decision whether to continue or stop will therefore depend strongly on time and duration of the present and previous activities. Thus, the probabilities of pursuing different activities and travel to different locations will change continuously over time. Both static and dynamic discrete choice models do not explicitly account for this duration dependence.

Recently, hazard models have gained increasing interest in transportation research as a means to describe the duration of processes such as activity performance (*10*). Hazard models therefore are promising tools for incorporating duration dependence into activity-based approaches and taking into account the continuous nature of the implied decision making. The specific contribution of this paper to the literature is the introduction of a competing risk hazard model to activity scheduling modeling to describe not only activity duration but also activity choice. Spatiotemporal constraints were incorporated by using specific individual data on available locations and hours obtained by using a computerized interactive data collection procedure. The results indicate that transitions between activity types can be described by a competing risk model with covariates accounting for spatiotemporal flexibility of activities.

The remainder of this paper is organized as follows. In the second section, hazard models are introduced and discussed. Special attention is given to competing risk models and issues of heterogeneity and risk interdependency. In the third section, the model that was used in the present research is discussed. The model structure and the covariates, representing spatiotemporal constraints, are outlined. In the fourth section, the data collection procedure, which was performed using a recently developed interactive computer procedure, is described. The fifth section describes the results of the analyses. Different specifications of the competing risk model are discussed. Finally, the sixth section summarizes the findings and addresses directions for future research.

THEORETICAL BACKGROUNDS OF HAZARD MODELS

Basic Concepts

In this paper, a series of hazard models is applied to describe and analyze activity scheduling processes. Because hazard models have

Urban Planning Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

not been widely used in transportation research the basic principles of hazard models will be discussed and summarized to allow a better understanding of the empirical findings of this study.

Although hazard models only recently have gained increasing popularity in transportation modeling, they have been applied for decades in other disciplines, such as industrial engineering, biology, medical science, and labor market research (*11,12*). Hazard models typically are applied to describe duration data such as machine failure times or patient survival times under different medications or unemployment periods. More specifically, hazard models describe the probability of occurrence of a certain event (machine failure, death, finding a job) within an interval $[t,t + dt]$, given that it has not occurred up to time t. This conditionality can be considered the key concept of hazard modeling and offers a natural framework for describing durations and intervals between the occurrence of events. For instance, in the case of activity duration, the probability of stopping an activity will be small when it has just started and will gradually increase with the time of execution. Hazard models offer the statistical tools to describe this conditional probability, which enables one to incorporate duration dependence into transportation modeling.

Mathematical Formulation

A number of functions are of particular interest with respect to the mathematical description of hazard models. First, a probability density function $f(t)$ giving an unconditional distribution of durations T within a population can be defined as

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \ T \leq t + \Delta t)}{\Delta t} \tag{1}$$

The probability that in a specific case the event will occur before time t is then

$$F(t) = P(T < t) = \int_0^t f(u)du \tag{2}$$

It follows that $f(t)$ is the first derivative of $F(t)$ with respect to time. A key function in hazard modeling is the survivor function $S(t)$, giving the probability that the process has survived until t:

$$S(t) = 1 - F(t) = P(T \geq t) = \int_t^\infty f(u)du \tag{3}$$

The hazard function $h(t)$, finally, describes the probability of occurrence at t conditional on survival until t:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \,|T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{4}$$

In principal, the hazard can take many different forms (see Figure 1). It can be monotonically increasing (*a*), U-shaped (*b*), monotonically decreasing (*c*), or constant (d). Lawless (*11*) and Kalbfleisch and Prentice (*13*) give examples of shapes that are typical for certain types of duration processes. Given that the shape of the hazard yields important information about the nature of the process under study, remarkably little attention has been paid to the specific shape of the hazard in transportation applications. The emphasis has been primarily on the influence of covariates influencing the scale of the hazard, indicating longer or shorter durations in general.

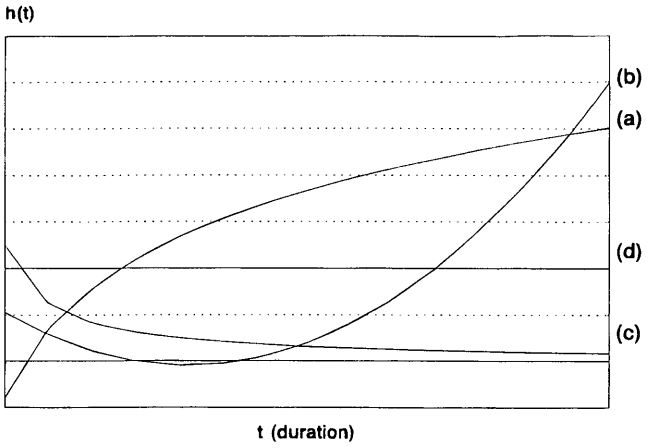

FIGURE 1 Some hazard functions.

