An additive risk based multistate model for activity chaining behavior analysis

Tai-Yu Ma, Iragaël Joly, Charles Raux

An additive risk based multistate model for activity chaining behavior analysis

Tai-Yu Ma*, Iragaël Joly, Charles Raux

Laboratoire d'Economie des Transports (CNRS, Université de Lyon, ENTPE)
* corresponding author: email: tai-yu.ma@let.ish-lyon.cnrs.fr,
address: LET, IH, 14 av. Berthelot, 69363 Lyon Cedex 07, France.

Abstract

The time-use pattern has been recognized for its correlation with socio-demographic characteristics, spatial settings and the institutional contexts. Although the correlation between the duration of single activity participation and explicative covariates has been well studied, the interdependency structure between travel/activities conducted in different episodes of an activity chain is still a relatively unexplored subject. We consider that activity pattern is resulted from a semi-Markov stochastic process, assuming that the activity-specific transition probability depends on its state type, sojourn times since entering occupied state and related covariates. Conditionally on the sequence of states visited previously, the semi-Markov property admits the estimation of state transition probability separately with cause-specific covariates. The multistate model provides a relevant framework to investigate the activity chaining behavior and related state transition probability estimates. To investigate the time-dependent effects of covariates on travel/activity duration, an Aalen's additive hazard model with competing risk is applied for model estimate. Different from usual Cox proportional hazard models (Cox 1972), the additive risk model incorporates the effects of covariates in an additive way relaxing the proportionality assumption of Cox proportional hazard models and provides a more flexible way to investigate the effect of temporal constraints on the activity chaining behavior. The estimation results of additive model are compared with Cox model and provide practical knowledge for model specification.

1. Introduction

Time-use pattern has been recognized for its correlation with socio-demographic characteristics, spatial settings and institutional contexts. Although the correlation between the duration of single activity participation and explicative covariates has been well studied, the interdependency between travel and activities conducted in different episodes of an activity chain is still a relatively unexplored subject. Empirical evidence suggests that the timing and duration of activity depends on its type and also on the activity previously conducted. The study on activity duration needs to take into account the dependency between activities conducted sequentially. If one neglects the dependency effects between activities, the duration model could be misspecified (Popkowski Leszczyc and Timmermans, 2002).

Previous empirical studies attempted to identify the interdependency of activity durations conducted in different episodes of activity chain. Ettema et al. (1995) investigated the effects of temporal constraints on activity choice, timing and its duration based on a parametric competing risk model. The results suggested that the activity choice, timing and duration are correlated with spatiotemporal constraints under activity chaining process. Bhat (1996) proposed an outcome-specific proportional hazard model to estimate the multiple types of activity durations by combining activity choice and duration in its specification. Srinivasan and Guo (2007) utilized a mixing distribution for the random error term in a joint hazard-based model to investigate the correlation between durations of adjoined activities. Similarly, Pendyala and Bhat (2004) applied a discrete-continuous simultaneous equation model to investigate the causal structure of activity timing and duration. They found that the activity timing and duration are closely related for non-commuters but loosely related for commuters. Popkowski Leszczyc and Timmermans (2002) utilized conditional and unconditional parametric competing risk models to estimate activity duration and the effects of socio-demographic covariates. The estimation results showed that conditional competing risk model fitted best, confirming that the dependency of activity choice and its duration for adjoining activity realization.

Although these studies provided some empirical evidence for temporal dependency on activity chaining behavior, most of them are still limited within one or two episodes and lack simultaneous consideration of the dependency of activity-type, timing and temporal constraints in the activity chaining process. Moreover, most studies utilized Cox proportional hazard models (Cox, 1972) to estimate activity-to-activity transition hazards by assuming relative risk is proportional with respect to explanatory variables. Although the assumption simplifies the interpretation of results and gives the parameter estimation easier, it is usually not verified for all covariates. Moreover, even the proportionality assumptions are checked, the Cox model may produce some misleading results (Amato, 1988). Other critics on Cox model includes its weakness in showing time-varying effects of covariates and also the lack of consistency in proportional risk assumption when deleting covariates or changing the precision of covariate measurements (Aalen, 1989, 1989 and 1993). To improve these shortages, an alternative additive regression model was developed by Aalen. The additive risk model incorporates the effects of covariates in an additive way, relaxing the proportionality assumption of Cox model. Moreover, it provides excess risk profile of the covariates allowing the investigation of the covariate-specific time-varying effects on survival times. Some recent applications of additive model in modeling competing risks in cancer studies can be found in Klein (2006).

In this work, we propose a multistate semi-Markov model to analyze individual's travel-activity duration pattern. The travel-activity duration formation is assumed following a semi-Markov stochastic process. Conditionally on the travel/activity previously conducted, the semi-Markov assumption admits the estimation of state transition (terminating one travel or activity participation and entering another one) probability separately with cause-specific covariates. The assumptions state that the state transition probability depends on its type, sojourn times since entering occupied state and related covariates. Based on the proposed multistate model, one can estimate the transition hazards over episodes with competing activity choice. The time-varying effects of covariates on activity duration are examined based on Aalen's additive model and compared for different classes of activities over episodes. The estimation results of the additive model are also compared with Cox model, providing empirical investigation of these two models.

2. Travel-activity duration pattern formation based on semi-Markov process

Consider an individual's observed travel-activity alternate duration sequence $S = \{s_1, ..., s_n\}$ within a period of time. This sequence composes of a set of events (a, T) with a being the event (state) and T being the event duration. In our context of analysis, the event means the performance of a trip or an activity chosen among a finite activity choice set. We assume individual's travel and activity time use pattern formation follows a stochastic semi-Markov process, i.e. the duration of travel/activity episode is time-dependant, conditional on its adjoining state. The duration of episode k $\tau^k = t_{k+1} - t_k$ is assumed to be a continuous random variable following some unknown probability distribution to be estimated. Based on the assumption of semi-Markov process, the probability distribution of travel/activity duration τ^k in episode k satisfies:

$$\mathrm{P}(t_{k+1} - t_k \le T, a_{k+1} = j | s_1, ..., s_k) = \mathrm{P}(t_{k+1} - t_k \le T, a_{k+1} = j | s_k) \tag{1}$$

