



MULTI-STATE NON-HOMOGENEOUS SEMI-MARKOV MODEL OF DAILY ACTIVITY TYPE, TIMING AND DURATION SEQUENCE

Tai-Yu Ma, Charles Raux, Eric Cornelis, Irigaël Joly

► To cite this version:

Tai-Yu Ma, Charles Raux, Eric Cornelis, Irigaël Joly. MULTI-STATE NON-HOMOGENEOUS SEMI-MARKOV MODEL OF DAILY ACTIVITY TYPE, TIMING AND DURATION SEQUENCE. 2008. halshs-00310900v1

HAL Id: halshs-00310900

<https://shs.hal.science/halshs-00310900v1>

Preprint submitted on 11 Aug 2008 (v1), last revised 16 Nov 2008 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTI-STATE NON-HOMOGENEOUS SEMI-MARKOV MODEL OF DAILY ACTIVITY TYPE, TIMING AND DURATION SEQUENCE

Tai-Yu Ma*

Transport economics laboratory (LET)

Fax (33) 4-72-72-64-48, Phone (33) 4-72-72-64-46

E-mail: tai-yu.ma@let.ish-lyon.cnrs.fr

14, Avenue Berthelot

F-69363 Lyon, France

Charles Raux,

Transport economics laboratory (LET)

Fax (33) 4-72-72-64-48, Phone (33) 4-72-72-64-54

E-mail: Charles.Raux@let.ish-lyon.cnrs.fr

14, Avenue Berthelot

F-69363 Lyon, France

Eric Cornelis

Groupe de recherche sur les transports, FUNDP

Fax: (81) 72-49-14, Phone (81) 72-49-22

E-mail: eric.cornelis@fundp.ac.be

Rue de Bruxelles 61

B-5000 Namur, Belgique

Iragael JOLY

Laboratoire d'Economie Appliquée de Grenoble

INPG - ENSGI – CUEFA

Fax (33) 4-76-82-54-55 Phone (33) 4-76-82-83-36

Email: Iragael.Joly@inpg.fr

Université Pierre Mendès France

BP47 38040 Grenoble, France

*Corresponding author

Submission date: 1 August 2008

Number of words : 4927 words + 5 Tables/Figures = 6177 words

ABSTRACT

Understanding travelers' daily travel-activity pattern formation is an important issue for activity-based travel demand analysis. The activity pattern formation concerns complex interrelations between household members and individual's socio-demographic characteristics, which is in general difficult to quantify its influence. To investigate the interrelationship between travel-activity pattern formation and related covariates, we apply a non-homogeneous multi-state semi-Markov model to analyze the timing and duration of daily travel and activity sequence. We consider the observed travel-activity chains as a result of a semi-Markov process with a finite state space, i.e. the sequence of travel-activity depends only on its adjoining states. To estimate the effect of covariates related to individual and household characteristics, a Cox regression model is adopted in a semi-Markov process. The statistical tests of significance suggest that the duration of activity depends not only on its beginning time-of-day but also on the duration of travel/activity in prior episode. Empirical estimates of state transition hazard based on Belgian Mobility survey suggest useful insight on the relationship between the duration of activity and related socio-demographic characteristics.

INTRODUCTION

The daily travel and activity patterns concern multidimensional facets encompassing activity type, duration of activities and the location of activities, etc. The formation of a sequence of travels and activities is general heuristical, based on individual schedule/reschedule process given an uncertain environment. The understanding of the complex travel-activity decision has important implications for the evaluation of transport policy and traffic demand management. Hence, it is interesting and also important to investigate individuals' travel-activity chaining behavior with respect to related household characteristics. The results of analysis can provide useful insights for travel and activity duration estimations and also for individuals' daily travel activity participation.