The shape of the hazard function is determined by the distributional assumptions that are made for the probability density function $f(t)$. A number of different distributions can be chosen (*11*), resulting in different hazard functions. Some distributions and their related hazard functions are listed below. For a detailed review of possible distributions the reader is referred to Lawless (*11*) and Kalbfleisch and Prentice (*13*).

1. Exponential distribution:

$$h(t) = \lambda \qquad t \geq 0 \tag{5}$$

2. Weibull distribution:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1} \qquad \lambda,\beta > 0 \tag{6}$$

3. Log normal distribution:

$$h(t) = \frac{\frac{1}{(2\pi)^{1/2}\,\sigma t} \exp\left[0.5\left(\frac{\log t - \mu)}{\sigma}\right)^2\right]}{1 - \int_{-\infty}^{t} \frac{1}{(2\pi)^{1/2}} e^{-\mu^2/2}\, du\left(\frac{\log t - \mu}{\sigma}\right)} \tag{7}$$

4. Log logistic distribution:

$$h(t) = \frac{\lambda\beta(\lambda t)^{\beta-1}}{1 + (\lambda t)^{\beta}} \tag{8}$$

The choice of a specific distribution and related hazard function usually will be made according to hypotheses based on existing theory. However, testing of different distributions with different scale and shape parameters may often lead to a better insight into the duration process under study.

Parametric Hazard Models

Apart from duration dependence other factors also influence activity duration and timing. For instance, the start of an activity may be

influenced by opening hours, time of day, or priority of the activity. To incorporate such explanatory variables into the model two model types can be used. The first is known as the proportional hazard model, which takes the following form:

$$h(t \mid X) = h_0(t)g(X) \tag{9}$$

where

X = a vector of explanatory variables and
$h_0(t)$ = the baseline hazard function.

The baseline hazard is the hazard function assuming that all covariates X have a value of 0. $g(X)$ is usually defined as $exp\ (\beta X)$, where β is a vector of parameters. The function g thus acts multiplicatively on the baseline hazard. This causes the property of proportionality, implying that the ratio of hazards for specific sets of covariates (h_1/h_2) remains constant over time. This assumption however can in some cases be undesired. For instance, Leszczyc and Timmermans (*14*) found that intershopping trip times differed, depending on the store chains that were visited. In this research, different duration processes may be expected for different types of activities. In such cases the proportion of hazards of different destinations is likely to vary over time. Accelerated lifetime models can be used to describe such cases. These models are log linear for T:

$$\log T = X\beta + \epsilon$$

The hazard function in this case can be shown to be

$$h(t|X) = h_0\,(te^{-X\beta})e^{-X\beta} \tag{11}$$

Thus, the effect of the covariates X is on t rather than on the baseline hazard. The models are not proportional and offer greater flexibility in modeling durations of alternative processes. This, however, comes at the cost that heterogeneity cannot be incorporated into the model. In both cases different forms of the hazard function are obtained by taking different distributions for the baseline hazard h_0 as described earlier.

Competing Risk Models

The previous description has considered durations of processes with only one exit. However, the ending of an activity can be the means to starting various new activities so that there will be different possible exits. A competing risk model was used to describe transition rates to these competing risks. In this case, hazards are defined specifically for different exits:

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t,\ D_k = 1 | T \ge t)}{\Delta t} \tag{12}$$

where $h_k(t)$ is the probability that exit k occurs at time t and D_k is a dummy variable indicating whether or not exit k was chosen.

The relation between hazards and survival functions for specific exits and joint hazard and survival functions is given simply by

$$h(t) = \sum_k h_k(t)dt \tag{13}$$

$$S(t) = \sum_k S_k(t) \tag{14}$$

Parametric versions of the proportional and accelerated time type are written as follows:

$$h_k(t \mid X) = h_{0k}\,(t)e^{X_k\beta_k} \tag{15}$$

$$h_k(t \mid X) = h_{0k}(te^{-X_k\beta_k})e^{X_k\beta_k} \tag{16}$$

where different distributional assumptions can be made for h_{0k}, the k-specific baseline hazard.