where T is a continuous random variable representing the sojourn times in kth episode of an travel-activity chain. The model specification under semi-Markov process allows us to estimate the transition hazards over episodes for which the travel-to-activity transition hazards estimation leads to a competing risk model specification, i.e. the subject enters into one of competing states at the end of one episode. As the transition hazard depends on its entering and exit states, its state-dependent hazard function needs to be specified with respect to related covariates. To this end, let $\tau^k(t) = t - t_k$ represents the sojourns times in episode k since entering current state until time t. When one transition occurs at the end of episode k, the sojourn time is evaluated with respect to the entering time t_k. We call the time elapsed from t_k to t as the *renewal time* $\tau^k(t)$ (sojourn time) with respect to episode k. The distribution of the renewal time is independent, conditional on the sequence visited by a Markov chain. The one-step transition hazard $\lambda_{ij}^k(\tau_{ij}^k(t))$ at time t, i.e. renewal time $\tau_{ij}^k(t)$, from state i to state j at the end of kth episode of the travel-activity chain is defined as:

$$\lambda_{ij}^k\left(\tau_{ij}^k(t)\right) = \lim_{h \to 0} \frac{P\left[\tau_{ij}^k(t) \le T \le \tau_{ij}^k(t) + h, \, a\left(\tau_{ij}^k(t) + h\right) = j \,\middle|\, a\left(\tau_{ij}^k(t)\right) = i\right]}{h} \tag{2}$$

where $a(\tau_{ij}^k(t))$ is the state at renewal time $\tau_{ij}^k(t)$. The transition rate $\lambda_{ij}^k(\tau_{ij}^k(t))$ represents the changing rate of the transition probability from state i to state j at the renewal time $\tau_{ij}^k(t)$. For simplifying the notation, $\tau_{ij}^k(t)$ is denoted as τ hereafter. The schematic representation of multistate model for travel-activity alternate duration sequence formation is illustrated in Fig. 1. Note that in activity episode, an individual chooses one activity among a limited activity choice set.

For the transition hazards estimation, two main reasons make us choose semiparametric (Cox) model and its extension (Aalen's additive hazard model). Firstly, there is no priori information about the hazards distributions, and secondly, the effects of covariates on transition hazards are of our interests. We detail the hazard function specifications of the two models and the estimation techniques in the next section.

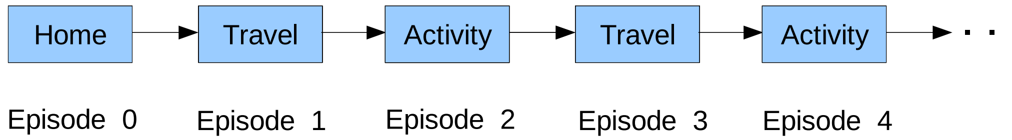

Fig. 1 Schematic representation of multistate model for travel-activity alternate duration sequence formation

2.1 Cox proportional hazard and Aalen's additive hazard model for transition hazard estimation

The survival data in the analysis consists of the duration of travel/activity with right censoring conducted in each episode. The survival data in question can be represented by a triplet of variables $(T_{ijm}^k, \delta_{ijm}^k, \mathbf{X}_{ijm}^k)$, $m = 1,\ldots,N$, where T_{ijm}^k denotes a positive random variable, representing individual *m's* sojourn times in state *i* until the next transition to state *j* in episode *k*. δ_{ijm}^k represents an indicator being 1 if the transition (*k*, *i*, *j*) is observed for individual *m*, 0 otherwise. The triplet (*k*, *i*, *j*) denotes a transition from state *i* to state *j* at the end of *k*th episode. $\mathbf{X}_{ijm}^k$ denotes the covariate column vector for the transition (*k*, *i*, *j*) of individual *m*. The transition hazards $\lambda_{ij}^k(\tau;\mathbf{X}_{ij}^k)$ can be modeled by a general regression model with respect to the transition-specific covariates $\mathbf{X}_{ij}^k$ as:

$$\lambda_{ij}^k(\tau;\mathbf{X}_{ij}^k) = v(\mathbf{X}_{ij}^k), \quad \forall i \neq j \tag{3}$$

where

$\lambda_{ij,0}^k(\tau)$: unspecific baseline hazard function with respect to transition (*k*, *i*, *j*) ,

$\mathbf{X}_{ij}^k$: column vector of transition-specific time-independent covariates for the transition (*k*, *i*, *j*)

If the regression model is specified as a non-parametric baseline hazard multiplied by the effects of covariates as

$$v(\mathbf{X}_{ij}^{k\,t}) = \lambda_{ij0}^k(\tau)\exp(\mathbf{X}_{ij}^{k\,t}\boldsymbol{\beta}) \tag{4}$$

, which leads to the Cox proportional hazard model with the hazards ratio between individual $\mathbf{X}_a$ and $\mathbf{X}_b$ being $\exp(\mathbf{X}_a)/\exp(\mathbf{X}_b)$. Note that the hazards ratio is constant, depending on the difference of the value of covariates. The estimation method is based on maximum partial likelihood estimator where each state transition contributes a partial likelihood. To derive the maximum likelihood estimate for $\boldsymbol{\beta}$, one can apply Newton-Raphson method to obtain the estimators and the covariance matrix (Kalbfleisch and Prentice, 2002).

Alternatively, the regression model in Eq. (4) can be specified as a term of time-dependent non-parametric baseline hazards $\beta_{ij0}^{k}(t)$ and a term of additive effects (excess risk) of covariates as:

$$\lambda_{ij}^{k}(\tau;\mathbf{X}_{ij}^{k}) = \beta_{ij0}^{k}(\tau) + \sum_{l=1}^{p}\beta_{ijl}^{k}(\tau)X_{ijl}^{k} \tag{5}$$

, which leads to Aalen's additive model. Note the additive effects represent the excess risk due to covariates with respect to the baseline hazard $\beta_{ij0}^{k}(\tau)$.