The early studies for activity patterns have focused on descriptive analysis of activity pattern characteristics and influence factors for its formation (1,2). The analysis framework emphasized on the concept that travel demand is derived from the need of activity participation. As numerous influence factors impose dynamical spatial-temporal constraints on activity pattern formation, it is generally difficult to investigate the relationships between these factors and the resulting travel-activity patterns. The progress in activity-based demand analysis has developed many modeling techniques trying to elucidate the process of activity and travel decisions. In general, these analytical approaches can be classified into: (a) sequential approaches, which decompose individuals' travel-activity choices into a sequential decision-making process and apply microsimulation techniques to entail the estimation of travel-activity patterns (see recent review in (3)); (b) simultaneous approaches, which aim to estimate simultaneously the entire activity patterns by utilizing econometrics or mathematical programming methods (4,5,6,7). An alternative class of modeling approaches is based on hazard models for the empirical analysis of activity duration and sequencing (8,9,10). These studies focused especially on the estimation of activity duration hazard function with the presence of individual's heterogeneity. As for the interdependency of durations between multiple activity types, Bhat (11) proposed an outcome-specific proportional hazard model, which generalizes the estimation of duration hazard function with multi-entrance/exit states. Another activity hazard-based study, proposed by Srinivasan and Zhiyong (12), encompassed the correlation of simultaneous duration process. This model utilized a mixed distribution combining baseline hazard function with a random error term to capture the unobserved correlation between two duration processes. This study provided empirical evidence of dependency on the durations of trips and activities. However, the considerations of the influence of activity sojourn time in hazard function specification need to be further investigated.

The time dependency effect plays an important role in the estimation of activity duration since it reflects the transition probability after certain duration from current activity to another. Recently, some research effort has been made to investigate the relationship of timing and duration of activity. Pendyala and Bhat (13) applied discrete-continuous simultaneous equation model to investigate the casual 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 (14) utilize conditional and unconditional parametric competing risk models to estimate activity duration and its relationship with socio-demographic covariates. The estimation results showed that the activity duration depends on the type and duration of the activity previous conducted. Parallel to the competing risk models for the estimation

the timing and duration of activity, Lee et al. (15) proposed a generalized multi-state hidden Markov extension model to analyze the duration of type-specific activities. Although the proposed model provided a flexible framework to capture the misclassification effect of observed activities, it did not incorporate the dependency effect of the duration of activity previously conducted. Recently, the advance in survival analysis has developed multi-state models to capture simultaneously several entrances/exits states over the duration process (see review in (16)). The multi-state models analyze event history data of individuals, focusing especially on the occurrence time, duration of type-specific events following a Markov process. The multi-state models consider event occurrence history based on Markov-type stochastic process, i.e. step transition probability depends only on the adjoining states. At each time instant, an individual occupied in one state possibly transfers to another state with respect to time dependent transition rates. In general, the multi-state model is based on the continuous-time Markov process for which time dependency effect on transition hazard function can be analyzed by incorporating convenient time-dependent covariates. As the activity duration depends on its type, timing and sojourn time prior to current state, it is interesting to investigate the influence of these factors on individual's activity chaining behavior. Such analysis of the duration dependency of a sequence of discrete events visited by Markov process is generally called *Markov renewal process*. The reader is referred to (17,18) for more detail descriptions and applications.

In this study, a non-homogeneous multistate semi-Markov model is developed to estimate the daily activity patterns by encompassing the interdependency of sequential activity types, timing and duration. The estimation of the transition hazard provides useful information for competing activity choices at the end of each activity. Moreover, the transition hazard rates between different states are estimated with respect to transition-specific time-dependent covariates. These estimated coefficients provide useful insight to investigate the influence of individual and household characteristics on the duration of activity.