The probability that, if an exit is chosen, this exit will be k, π_k can be calculated as follows:

$$\pi_k = \int_0^\infty S(s)\, h_k(S)\, ds \tag{18}$$

Lancaster (*12*) shows that under the assumption of stationarity ($h_k(t)/h(t) = m_k$ for all t) and a Weibull distribution for the probability density function π_k can be written as follows:

$$\pi_k = \frac{\exp\,(x_k\,\beta_k)}{\sum_j \exp\,(x_j\,\beta_j)} \tag{19}$$

Thus, the well-known logit model can be regarded as a competing risk model under strong assumptions. This example clearly illustrates that competing risk models offer the attractive opportunity of relaxing the static assumption underlying discrete choice modeling and incorporating dynamic aspects into consumer behavior research. In a number of transportation applications, especially where the timing of travel decision is concerned, this might be a valuable contribution.

Notwithstanding these attractive features, some issues should be addressed in the application of competing risk models. Competing risk models fall in the class of models with multivariate lifetime distributions, with different distributions of lifetimes T_i according to the competing risks. If information on all lifetimes T_i is available, one can test for independence of the various lifetime distributions. However, in the case of competing risks only min ($T_1, \ldots, T_k$) is observed so that the assumption of independence cannot be tested. This is caused by the fact that it is principally impossible to discriminate between different multivariate distributions $f(t_1, \ldots, t_k)$ that give rise to the same cause-specific hazard functions based on min ($T_1, \ldots, T_k$) only (*11*). Recently, Han and Hausman (*15*) introduced a proportional hazard model that allows for testing of independence among risks. In their approach, time is divided into T discrete periods and a proportional hazard model is formulated in an ordered logit or ordered probit form. Interdependency can then be incorporated by correlations in the stochastic terms of the model.

A second issue that should be addressed is the problem of heterogeneity within the sample. In case of observed heterogeneity, characteristics of subjects that can easily be measured, such as sociodemographics, influence the observed behavior. Heterogeneity can then be accounted for by including the sociodemographics as explanatory variables in the model. Unobserved heterogeneity

exists when unobserved characteristics of subjects in the sample (e.g., motivations, tastes, and preferences) correlate with the observed behavior—in this case activity choice and duration. Not accounting for heterogeneity may lead to biased results. For instance, Meurs (*9*) found that linear regression models without heterogeneity lead to underestimation of elasticities. The effects of ignoring heterogeneity in duration models are less clear cut. Studies by Hensher (*16*) and De Jong et al. (*17*) seem to suggest that including heterogeneity does not have a dramatic effect on the parameter estimates of the explanatory variables but has a larger impact on the shape and scale parameters of the distribution of the baseline hazard. An additional complication arises when multiple observations for one subject are included in the sample, for example, if panel data or multispell duration data are used. If heterogeneity exists, the observations of one subject will be interdependent. By treating the observations as independent, one can easily overestimate the effects of state and time dependence and habit persistence (*18*).

To account for heterogeneity in proportional hazard models usually a heterogeneity term is introduced, which is a random variable with a certain (often gamma) distribution (*13,16*). Lancaster (*12*) and Sueyoshi (*19*) extend the inclusion of a mixing distribution to the competing risk case. By specifying mixing distributions V_j for competing risks, the joint distribution can be used to account for interdependency between risks. However, in the case of accelerated time models, introduction of a heterogeneity term is not possible because of identification problems (*20*).

COMPETING RISK MODEL OF ACTIVITY CHOICE, SEQUENCING, TIMING, AND DURATION

In the current research, the sequencing and timing of activities during the course of a day are of interest. Two models were estimated to describe this process. Both models describe the transition from one activity to another. The competing risks by which the origin activity can end are the possible following activities. The dependent variable in the models is the duration of the first activity that is equivalent to the time until a transition to another activity takes place. However, the covariates used to explain the duration are generic in one model and specific for various activity types in the other model. For this application, models should be obtained in which the proportion of transition probabilities to different activity classes can change over time. This implies an accelerated time formulation for each model. The generic model can be specified as

$$h_k(t) = h_0(te^{-\beta X})\, e^{-\beta X} \tag{20}$$

where

$h_k(t)$ = hazard function for transition to any activity k,
h_0 = baseline hazard,
X = vector of generic covariates, and
β = vector of generic parameters.