2.2 Parameter estimations and model fit test

To estimate the dynamical coefficient $\boldsymbol{\beta}_{ij}^{k}(\tau)$, Aalen (1989) proposed to a construct one counting process for each individual and apply a least-square technique for the parameter estimation. The individual's at-risk counting process is designed as a $N_{i}^{k} \times (p+1)$ matrix with N_{i}^{k} being the number of subjects under observation in state i and episode k. p is the number of covariates. To estimate the time-varying coefficient $\beta_{ij}^{k}(\tau)$, it is easier by constructing the cumulative hazard function $B_{ij}^{k}(\tau)$, where its slope represents the crude estimate of $\beta_{ij}^{k}(\tau)$. The cumulative hazard function $B_{ij}^{k}(\tau)$ is written as:

$$B_{ij}^{k}(\tau) = \int_{0}^{\tau}\beta_{ij}^{k}(u)du, \quad l = 0,1,2,...,p \tag{6}$$

The estimation of $B_{ij}^{k}(\tau)$ is based on the least-square technique. Let the risk indicator $Y_{ijm}^{k}(\tau)$ be 1 if individual *m* under observation has yet experience any of competing causes (travel/activities) and am experiencing the transition (*k*, *i*, *j*) at renewal time τ, and 0 otherwise. We construct a $N_{ij}^{k} \times (p+1)$ matrix $\mathbf{M}_{ij}^{k}(t)$ for which the *m*th row is set as $Y_{ijm}^{k}(\tau)(1, X_{ij1}^{k}, X_{ij2}^{k},..., X_{ijp}^{k})$, representing individual m's at-risk process. Note that N_{ij}^{k} is the number of independent observed survival times (travel or activity durations) without ties for the transition (*k*, *i*, *j*). The least-square estimates of the vector $\mathbf{B}_{ij}^{k}(\tau) = (B_{ij0}^{k}(\tau), B_{ij1}^{k}(\tau), B_{ij2}^{k}(\tau),..., B_{ijp}^{k}(\tau))^{t}$ can be obtained as:

$$\hat{\mathbf{B}}_{ij}^{k}(\tau) = \sum_{T_m \le t}[(\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{M}_{ij}^{k}(T_m)]^{-1}\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{I}_{ij}^{k}(T_m) \tag{7}$$

where T_m denotes individual *m's* sojourn times in state *i* until the next transition to state *j* in episode *k* and $\mathbf{I}_{ij}^{k}(\tau)$ be an $N_{ij}^{k} \times 1$ vector with *m*th element being 1 if individual *m* is at risk at renewal time τ, and 0 otherwise. Note that the cumulative hazard function exists only when the matrix $[\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{M}_{ij}^{k}(T_m)]$ is not singular. Hosmer and Lemeshow (1998) notes that if there are fewer than p+1 individuals in risk set, the matrix will be singular.

The variance-covariance matrix of $\mathbf{B}_{ij}^{k}(\tau)$ can be obtained as:

$$\mathrm{Var}(\hat{\mathbf{B}}_{ij}^{k}(\tau)) = \sum_{T_m \le t}[(\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{M}_{ij}^{k}(T_m)]^{-1}\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{I}_{ij}^{k\,D}(T_m)\mathbf{M}_{ij}^{k}(T_m)[[\mathbf{M}_{ij}^{k\,t}(T_m)\mathbf{M}_{ij}^{k}(T_m)]^{-1}]^{t} \tag{8}$$

, where $\mathbf{I}_{ij}^{k\,D}(\tau)$ is the diagonal matrix with diagonal element being $\mathbf{I}_{ij}^{k}(\tau)$.

For the technique details, the reader is referred to Aalen (1989, 1993) and Klein and Moeschberger (2003).

Based on Eq. (7), the estimator of the cumulative hazard function and survival function for transition (*k*, *i*, *j*) can be obtained respectively as:

$$\hat{\Lambda}_{ij}^{k}(\tau;\mathbf{X}_{ij}^{k}) = \sum_{l=1}^{p}\hat{\mathbf{B}}_{ij}^{k}(\tau)\mathbf{X}_{ij}^{k} \tag{9}$$

and

$$\hat{S}_{ij}^{k}(\tau;\mathbf{X}_{ij}^{k}) = \exp[-\hat{\Lambda}_{ij}^{k}(\tau;\mathbf{X}_{ij}^{k})] \tag{10}$$

As we assume that there is no parallel activity participation at same time, the transition hazard can be estimated by considering the other competing states as censored data. The transition hazard at state *i* on *k*th episode is simply the summation of that of competing risk *j*:

$$\lambda_{im}^{k}(\tau,\mathbf{X}_{im}^{k}) = \sum_{j \in A_i^k}\lambda_{ijm}^{k}(\tau,\mathbf{X}_{im}^{k}) \tag{11}$$

where A_i^k denote the set of possible exit states at the end of *k*th episode, depending on its current state *i* and currently occupied *k*th episode.

The investigation of time-varying effects of covariates can be easily conducted by plotting the cumulative hazard function $\hat{\mathbf{B}}_{ij}^{k}(\tau)$ over renewal time τ, where the slope presents the effects of covariates over time. The confidence interval of the cumulative hazard function can be obtained as:

$$\hat{\mathbf{B}}_{ij}^{k}(\tau) \pm z_{\alpha/2}\mathrm{S.E.}(\hat{\mathbf{B}}_{ij}^{k}(\tau)) \tag{12}$$

where $z_{\alpha/2}$ S.E.$(\hat{\mathbf{B}}_{ij}^{k}(t))$ is the estimator of the standard error of $\hat{\mathbf{B}}_{ij}^{k}(t)$ from (8).

For the test of whether a covariate has specific effect on transition hazards, Aalen proposed a test statistic based on the ratio of weighted least-square estimates and its standard deviation. An alternative test statistics due to Scheike (2002) is based on computing the absolute value of supremum of scaled cumulative martingale residuals, which represent the difference between expected (under the model) and observed number of events over time. For accessing the goodness-of-fit of the additive model, Aalen suggests using Arjas plot (Arjas,

1988) or martingale residuals (Therneau et al., 1990), which have been widely utilized for checking Cox model. The martingale residuals represent the difference of the observed number of events and the number of expected events of the assumed model over time. Hence, the martingale residuals should frustrate over the zero line if the additive model is fitted well to the observed data. As sometimes it is difficult to check the goodness-of-fit of the model by the martingale residuals plot, one can utilize the supremum test statistics to see how the summarized difference is far from zero (Scheike, 2004; Martinussen and Scheike, 2006).