The remainder of this paper is organized as follows. Firstly, a multi-state semi-Markov model is proposed to analyze activity-sequencing dependency on the basis of activity chaining. In section 2, we define the state structure of activity types and specifying its transition hazard functions based on Cox relative risk model (19). It follows with the hazard function specifications associated with related covariates under the Markov renewal process. To investigate the time dependency effect of activity duration with respect to its beginning time-of-day and the duration of prior activity conducted, the Cox model is applied to investigate the effect of time-dependent covariates on multiplicative activity transition hazard. The likelihood function and its parameter estimation method are also illustrated. The summarized statistics about the frequencies of activities conducted on a sequence of individual's travel-activity pattern and related socio-demographic covariates are presented in section 3. In section 4, the estimation results of activity transition hazard function and its relationship with socio-demographic covariates are presented. Finally, we summarized this study and discuss some future extensions.

MULTI-STATE NON-HOMOGENEOUS SEMI-MARKOV MODEL

Consider a *travel-activity sequence* of an individual observed over a whole day for which the types of activity, timing and duration are of interest. Assuming that the evolution of individual's activity states in time is resulted from a semi-Markov process with finite state

space, representing an individual's travels/activities, denoted as $\{a(t), 0 \leq t \leq T\}$, $\forall a(t) \in A$, with T being the ending time of a day (24 hours) and A the finite state space of one transition, assumed uniform for all transitions and all individuals. This sequence of events corresponding to individual's travel/activity occurrences over a whole day is denoted as (see Fig. 1):

$$q = (a^{(n)}, t^{(n)}), n \in \{0, 1, 2, \dots, N\} \quad (1)$$

where

$a^{(n)}$: type of the n^{th} travel/activity, $\forall a^{(n)} \in A$

$t^{(n)}$: beginning (transition) time of the n^{th} travel/activity, with $0 \leq t^{(0)} < t^{(1)} < \dots < t^{(n)} \leq T$.

The initial state of a travel-activity chain is set as $t^{(0)} = 0$, $a^{(0)} = H$, being home-based activity H .

The duration of the $n-1^{\text{th}}$ travel/activity $\tau^{(n)}$ is defined as $t^{(n)} - t^{(n-1)}$. Individuals' occupied state history on the process until time t is denoted as:

$$H_t = \{a(s), 0 \leq s \leq t\} \quad (2)$$

At each transition step, it represents the occurrence of a change of travel/activity type. If the observations are based on the semi-Markov process, the transition rate from state i to state j depends only through current state i and the sojourn time until t . We call the sojourn time until t since entering current state τ as *renewal time*. The transition rate from state i to state j at the sojourn times τ $\lambda_{ij}(\tau)$ is defined:

$$\lambda_{ij}(\tau) = \lim_{h \rightarrow 0} \frac{P[a(t^- + h) = j \mid a(t^-) = i]}{h} \quad (3)$$

with τ is the sojourn time at state i until t since the individual entered occupied state i at time t^- and $a(t^-)$ is the state at time t^- . The semi-Markov model assumes that the duration of one event is dependent only on its adjoining states. As our interest is to estimate type-specific activity survival functions and investigate the influence of related covariates, the semi-parametric regression model is preferred. In the literature, Cox proportional hazard model is very widespread for many applications. The advantage of Cox model resides on its flexibility since it does not need to specify the underlying survival distribution and can incorporate the covariates of interest. The incorporation of Cox model with semi-Markov model for the model parameter estimation is generally called modulated Markov renewal process (20,21). The one-step baseline transition hazard is specified as function of the sojourn times τ . However, this specification of time-dependent baseline hazard function is more complicate to estimate since Aslen's multiplicative intensity models cannot be applied by incorporation of the sojourn duration (17). To capture the dependency effect of activity timing and duration and avoid the estimation of the renewal time-dependent baseline hazard function, we include related time-dependent covariates in the Cox model to capture the time-dependency effect.