The specific model is given by

$$h_k(t) = h_0(te^{-\beta_k X_k})e^{-\beta_k X_k} \tag{21}$$

where

$h_k(t)$ = hazard function specific for transition to activity k,
X_k = vector of covariates specific for activity k, and
β_k = vector of parameters specific for activity k.

Obviously, the choice for the accelerated time model has some implications. First, it does not allow incorporation of heterogeneity into the model. In addition, it is not possible to readily test for independence of activity choices. Hence, the choice of which model to use was based on a tradeoff between incorporating heterogeneity and interdependence between risks and the flexibility to allow the ratio of transition rates to different risks to vary over time. The latter should receive the priority that led to the choice of the accelerated time model in this project.

The following activities were distinguished in the specific model:

1. In-home leisure activities,
2. In-home task activities,
3. Work/education,
4. Shopping,
5. Personal business out of home (not item 3 or 4), and
6. An end state in which no further activities are performed.

The covariates X used in both models to explain activity duration and transition to other activities are derived from previous research (*21*), which revealed that spatiotemporal constraints and general characteristics of activity performance were relevant for activity scheduling behavior. The following covariates describing spatiotemporal flexibility were used:

1. The activity from which the transition takes place. Five dummies are used to represent the possible activities, being the first five activity types mentioned earlier. These are generic variables in both models. The dummies represent differences in average activity duration between different classes of activities.
2. The activity to which the transition takes place. Dummy coding was used in a similar way to represent the six possible destination states. The dummies represent the effect of the destination activity on the duration of the preceding activity.
3. START: the start time of the first activity in minutes. It is assumed that the time of day at which activities start may influence the probability of transition to another activity. For instance, the probability of switching to leisure activities may be higher at the end of the day, whereas switching to work is more likely at the beginning of the day.
4. TILSTART: the time until the next activity can start in minutes. This factor represents the influence of opening hours of facilities or the influence of fixed hours for certain activities, such as work or education. It is hypothesized that if less time remains until an activity can start, a transition to this activity is more likely to take place. If PS2 is the earliest possible start time of the destination activity, TILSTART is calculated as follows:

TILSTART = PS2 − START

This measure takes a value of 0 if the activity can start before START.

5. TILCLOSE: the time until the next activity can end at the latest in minutes. This factor represents the effect of closing times or the end of fixed hours for certain activities. The effect can be

twofold: if less time remains for the execution of an activity, it becomes more urgent so that transition to this activity is more likely to take place. However, if too little time remains for the execution of an activity a transition will become less likely. If PE2 is the latest possible endtime of the destination activity, TILCLOSE is calculated as follows:

TILCLOSE = PE2 − START

If START > PE2, then TILCLOSE is set to 0.

6. PRIOR1: the priority of the first activity on a 0 to 10 scale. It can be hypothesized that the priority of the first activity will influence the duration of this activity, in the sense that activities with lower priority are more likely to be ended to pursue other activities.

7. PRIOR2: the priority of the next activity on a scale of 0 to 10. Analogous to PRIOR1, a transition to an activity with higher priority is more likely to take place if the priority of this activity is higher.

8. TRAVTIME: the travel time between the origin and the destination activity in minutes. This factor represents the distance decay over time of switching to different activities.

9. TIMESPENT: the time spent on the destination activity type at earlier occasions during the same day in minutes. This factor represents history dependence. That is to say, the amount of time spent on an activity earlier in the day is likely to influence the probability of switching to the activity once more.

DATA COLLECTION

The competing risk model was estimated using activity scheduling data that were collected in January 1994. Subjects were 39 students of Eindhoven University of Technology, Eindhoven, The Netherlands. The data were collected using the interactive computer procedure MAGIC (*22*), which consists of two parts. In the first part general information on activity performance and spatiotemporal constraints is collected. For 31 activities the following information is recorded for each subject:

1. Will the activity be performed on the planning day according to an arrangement in which other people are involved (yes/no)?
2. What was the last time the activity was performed (days ago)?
3. What is the average frequency of performance of the activity (times per month)?
4. How long does is take to perform the activity (minimum time, average time, maximum time)?
5. How likely is it that the activity will be performed on the target day (on a 0 to 10 scale)?
6. What are the locations at which an activity takes place? Of each location the subject is asked to provide the following information: (*a*) the name of the location, (*b*) the hours at which the subject would consider performing the activity at this location (this may be a smaller range than is implied by strict opening hours), (*c*) the attractiveness of the location on a 0 to 10 scale, indicating how pleasant the location is to stay at, and (*d*) The address of the location.