3. The data

The data used in this analysis is based on household mobility surveys recently available in Lyon (2006) with the sample size of 11234 households and 27573 individuals. The survey was designed to investigate individual's travel behavior on weekdays. The travel survey was conducted for all individuals of age more than 5 years in the household. All trips conducted on the previous day before interviews are collected. The data contains individual's socio-demographic characteristics and related trip attributes. The interests of this study is to compare the Aalen's additive hazard model and Cox model for travel-activity state transition probability estimates for individual's travel-activity sequence. Given that the daily travel-activity sequence is composed of numerous episodes of travels and activities, we restrict ourselves only on the first four out-of-home activity episodes. However, one can investigate all travel and activity sequence based on the same estimation procedure.

For explicative covariate settings, individual's socio-demographic attributes, spatial and transport availability and the dependency between episodes are taken into account. The socio-demographic attributes contain gender, household type, the presence of children of age less than 12 years, worker status, and car ownership. The spatial and transport availability covariates contain the population density of the zone of household location, car ownership, distance from the zone center of household location to the nearest interchange of divided highway and distance to the nearest station of metro or tramway. The summary statistics of the covariates are listed in Table 1. As previous empirical study showed activity durations depends on its starting time of day and its state (Ma et al., 2009), these two covariates are included in the model specification. The initial out-of-home activities in parenthesis are reclassified into four categories of interest: subsistence activities (habitual work and non-habitual work), maintenance activities (daily/weekly purchase, looking for a job, administration, health, purchase of equipment, clothing or leisure) and discretionary activities (walk, sports, culture and associative activities, out-of-home eating, visit to the family or to friends) and other activities, similar as the activity categorization in the study of Bhat and Misra (1999). The sample is selected from the agglomeration of "Grand Lyon". After a data cleaning process, 9465 individuals and 36656 trips are utilized for the analysis.

Table 1 Covariate settings with respect to individual's socio-demographic, spatial and transport supply characteristics

Variable	Definition	Mean	Std. dev.
Socio-demographic characteristics			
Gender	Gender (1 if male, 0 female)	0,49	0,50
H_type	1 if individual lives in couple, 0 otherwise	0,79	0,41
Children	1 if the presence of young children with age less than 12 years in the household, 0 otherwise	0,36	0,48
Work_S	Employment status (1 have a job, 0 otherwise)	0,47	0,50
Spatial and transport availability characteristics			
Car_O	1 if one or more cars are available in the household, 0 otherwise	0,88	0,32
Density	Population density of the zone of residence location (1000 hab./km^2)*	8,81	7,60
Dist_I	Distance to the nearest interchange of divided highway (by 100m)	19,00	10,86
Dist_P	Distance to the nearest station of metro or tramway (by 100m)	15,11	15,94
Timing and duration characteristics of state/episode			
Duration_E	Duration of travel or activity conducted in previous episode (in minute)	N/A	N/A
Entering_T	Entering time of current state *i* (in minute)	N/A	N/A

Remark: 1. Motorway is defined as a road or highway in which two directions of traffic are separated by a central barrier or strip of land without direct access (neither stops, nor traffic lights).
2. The distance is calculated as the Euclidian distance of geographical centers between zone center and station/interchange of rail/road network. The agglomeration of Lyon under study is divided into 76 zones, with median zone surface and zone density being 3 km^2 and 3816 habitants/km^2, respectively.

4. Estimation results

In this section, we provide the estimation results based Aalen's additive hazard model and compare them with the Cox model for transition hazard estimation. As the number of episodes is large, we limit ourselves only on the analysis of the first four episodes. For the treatment of tied duration data, a usual technique consists of adding a uniform-distributed small random variable to survival data to avoid this problem. The Aalen's additive model and Cox model are estimated by using the Timereg package in R (R Development Core Team, 2005) and Phreg procedure in SAS, respectively.

The table 2 shows the average starting time and average activity duration over the seven main episodes in the daytime. For subsistence activity, the average starting time for the first episode is about 8 a.m. The average activity duration is about 400 minutes. Different with subsistence activity, the maintenance and discretionary activity start near 11 a.m. and 12 a.m, respectively. It is shown that later the starting time of activity, shorter its activity duration is. As for the average activity duration, the subsistence activity duration is in average 400 minutes for the first episode. The duration of maintenance activity is about 30-40 minutes for most episodes. The duration of discretionary activity is 70-80 minutes for most episodes.

Table 2 Average duration and starting time of non-travel activities over episodes

Episode \ Activity type		Subsistence		Maintenance		Discretionary	
		Starting time (hour)	Duration (minute)	Starting time (hour)	Duration (minute)	Starting time (hour)	Duration (minute)
AEP1	Mean	8.3	396.9	11.3	55.7	12.5	167.9
	Std. dev.	1.8	185.1	2.6	60.5	3.2	140.5
AEP2	Mean	10	300.2	13.1	38.2	13.7	96.6
	Std. dev.	3	196.4	3.2	46.3	2.7	91.9
AEP3	Mean	13.5	234	14.6	47.9	15.9	117.7
	Std. dev.	2.2	116	2.9	50.6	3	83.5
AEP4	Mean	14.1	203.4	15.6	38.6	16.7	112.2
	Std. dev.	2.5	119.1	2.6	37.6	2.7	82.8
AEP5	Mean	14.7	162	16.4	37.3	18	110.1
	Std. dev.	2.4	131.7	2.4	33.4	2.5	74.9
AEP6	Mean	15.1	154.4	16.7	33.2	18.5	114.4
	Std. dev.	2.7	134	2	29.7	2.6	78
AEP7	Mean	16.3	126.6	16.9	43.8	18.5	113.5
	Std. dev.	2.6	118.7	1.9	44.7	2.5	77.1

Remark: 1. AEP: non-travel activity episode

The results of the Cox model and Aalen's additive model for the transition hazard estimation over the first four episodes are shown in Table 3 and Table 4. In Table 4, we report the covariates parameter estimates of Cox model. For Aalen's additive model, as the covariates parameters are time-dependent, it is far easier to investigate covariate-specific effects by plotting cumulative hazard functions.