The Cox model with respect to transition from state i to j associated with time-dependent covariates of individual l \mathbf{X}_l during the process can be defined as

$$\lambda_{ijl}(\tau) = \lambda_{0,ij} \exp(\mathbf{X}_l'(\tau) \boldsymbol{\beta}_{ij}), \quad \forall i \neq j \quad (4)$$

where

$\lambda_{0,ij}$: unspecific baseline hazard function with respect to transition (i, j) , which can be obtained by empirical Nelson-Aalen estimator of cumulative hazard function
 $\mathbf{X}_l(\tau)$: column vector of time-dependent covariates of individual l
 $\boldsymbol{\beta}_{ij}$: column vector of transition-specific regression coefficients with respect to (i, j) .

Note that a class of models can be tested by the specification of covariates. For instance, one can investigate the time effect on hazard rate variation or test the interaction effect between covariates. As we aim to investigate the time dependency effect over activity duration, three models are specified:

- Model 1: including all individual's socio-demographic covariates and then selected based on stepwise regression method.
 Model 2: extending model 1 by incorporating entering time of current travel/activity episode.
 Model 3: extending model 2 by adding the duration of previous travel/activity episode.

The ratio of two hazard rates represents the relative risk on the basis of baseline hazard function. If regression coefficients are positive (negative), the relative risk is increasing (decreasing), i.e. activity duration is decreasing/increasing. Besides, one can also incorporate a discrete choice model to analyze the outcome state choice (11) or investigate the effect of individual's heterogeneity in Cox model (9,10).

The survival function for transition (i, j) at renewal time τ can be derived as:

$$S_{ijl}(\tau) = S_{0,ij}^{\exp(\mathbf{X}_l'(\tau)\boldsymbol{\beta}_{ij})} \quad (5)$$

where $S_{0,ij}$ is the baseline survival function for transition from state i to j .

The estimation of parameter vector $\boldsymbol{\beta}_{ij}$ for all $i, j, \forall i \neq j$ can be obtained by the maximization of partial log likelihood as:

$$\log L(\boldsymbol{\beta}_{ij}) = \sum_{l=1}^L \int_0^T \mathbf{X}_l \boldsymbol{\beta}_{ij} dN_{ijl}(\tau) - \int_0^T \log \left\{ \sum_{m=1}^L Y_{im}(\tau) \exp(\mathbf{X}_m'(\tau)\boldsymbol{\beta}_{ij}) \right\} dN_{ij.}(\tau) \quad (6)$$

where

$N_{ijl}(\tau)$: observed number of transitions from i to j for individual l within time interval $[t_{il}, t_{il} + \tau]$. t_{il} is the entering time of current state of individual l .

$N_{ij.}(\tau) = \sum_{m=1}^L N_{ijm}(\tau)$ representing the observed number of individuals at state i at risk within time interval $[t_{il}, t_{il} + \tau]$

$Y_{im}(\tau)$: indicator being 1 if individual m is at state i under risk within $[t_{il}, t_{il} + \tau]$; 0 otherwise.

L : total number of individuals

Conventional Newton-Raphson technique can be applied to derive $\boldsymbol{\beta}_{ij}$ and generally its convergence can be obtained quickly. Related statistical test methods are also available to

investigate its proportional hazard assumption with respect to each covariate (22).

The estimation of transition probability function and cumulative hazard function is related to the counting process with respect to the observed number of transitions from state i to state j . The estimates of empirical baseline cumulative hazard function $\hat{\Lambda}_{0ij}(t)$ can be obtained by Nelson-Aalen estimator, which is right-continuous step function.

$$\hat{\Lambda}_{0ij}(\tau) = \int_0^\tau \sum_{l=1}^L \frac{Y_{ijl}}{\sum_{m \in Y_i(\tau)} \exp(\mathbf{X}'_m(\tau) \boldsymbol{\beta}_{ij})} d\tau \quad (7)$$

where

$\hat{\Lambda}_{0ij}(\tau)$: the estimate of the cumulative baseline hazard of transition (i, j) for renewal time τ .

$Y_{ijl}(\tau)$: indicator being 1 if individual l is at state i under risk within $[t_{il}, t_{il} + \tau]$; 0 otherwise.