The list of 31 activities is designed to cover the spectrum of daily and incidental activities and includes both in-home and out-of-home activities; it includes the following:

- Taking an educational course;
- Studying at home;
- Practicing hobbies at home;
- Buying provisions;
- Visiting post office or bank;
- Visiting a cafe, bar, or disco;
- Visiting a sports match;
- Sightseeing;
- Eating breakfast;
- Housekeeping;
- Visiting someone;
- Performing work;
- Visiting cashpoint;
- Participating in sports;
- Attending the theater or a concert;
- Eating lunch;
- Reading;
- Having visitors;
- Buying clothes or shoes;
- Engaging in club activities;
- Volunteering;
- Attending a museum or exhibition;
- Eating supper;
- Watching television;
- Getting food (snack bar);
- Visiting a specialty shop;
- Going to the movies;
- Visiting the library; and
- Visiting a restaurant.

In addition, travel distances between the locations mentioned by the subjects are requested. These data enable the calculation of the covariates described earlier.

In the second part of the procedure subjects are asked to perform a scheduling task; that is, they are requested to list all activities they plan to perform the day after the experiment. These activities are all on the list of activities used in the first part, so that detailed information on each selected activity is available. The schedule encompasses the planned activities and the sequence in which they are executed, the locations at which the activities take place, travel modes used, and the start and end times of activities. From these schedules the data used for estimation of both the models was derived. For each observed transition from one activity to another all competing risks, that is, all possible destination activities, were included in the data set. The set of alternative destination activities for a transition encompasses all activities from the list of 31 that were assigned a likelihood greater than 0 in the first part of the procedure. For each competing risk the values of the covariates in the generic and specific model were calculated on the basis of the information supplied in the first part of the procedure. The destination activity that was chosen by the subject was coded as an observed transition; the other competing risks were coded as right censored. The data set consisted of a total of 256 observed transitions and 7,041 right-censored cases.

The study described in this paper is exploratory in nature. A small sample that is not representative of the population of some geographic area as a whole has been used. The sample is homogeneous with respect to age (18 to 25 years), main occupation, and income (students). Therefore, sociodemographic variables are not included in the model. Furthermore, a data set that was collected using this procedure, as detailed information about spatiotemporal constraints on an individual level was obtained, was preferred in

this case but is usually not the case with existing time budget and travel surveys.

ESTIMATION AND RESULTS

Estimation Procedure

The models described earlier were estimated using the SAS package. Independence between competing risks and homogeneity was assumed. This implies the following likelihood function:

$$L = \prod_{i=1}^{N} \prod_{a=1}^{A_i} \prod_{c=1}^{C_{ai}} f_c(t_{ia}|x_{iac})^{d_{iac}} S_c(t_{ia}|x_{iac})^{1-d_{iac}} \quad (23)$$

where

N = number of individuals in sample,
A_i = number of activities performed by individual i,
C_{ai} = number of possible exits for activity a of individual i,
f_c = probability density function of duration times for Exit c,
S_c = survivor function for Exit c,
t_{ia} = time at which activity a of individual i is ended,
X_{iac} = vector of covariates associated with Exit c from activity a of individual i, and
d_{iac} = dummy variable that indicates whether Exit c was chosen for ath activity of individual i.

As noted earlier, the fact that heterogeneity is not included in the model may affect the scale and shape parameters of the baseline hazard and the estimation of lagged effects. However, the heterogeneity in the sample used for this study is diminished by the fact that the subjects were all students who differed little with respect to sociodemographic characteristics. Further, the effect of heterogeneity on the estimation of state dependence in duration models is less clear, compared with dynamic models based on panel data. Nevertheless, heterogeneity will have some effect, and this should be considered when interpreting parameter estimates.

Generic Model

The generic model was estimated with various distributions assumed for the baseline hazard. The goodness-of-fit measures for the various distributions are indicated in Table 1. As indicated by the goodness-of-fit measures, the log-normal distribution describes the transition probabilities best.