First, we compare the results of model estimates by examining the test statistics for the null hypothesis of $\beta_{ijl}^{k} = 0$ and $\beta_{ijl}^{k}(\tau) = 0$ with $l = 0,1,..., p$ for Cox model and additive model, respectively. The aim of the comparative study aims to investigate the estimates difference and learns about some practical knowledge for model estimations. The test statistics (p-value) for the two models are reported in Table 3. It is interesting to find that the two models show similar test results for most state transitions over episodes except Maintenance-Trip transition in episodes 1 and 4, and Discretionary-Trip transition in episode 1. The results are consistent with previous comparative studies in the literature (Aalen, 1989).

The estimated effects of covariates on transition hazards based on Cox model for different episodes are shown in Table 4. As the number of estimated episodes is large, we limit our analysis for covariates effects on the first episode. However, one can conduct similar analysis for each of activity types over episodes. For the first episode, the results indicate that men have significant longer durations of subsistence activity. Later the entering (starting) time of subsistence activity is, longer its duration is conducted. Couple, density of zone of residence location, distance to interchange of the nearest divided highway and duration of trip have less significant effects due to its coefficients close to 0. For maintenance activity, the results indicate that couple and the later starting time of activity have significant longer activity duration. The presence of young children reduces significantly maintenance activity duration. For discretionary activity, the results indicate the starting time is determinant for the duration of this activity. Later the discretionary activity begins, longer its duration is (for episode 1).

As aforementioned, Cox model assumes a constant covariate effects, which is not always true for state transition hazards estimates. To response to this research question, the test statistic for time-varying effects of additive model is conducted. The null hypothesis is stated as $\beta_{ijl}^{k}(\tau) = r$ for $l = 0,1,...,p$ to examine whether the additive covariate effects is constant. The test results are shown in Table 5. The results indicate that the baseline hazards are time-dependent for all state transitions except Discretionary-Trip transition in episode 1 and Maintenance-Trip in episode 3. In addition, the starting time of activity, the trip duration previously conducted, car ownership and the distance to the nearest station of metro or tramway have significant time-dependent effects for most of state transitions. To examine the time-varying effects, one need to check the cumulative hazard functions over the observation period. As the number of episode is large, we limit ourselves in the analysis of episode 2 conducted in the morning. The estimated cumulative hazard functions for three activity types are shown in Fig. 2 to Fig. 4. For subsistence activity, estimated cumulative hazard functions are shown for the covariates of the presence of young children, car ownership, starting time of activity and trip duration with pointwise 95% confidence interval. As shown in Fig 2, the cumulative baseline hazard increases throughout the observation period. The baseline transition hazard has substantially higher value in [240, 270] minutes compared to the period of [0, 240] and accelerates rapidly after 480 minutes. This reflects that general temporal rhythm of subsistence activity. For the significant time-varying effects of covariates, the positive/negative cumulative hazards reflect excess/reduced risk to stop an activity. The results reveal that the presence of young children has a negative effect on the durations of subsistence activity with two inverse (positive) effects within [200, 250] and after 460 minutes. This result coincides with the parameter estimator (-0.14) of Cox model in Table 4 for its negative effects. For car ownership, the result indicates the effects are not significantly differently with 0 except for marginal examples with higher subsistence durations more than 500 minutes, coincided with the parameter estimates of Cox model (Table 4). As for the effects of starting time and trip duration of precedent episode, the results conform also with the parameter estimates of Cox model, revealing that the significant negative effect of starting time and slight positive effect of trip durations. The cumulative hazard functions indicate that the hazard rate decreases rapidly for the durations of subsistence activity more than 500 minutes.

The time-varying effects of covariates on maintenance activity duration are shown in Fig. 3. The cumulative hazard functions are shown for the covariates of gender, the presence of young children, work status, distance to the nearest interchange of divided highway and the starting time of activity. The results indicate that gender has no significant influence on the durations of maintenance activity since the slope of cumulative hazard function is near to 0, but it appears its negative/positive influence for activity durations more than 150 minutes. For the effect of work status and distance to the nearest station of metro/tramway, the results indicate that they have significant positive effects on the durations of maintenance activity, and work status has larger influence than the distance to the nearest station of metro/tramway. Similar with subsistence activity, the effect of starting time has negative effect on the durations of maintenance activity. The comparison with the parameter estimates of Cox model (Table 4) for these covariates indicates that the two models coincide not only in the positive/negative effect but also in the magnitude of influence of covariates.

Finally, we investigate the time-varying effects of covariates on the durations of discretionary activity (Fig 4). The cumulative hazard functions are estimated for the covariates of household type, density of zone of residence, distance to the nearest station of metro/tramway and the starting time of discretionary activity. The results indicate that the baseline hazard rate increase constantly in the period of [0, 200] minutes, but then it seems to disappear. The effects of household type, density of zone of residence, distance to the nearest station of metro/tramway have slight influence on the durations of discretionary activity since the slopes of corresponding cumulative hazard functions is quite flat. The results compared with the parameter estimates of Cox model (Table 4), indicating that they have no significant effects on the durations of discretionary activity. Finally, the effect of starting time have significant decreased negative effects on the durations of discretionary activity, but it seems to disappear after 200 minutes.

By comparing the results of parameter estimates based on Cox model and Aalen's additive model, we found that the two models give similar estimation results. The advantage of applying additive model is that the time-varying effects of covariates on transition hazard rate can be easily investigated. It provides more information to the interpretation of covariates effects compared with Cox model. Interesting, although the model specifications of Cox model (multiplicative effects of covariates on baseline hazard) and additive model (additive effects of covariates on baseline hazard) are different, the empirical results reveal their complementarities for investigating the effects of covariates on the durations of activity.