Note that the estimates of cumulative baseline hazard function are carried out with respect to strata of the same covariates. Its plots provide useful insights to check its appropriateness of such stratification.

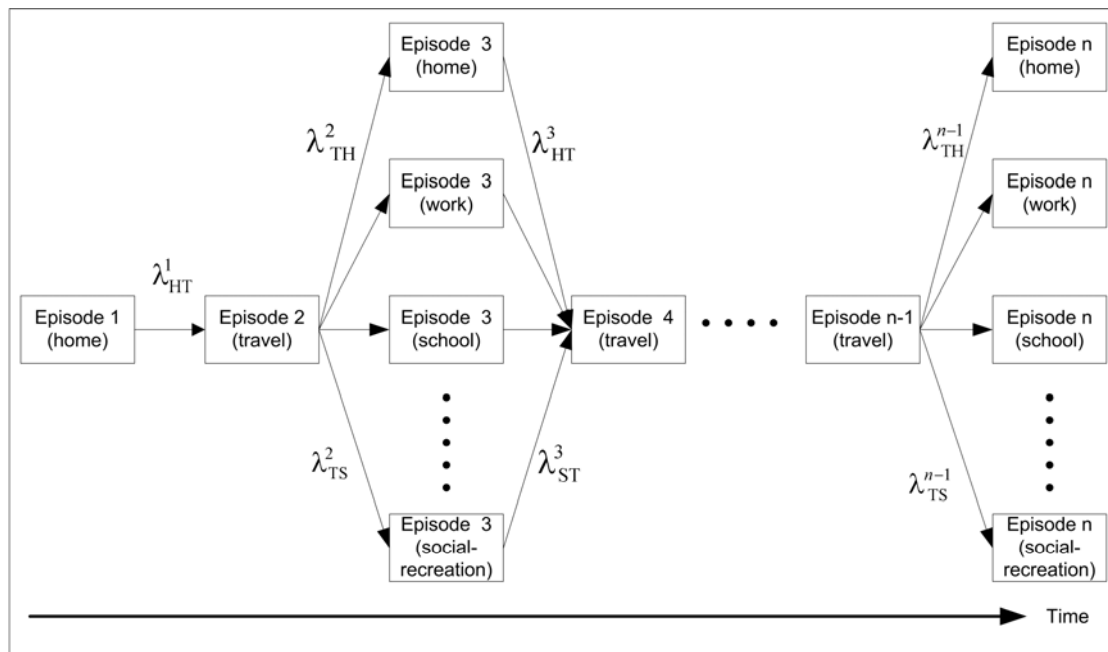


FIGURE 1 Non-homogeneous semi-Markov process for daily travel-activity pattern formulation

DATA DESCRIPTION

The data utilized is based on the Belgian Mobility Survey collected in 1998-1999 for investigation of household members' mobility behavior. 7025 individual's daily travel-activity patterns are investigated. As the analysis of interest resides on the investigation of travel-activity chaining, our focus is on the sequence of travels and out-of-home activities

realized on a day. Hence, samples without any trip occurred are not included in the analysis data set. After investigating the different activity type participations over weekday and weekend, we found that these two activity type choice patterns are quite different. The former is work/school oriented. However, the latter is shopping and social-recreation dominated. This study focuses only on travel-activity pattern over workday. One can, however, extend similar analysis for data on weekend. It is also interesting to investigate the influence of covariates for different types of travel and activity transitions. Based on previous studies (Lee et al. 2007), some individuals (sex, age and profession status) and household socio-demographic characteristics (household income, household type, and presence of children) are included in the model specification. Besides, spatial and modal related variables (household location, transport mode availability and proximity to public transport) are also included in the model specification. The summary statistics of these covariates is listed in Table 1.