TABLE 1 Goodness-of-Fit Measures of Generic Models with Various Distributional Assumptions

distribution	loglikelihood
weibull	-1151.61
exponential	-1163.90
lognormal	-1125.08
loglogistic	-1151.00

The parameters that were estimated assuming the log-normal distribution are indicated in Table 2. Positive parameter values indicate a positive effect on the duration of the origin activity, whereas negative parameter values suggest a shorter duration. The origin dummies thus suggest that work/education usually has a relatively long duration. The positive effect on the durations of in-home leisure, in-home task activities, and personal activities out of home is smaller, whereas shopping is the activity with the shortest duration. The positive and significant sign of STARTTIME suggests that if an activity starts later, it will have a longer duration. Apparently, the probability of starting a new activity is smaller later in the

TABLE 2 Parametric Estimates of Log-Normal Generic Model

variable name	parameter (t-value)
intercept	-0.08 (-0.18)
in home leisure[1]	0.93 (3.65)
in home task[1]	0.73 (3.08)
work/education[1]	2.22 (8.70)
shopping[1]	-0.08 (-0.24)
pers. act. out of home[1]	0.88 (3.75)
starttime	0.10 (4.02)
tilstart	0.07 (1.33)
tilclose	0.06 (2.70)
prior1	0.05 (2.05)
prior2	-0.08 (-5.71)
travtime	0.27 (9.50)
timespent	0.13 (2.25)
scale	1.60 (22.22)

[1] dummy for origin activity

day. TILSTART does not have a significant effect. TILCLOSE, however, has a significant and positive effect, suggesting that transitions are postponed if more time remains for the destination activity. Thus, if there is less time pressure for the destination activity, the preceding activity will have a longer duration, as one would expect. PRIOR1 has a significant and positive value. This suggests that if the priority of the origin activity is higher, it will have a longer duration. However, a higher priority of the destination activity has the reverse effect, as indicated by the negative value of PRIOR2. Thus, if the destination activity has a higher priority, the preceding activity will have a shorter duration. TRAVTIME has a positive and significant value. Thus, if travel time to the destination activity increases, this will postpone the transition to this activity, resulting in a longer duration of the preceding activity. TIMESPENT, finally, has a positive and significant value: that is, that the more time one has already spent on an activity, the less likely one is to switch to this activity. Apparently, time budgets exist for various types of activities that set limits to the amount of time spent on one activity.

Specific Model

The specific model was also estimated with different assumptions of the distribution of the baseline hazard. The goodness-of-fit measures are indicated in Table 3. Again, the best performance is achieved assuming a log-normal distribution.

The parameters of the log-normal model are indicated in Table 4. In the specific model the origin activity was represented by a set of generic dummy variables; furthermore, an intercept was estimated for each destination activity. The parameters for the origin activity all have a significant positive value, so that the effects can be interpreted only relatively to each other. The estimated values suggest that work/education usually has the longest duration, whereas shopping has the shortest duration on average. The intercepts represent the effect of the destination activity on the duration of the preceding activity. A significant negative parameter value suggests that a transition to personal activities out of home will shorten the preceding activity. In-home task activities, however, are usually postponed, resulting in a longer duration of the preceding activity.

Positive and significant STARTTIME parameters were estimated for work/education and personal activities. Transitions to these activities are thus postponed if the preceding activity starts later. However, the negative value for the end state indicates that transitions to this category are more likely to happen later in the day. A significant parameter for TILSTART was found only for work/education. Thus, activities followed by work/education activities have longer durations if more time remains until this activity can start. The effect of TILCLOSE is significant only for in-home task activities. Contrary to the expectation and to the findings of the generic model, transition to this activity takes place earlier if time pressure is less, resulting in a shorter duration of the preceding activity. Parameters for PRIOR1 were significant only at the 10 percent confidence level for work/education and personal activities out of home. The signs indicate that a transition to work/education takes place earlier if the origin activity has a higher priority, which is contrary to the expectation in this paper. However, the opposite holds for personal activities out of home. The positive sign indicates a longer duration of the preceding activity if the priority is higher. Positive and significant parameters for PRIOR2 were found for in-home task activities, work/leisure, and shopping. If the priority of these activities increases, transitions to these activities will take place earlier, resulting in shorter durations of the preceding activities. A positive and significant parameter value for TRAVTIME was found for all activities, except the end state. Apparently, all are postponed if travel time increases. This effect is strongest for in-home leisure and relatively weak for work/education and in-home task activities. This indicates a weaker distance decay for obligatory activities, as one would expect. Significant parameters for TIMESPENT were found for work/education and personal activities out of home, indicating that transitions to work/education and personal activities out of home are postponed if more time already has been spent on these activities. This holds to an extreme extent for personal activities. Apparently, there are strict time budgets for personal activities. So, with few exceptions (PRIOR1 and TILCLOSE) the parameter signs are in line with common sense.