Table 3 Test statistics for Cox model and Aalen's additive hazard model for the covariates parameter estimates (p-value)

Episode	State transition	Model test	Gender	H_type	Children	Work_S	Car_O	Density	dist_I	dist_P	Entering_T	Duration_E
AEP1	S-Trip	Cox	0.01	0.08		NA		0.01	0.03		0.01	0.01
		Additive	0.01			NA		0.01			0.01	0.01
	M-Trip	Cox		0.05	0.09						0.01	0.01
		Additive	0.01			0.01					0.01	0.01
	D-Trip	Cox									0.01	0.01
		Additive		0.01		0.07	0.01	0.01	0.01		0.01	0.01
AEP2	S-Trip	Cox			0.09	NA		0.01	0.04		0.01	0.01
		Additive	0.01		0.01	NA					0.01	0.01
	M-Trip	Cox		0.08	0.01	0.01	0.01			0.02	0.01	0.01
		Additive		0.06	0.02	0.01	0.07				0.01	0.01
	D-Trip	Cox				0.01	0.01				0.01	0.02
		Additive	0.07		0.06	0.01	0.04				0.01	0.01
AEP3	S-Trip	Cox	0.06	0.05	0.01	NA					0.01	0.01
		Additive	0.01	0.03	0.01	NA		0.08			0.01	0.01
	M-Trip	Cox				0.04		0.01	0.04		0.01	0.01
		Additive			0.08	0.05		0.01			0.01	0.01
	D-Trip	Cox			0.01	0.01	0.04	0.02			0.01	0.01
		Additive			0.01	0.01	0.05		0.07		0.01	0.01
AEP4	S-Trip	Cox			0.01	NA	0.06		0.06		0.01	0.01
		Additive	0.01		0.04	NA				0.01	0.01	0.01
	M-Trip	Cox								0.01	0.01	0.01
		Additive			0.08	0.01			0.05		0.04	0.01
	D-Trip	Cox			0.02	0.01					0.01	0.03
		Additive				0.01					0.01	0.01

Remark: 1. AEP: non-travel activity episode
2. S: subsistence activity, M : maintenance activity, D : discretionary activity
3. The test of the significance of covariate-specific effects of Cox model and Aalen's additive model is based on Wald test and Scheike's supremum test, respectively.

Table 4 Covariates parameter estimates of Cox model for state transition hazard estimates

Episode	State transition	Gender	H_type	Children	Work_S	Car_O	Density	dist_I	dist_P	Entering_T	Duration_E
EP1	S-trip	-0.34	-0.08		NA		0.01	0.004		-0.71	0.02
	M-trip		-0.10	0.11						-0.16	-0.02
	D-trip									-0.20	-0.01
EP2	S-trip			-0.14	NA		0.02	0.01		-0.56	0.02
	M-trip		0.14	-0.24	0.18	-0.26			0.005	-0.07	-0.01
	D-trip			-0.12	0.41	0.34		0.01		-0.19	-0.01
EP3	S-trip	0.12	0.15	-0.52	NA					-0.33	0.01
	M-trip	0.09			0.13		0.01	0.01		-0.10	-0.02
	D-trip			0.24	0.39	0.19	0.01			-0.16	-0.01
EP4	S-trip			-0.41	NA	-0.58		-0.01		-0.40	0.02
	M-trip		-0.16		0.15				0.01	-0.05	-0.02
	D-trip			0.27	0.32					0.17	-0.01

Remark: NA means this covariate is not included in the model specification

Table 5 Test for time-varying effect of covariates in the additive models for each of activity types over episodes

Episode	State transition	Intercept	Gender	H_type	Children	Work_S	Car_O	Density	dist_I	dist_P	Entering_T	Duration_EPI
AEP1	S-Trip	0.01	0.06			N/A	0.02				0.01	0.09
	M-Trip	0.01			0.01			0.07			0.01	0.01
	D-Trip			0.01		0.08	0.01	0.01	0.05	0.01		
AEP2	S-Trip	0.01			0.01	N/A	0.03				0.01	0.01
	M-Trip	0.01	0.01		0.04	0.04				0.02	0.01	
	D-Trip	0.01		0.08				0.03		0.01	0.07	
AEP3	S-Trip	0.01	0.04			N/A	0.01	0.01			0.01	0.01
	M-Trip			0.07	0.01		0.02		0.06	0.07		0.01
	D-Trip		0.02	0.03		0.01			0.04	0.01		0.01
AEP4	S-Trip	0.01				N/A	0.01				0.06	
	M-Trip	0.01					0.04	0.01	0.01		0.01	0.01
	D-Trip	0.01	0.10			0.07	0.01		0.04	0.04	0.01	

Remark: 1. S: subsistence activity, M : maintenance activity, D : discretionary activity
2. p-values are reported only for significance at 0.1 level