As for the travel-activity sequence data, individuals are asked to report his/her sequence of trips and its related trip motivation on the previous day. For each reported travel/activity episode, trip departure/arrival time information is utilized to calculate trip duration. Original trip motivation is distinguished into 12 types for which a regroupment in 6 types is undertaken to facilitate the analysis (see Table 2). For simplification, we assume that the activity within two consecutive trip episodes is homogeneous. As a result, the activity duration is calculated based on the arrival/departure time of its prior/posterior trip. As the model estimation requires the full information for covariates and travel and activity related variables, a complete data set is obtained by eliminating observations containing missing or incorrect data. The resulted data set contains 2849 observations and its related descriptive statistics is listed in Table1. For activity chaining behavior, the residual observations at each activity episode is shown in Table2. It is clear that home-work-home constructs the main travel-activity pattern on workdays and the number of activity participations is less than 7 for 80% observations.

TABLE 1 Summary statistics of individual's socioeconomic covariates (n=2849)

Variable	Definition	Means	Std. dev.
<i>Individual characteristics</i>			
Sex	Sex (1 if male, 0 female)	0.5117	0.4999
Age	Age of individual	37.4176	19.6528
Prof	Employment status (1 have a job, 0 otherwise)	0.4689	0.4991
<i>Household characteristics</i>			
H_type	Household type (1 if couple, 0 if single)	0.1712	0.3768
H_location	Location of household (1 center city/village, 0 otherwise)	0.6093	0.4879
N_children	Number of young children less than 16 years of age in the household	0.8961	1.2543
N_car	Number of cars in the household	1.3134	0.7751
PT_proximity	Proximity to public transport (1 less than 250m, 2 250m ≤ distance < 500, 3 500m ≤ distance < 1km, 4 1km ≤ distance < 2km, 5 km ≤ distance < 5km, 6 more than 5km)	1.9589	1.1240

TABLE 2 Number of observations in sequential non-travel activity episodes

Type of activity	AEP1	AEP2	AEP3	AEP4	AEP5	AEP6	AEP7	AEP8	AEP9	AEP10	AEP11	AEP12	AEP13
Home	2849	40	1707	430	737	257	269	109	99	34	32	9	14
Work	-	927	260	239	108	69	38	20	17	6	4	3	1
School	-	645	28	103	8	12	5	5	0	0	0	1	0
Shopping	-	484	242	195	99	90	28	29	9	7	2	3	2
Personal business	-	442	237	259	132	109	59	52	18	16	5	6	1
Social-recreation	-	304	231	305	124	121	65	43	17	15	3	8	2
Other	-	7	8	3	1	3	0	0	1	1	0	0	0
Total	2849	2849	2713	1534	1209	661	464	258	161	79	46	30	20

* AEP : non-travel activity episode

In the following, we estimate activity duration hazard functions and related survival functions for state transition based on Cox model. The estimation process consists in establishing a class of models with covariates of interests and then selecting one model based on model fit statistics. The stepwise regression method is utilized to select predictive covariates for which all candidates are included at the beginning and then tested one-by-one and eliminated if non significant according to its statistical significance

For covariates selection, 5% significance level is utilized for all classes of models in this analysis. Moreover, as the number of travel-activity episodes is large, a reduced analysis focused on the previous episodes, accounting for at least 58% observations, is conducted.

ESTIMATION RESULTS

The parameter estimation is conducted by the Proc Phreg in SAS. The model fit statistics is shown at Table 3. We find that the model 3 incorporating logarithm of the entering time of current travel/activity and logarithm of the duration of travel/activity conducted in previous episode performs best over the other two models in most episode transitions. The results show that the duration of travel/activity closely related to its type, beginning time-of-day and the duration of prior activity conducted, accordant with the empirical estimation results of activity duration of Popkowski Leszczyc and Timmermans (14). The coefficient estimates of individual and household characteristics are shown in Table 4. As mentioned previously, a positive coefficient on a covariate increases the hazard rate, i.e. decrease the duration of travel/activity.