TABLE 3 Goodness-of-Fit Measures of Specific Models with Different Distributional Assumptions

distribution	loglikelihood
weibull	-1073.56
exponential	-1106.82
lognormal	-1050.57
loglogistic	-1060.04

COMPARISON OF GENERIC AND SPECIFIC MODEL

In terms of interpretation of the parameters, the two models are largely consistent. Shifts in the sign of parameters appeared only in the case of TILCLOSE and PRIOR1. However, parameter values vary in sign and magnitude between destination states in the specific model, indicating that different types of activities are planned according to different criteria. Therefore, the extension from a generic model to a specific model is regarded as an important improvement. To test whether the specific model performed statistically better than the generic model, a likelihood ratio test was performed. The chi-square statistic of 149.02 with 35 degrees of freedom indicates that the specific model performs significantly better at $\alpha = 0.005$.

CONCLUSION

In this paper an alternative method of modeling activity choice, timing, and duration has been described. Competing risk hazard models of the accelerated time type were used to describe the duration of an activity, the choice of a next activity, and their mutual dependency. The estimated models performed satisfactorily, suggesting that competing risk models are a useful tool for incorporating duration dependence into discrete choice modeling. This conclusion is particularly relevant as timing of activities and trips

TABLE 4 Parametric Estimates of Log-Normal Specific Model

	DESTINATION ACTIVITIES					
	in home leisure	in home task	work education	shopping	pers. act. out of home	end state
in home leisure[1]	2.21 (13.23)[2]					
in home task[1]	1.97 (13.83)					
work education[1]	3.49 (21.17)					
shopping[1]	1.08 (4.82)					
pers. act. out of home[1]	2.34 (16.69)					
intercept for destination	-3.23 (-1.21)	1.93 (2.03)	-2.01 (1.85)	-1.32 (-0.75)	-2.34 (-3.66)	4.62 (3.70)
starttime	0.20 (1.09)	-0.11 (-1.83)	0.31 (4.25)	0.22 (1.85)	0.14 (2.98)	-0.55 (-6.88)
tilstart	0.21 (0.81)	0.09 (0.91)	0.35 (2.06)	--	0.12 (1.78)	--
tilclose	0.22 (1.24)	-0.17 (-3.40)	0.05 (0.64)	0.11 (1.22)	-0.02 (-0.41)	--
prior1	-0.03 (-0.54)	0.06 (1.39)	-0.15 (-1.93)	0.03 (0.23)	0.06 (1.84)	0.05 (0.60)
prior2	0.00 (0.56)	-0.16 (-4.06)	-0.11 (-2.59)	-0.25 (-3.29)	-0.05 (-1.93)	--
travtime	0.41 (3.52)	0.15 (2.46)	0.10 (1.77)	0.24 (2.53)	0.26 (6.77)	-0.04 (-0.54)
timespent	0.08 (0.28)	0.14 (1.08)	0.27 (2.32)	--	15.51 (4.06)	--
scale = 1.34 (24.81)						

[1] dummy for origin activity
[2] t-values in parentheses

becomes the subject of policy making to reduce congestion and preserve mobility.

In this application two models were estimated: a generic model and a specific model, which is conditional on the destination state. The specific model was superior to the generic model because the goodness of fit of this model was significantly higher. Furthermore, the estimated parameters reflected differences between scheduling criteria for different activity types, which would remain unrevealed in the generic model. Parameter estimates suggest that spatiotemporal constraints such as time of day, opening hours, and travel time play an important role in activity scheduling and timing. Also the history of the pattern and priorities of activities influence timing and choice of activities. The inclusion of these covariates was enabled by a specific computerized data collection procedure, which provides an extensive record of individual activity scheduling processes. Application of such a procedure can be considered a prerequisite if one wants to obtain models that are capable of describing travel behavior at a detailed level. Both the generic and the specific models were estimated by assuming different specifications of the baseline hazard. Of the tested distributions, the log-normal distribution provided the best fit in predicting activity transitions.