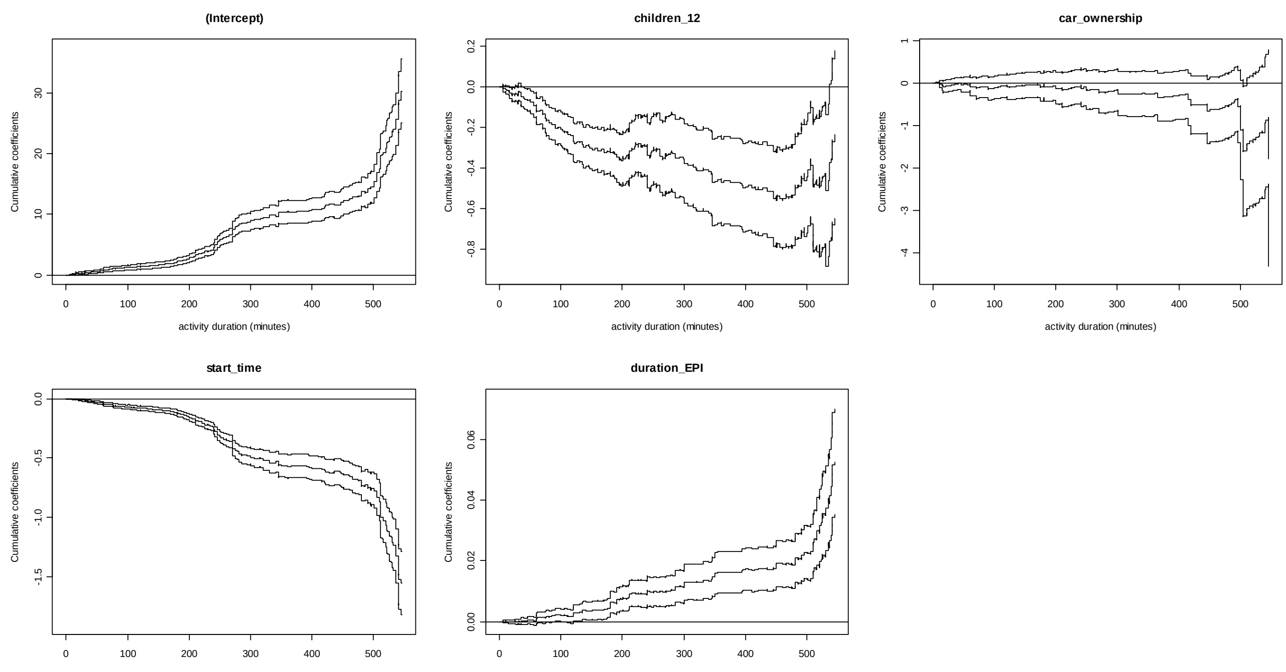

Fig. 2 Additive model: estimate of the cumulative effect of covariates and a 95% pointwise confidence interval for subsistence activity durations for AEP2 (the first non-travel activity episode)

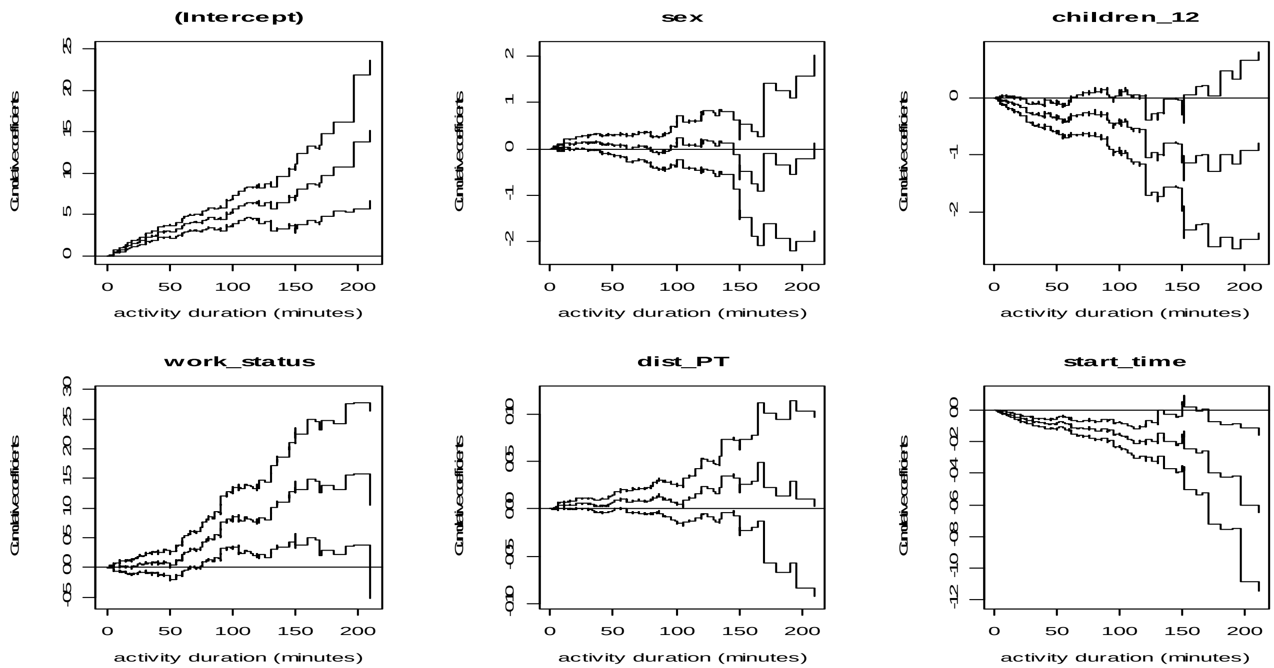

Fig. 3 Additive model: estimate of the cumulative effect of covariates and a 95% pointwise confidence interval for maintenance activity durations for AEP2