The modeling approach described in this paper is a first step in a new direction of modeling and simulating the performance of activity patterns. However, improvements still need to be made. An issue already raised is how heterogeneity and interdependency of risks should be handled. In this study flexibility of the model structure was allowed to prevail over heterogeneity and interdependency. Future research, however, should address possible ways of unifying the above properties into one model structure. Another development would be the extension of the unconditional competing risk models described above to conditional models, where transition probabilities are dependent on both origin and destination states.

REFERENCES

1. Jones, P. *New Developments in Dynamic and Activity-Based Approaches to Travel Analysis.* Gower, Aldershot, England, 1990.
2. Algers, S. Integrated Structure of Long-Distance Travel Behavior Models in Sweden. In *Transportation Research Record 1413,* TRB, National Research Council, Washington, D.C., 1993, pp. 141–149.
3. Kitamura, R., and M. Kermanshah. Sequential Model of Interdependent Activity and Destination Choices. In *Transportation Research Record 987,* TRB, National Research Council, Washington, D.C., 1984, pp. 81–89.
4. Kitamura, R. Incorporating Trip Chaining into Analysis of Destination Choice. *Transportation Research,* Vol. 18B, 1984, pp. 67–81.
5. Timmermans, H., and P. Van der Waerden. Modelling Sequential Choice Processes: The Case of Two-Stop Trip Chaining. *Environment and Planning A,* Vol. 24, 1992, pp. 1483–1490.
6. Recker, W. W., M. G. McNally, and G. S. Root. A Model of Complex Travel Behavior: Part 2: An Operational Model. *Transportation Research,* Vol. 20A, 1986, pp. 319–330.
7. Ettema, D., A. Borgers, and H. Timmermans. Simulation Model of Activity Scheduling Behavior. In *Transportation Research Record 1413,* TRB, National Research Council, Washington, D.C., *1413,* 1998, pp. 1–11.
8. Heckman, J. J. Statistical Models for Discrete Panel Data. (C. F. Manski and D. McFadden, eds.), In *Structural Analysis of Discrete Data with Econometric Applications* MIT Press, Cambridge, Mass., 1981.
9. Meurs, H. *A Panel Data Analysis of Travel Demand.* University of Groningen, The Netherlands, 1991.
10. Niemeier, D. A., and J. Morita. Duration of Trip-Making Activities by Men and Women: A Survival Analysis. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
11. Lawless, J. F. *Statistical Models and Methods for Lifetime Data.* Wiley, New York, 1982.
12. Lancaster, T. *The Econometric Analysis of Transition Data.* Cambridge University Press, England, 1990.
13. Kalbfleisch and Prentice. *The Statistical Analysis of Failure Time Data.* Wiley, New York, 1980.
14. Popkowski Leszczyc, P. T. L., and H. J. P. Timmermans. Consumer Store Choice Dynamics: An Analysis of the Competitive Market Structure for Grocery Stores. *Environment and Planning A* (in press).
15. Han, A., and J. A. Hausman. Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics,* Vol. 5, 1990, pp. 1–28.
16. Hensher, D. The Timing of Change for Automobile Transactions: A Competing Risk Multispell Specification. Presented at 7th International Conference on Travel Behavior, Valle Nevado, Santiago, Chile, 1994.
17. De Jong, G., R. Kitamura, and J. Klooster. A Disaggregate Model of Vehicle Holding Duration, Type Choice and Use. Presented at 7th International Conference on Travel Behavior, Valle Nevado, Santiago, Chile, 1994.
18. Hensher, D. A., and F. L. Mannering. Hazard-Based Duration Models and Their Application to Transport Analysis. *Transport Reviews,* Vol. 14, 1994, pp. 63–82.
19. Sueyoshi, G. T. Semiparametric Proportional Hazards Estimation of Competing Risk Models with Time-Varying Covariates. *Journal of Econometrics,* Vol. 51, 1992, pp. 25–58.
20. Ridder, G. The Non-Parametric Identification of Generalized Accelerated Failure-Time Models. *Review of Economic Studies,* Vol. 57, 1990, pp. 167–182.
21. Ettema, D., A. Borgers, and H. Timmermans. Using Interactive Computer Experiments for Identifying Activity Scheduling Heuristics. Presented at 7th International Conference on Travel Behavior, Valle Nevado, Santiago, Chile, 1994.
22. Ettema, D., A. Borgers, and H. Timmermans. Using Interactive Computer Experiments for Investigating Activity Scheduling Behaviour. *Proc., PTRC 21st Summer Annual Meeting,* University of Manchester, England, 1993, pp. 267–282.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.