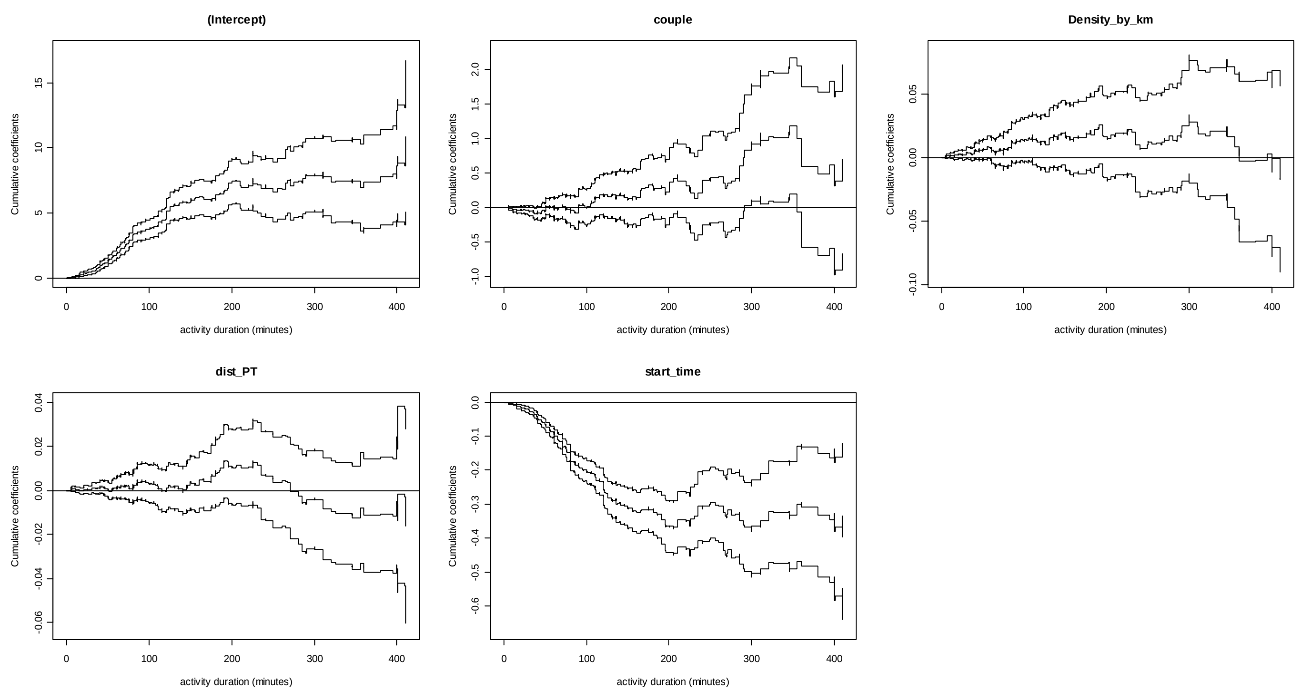

Fig. 4 Additive model: estimate of the cumulative effect of covariates and a 95% pointwise confidence interval for discretionary activity durations for AEP2

5. Conclusion

In this study, a multistate competing risk model is proposed to investigate the effects of covariates on the durations of travel/activity conducted in different episodes of an activity chain. The state transition hazards are estimated for the first four episodes of activity chain based on Cox proportional hazard model and Aalen's additive model. The test statistics are provided to verify the additive effects and the time-varying effects of Aalen's additive model. We compare the results of parameter estimates of the two models and give detailed interpretation for the effects of covariates for one episode.

The results of the empirical study can be summarized as follows. Firstly, we found that the effects of covariates on the durations of activities vary with respect to activity type and episode, but the starting time of activity has negative influence on almost all types of activities and all episodes. Interesting, we find that its influence is slighter on the durations of maintenance activity compared with subsistence and discretionary activity. The duration of trip previously conducted has also significant effects for almost all activities and all episode, but it reveals slight positive effects for subsistence activity and slight negative effects for maintenance and discretionary activities.

Secondly, we found similar results of the effects of covariates on the durations of activity based on Cox model and Aalen's additive model. The comparison shows that the estimation results of the two models coincide mutually in not only the positive/negative effects but also in the magnitude of influence on the durations of activity. As for the interpretation, we argue that Cox model provides useful information to investigate the hazard ratio between different values of covariates settings. However, Aalen's model removes the proportionality assumption of Cox model and provides an easier way to investigate the time-varying effects of covariates on the durations of activity. In practice, the fitting of the two models provide more comprehensive knowledge about the effects of covariates and it seems to be appealing to apply Aalen's model to investigate the time-varying effects, which are usually neglected in activity duration analysis.

Acknowledgements

This research has benefited from a grant of ANR (the French Agency for Research). Project EuroCities-DATTA n° ANR-07-BLAN-0032-01

Reference

1. Aalen, O.O. 1980. A model for nonparametric regression analysis of counting process. In: Klonecki. N. Kosek. A. Rosinski. J. (Eds.). *Survival Analysis: State of the Art*. 2nd Ed. Springer. New York. 1–25.
2. Aalen, O.O. 1989. A linear regression model for the analysis of lifetimes. *Statistics in. Medicine* 8 907-925.

3. Aalen, O.O. 1993. Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine* 12 1569-1588.

4. Amato, D.A. 1988. A generalized Kaplan-Meier estimator for heterogeneous populations. *Communications in statistics. Theory and methods* 17 263-286.

5. Arjas, E. 1988. A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association* 83 (401) 204-212.

6. Bhat, C.R. 1996. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. *Transportation Research Part B* 30(6) 465-480.

7. Bhat, C.R., Misra, R., 1999. Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. *Transportation* 26 (2), 193–209.

8. Cox, D.R. 1972. Regression Models and Life Tables. *Journal of the Royal Statistical Society B* 34 187-200.

9. Efron, B. 1977. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* 72 557-565.

10. Ettema, D., Borgers, A. Timmermans, H.J.P. 1995. Competing risk hazard model of activity choice, timing, sequencing and duration. In: *Transportation Research Record: Journal of the Transportation Research Board. No. 1493. Transportation Research Board of the National Academies*. Washington. D.C. pp.101-109.

11. Hosmer, D.W., Lemeshow, S. 1998. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley-Blackwell.

12. Kalbfleisch, J.D., Prentice, R.L. 2002. The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons. Inc. 2nd Edition.

13. Klein, J.P. Moeschberger, M.L. 2003. *Survival analysis—techniques for censored and truncated data*. 2nd edition. Springer. New York.

14. Klein, J.P. 2006. Modelling competing risks in cancer studies. *Statistics in Medicine* 25 1015-1034.

15. Ma, T-Y. Raux. C., Cornelis, E., Joly, I. 2009. Multi-state non-homogeneous semi-markov model of daily activity type. timing and duration sequence. *Proceedings of The Transportation Research Board (TRB) 88th Annual Meeting*. Washington. D.C. USA.

16. Martinussen, T., Scheike, T.H. 2006. *Dynamic Regression Models for Survival Data.* Springer.

17. Pendyala, R.M. Bhat C.R. 2004. An exploration of the relationship between timing and duration of maintenance activities. *Transportation* 31 429-456.

18. Popkowski Leszczyc, P.T.L. Timmermans, H. 2002. Unconditional and conditional competing risk models of activity duration and activity sequencing decisions: An empirical comparison. *Journal of Geographical Systems*. Vol. 4. pp.157-170.

19. R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

20. Scheike, T. 2002. The additive non-parametric and semi-parametric Aalen model as the rate function for a counting process. *Lifetime Data Anal* 8 247–262.

21. Scheike, T.H. 2004. Time-Varying Effects in Survival Analysis. In: Balakrishnan, N., Rao, C.R. (Eds.), *Handbook of Statistics Vol. 23: Advances in Survival Analysis*. Elsevier.

22. Srinivasan, K.K. Guo. Z. 2007. Analysis of trip and stop duration for shopping activities: Simultaneous hazard duration model system. *Transportation Research Record: Journal of the Transportation Research Board*. No 1854. Transportation Research Board of the National Academies. Washington. D.C.

23. Therneau, T.M., Grambsch, P.M., Fleming, T.R. 1990. Martingale-based residuals for survival models, *Biometrika* 77 147–160.