As the number of travel/activity episodes is numerous, we focus here only on the analysis of the duration of activities conducted in different episodes. For the first episode, it shows that the duration of home-based activity decreases with men, workers, and households with young children less than 16 years of age but it increases with the household type of couple and household location at center city/village. The estimation results are quite reasonable. For the third episode, it shows that the duration of home activity decreases with household location at center city/village. It implies that individuals living in center city tend to conduct another out-of-home activities at early episode. The time-of-day decreases the duration of home-based activity, work, school and social-recreation activities in this episode. However, the time-of-day effect on the duration of activities of other episodes is variant and depends on its type and the chronological time-of-day. For the influence of the duration of previous travel/activity conducted, the estimation results show that the duration of current activity increases if one spends more time of travel to conduct current activity. This implies that people spend more time on

travel to engage an activity with longer duration.

As for work activity, it shows that men have longer duration. However, people perform shorter duration of work if they are older, living in city center, with young children in the household and with public transport proximity. For shopping activity, it shows that workers perform shorter duration but couple with young children conduct longer duration of shopping. For personal business, the results imply that young people and people, living near public transport station, conduct longer duration of personal business. However, workers, people of the location of household at city center, with car and children, conduct shorter duration of personal business. As for social-recreation activity, the results suggest that workers and people with location of household at city center spend less time but households of couple spend more time on social-recreation activity. Finally, it is interesting to note that most transition-specific covariates have variant effect on the duration of activities during different episodes with exceptions of number of car and the duration of previous travel/activity conducted.

TABLE 3 Model fit statistics

Episode	Current state	Next state	-2 Log-likelihood	Test	Likelihood Ratio	DF	P value
EP 1	Home	Trip 1	39123,939			6	--
EP 2	Trip 1	Activity 1	30682,723			48	
			30600,934	2 vs 1	81,789	54	<0.0001
EP 3	1 Home	Trip 2	206,563			2	
			200,954	2 vs 1	5,609	3	0.0101
			200,704	3 vs 2	0,25	3	--
	2 Work	Trip 2	10782,833			5	
			10719,078	2 vs 1	63,755	6	<0.0001
			10702,179	3 vs 2	16,899	8	<0.0001
	3 School	Trip 2	6991,849			5	
			6981,946	2 vs 1	9,903	3	0.0035
			6956,491	3 vs 2	25,455	5	<0.0001
	4 Shopping	Trip 2	4985,217			3	
			4985,217	2 vs 1	0	3	--
			4935,382	3 vs 2	49,835	4	<0.0001
	5 Personal business	Trip 2	4461,838			6	
			4458,627	2 vs 1	3,211	7	0.04470
			4445,904	3 vs 2	12,723	7	--
	6 Social-Recreation	Trip 2	2858,153			3	
			2850,263	2 vs 1	7,89	4	0.00274
			2830,545	3 vs 2	19,718	5	<0.0001
EP 4	Trip 2	Activity 2	31023,215			48	
			30941,254	2 vs 1	81,961	54	<0.0001
			30894,857	3 vs 2	46,397	60	<0.0001
EP 5	1 Home	Trip 3	21823,221			4	
			21553,688	2 vs 1	269,533	4	--
			21546,158	3 vs 2	7,53	5	0.0033
	2 Work	Trip 3	2373,433			3	
			2329,702	2 vs 1	43,731	4	<0.0001
			2327,431	3 vs 2	2,271	5	0.0850
	3 School	Trip 3	124,116			2	
			123,363	2 vs 1	0,753	1	0.3155
			115,091	3 vs 2	8,272	3	0.0079

TABLE 3 Model fit statistics (Continuous)

Episode	Current state	Next state	-2 Log-likelihood	Test	Likelihood Ratio	DF	P value
	4 Shopping	Trip 3	2179,069			3	
			2179,069	2 vs 1	0	3	--
			2147,583	3 vs 2	31,486	3	--
	5 Personal business	Trip 3	2087,718			4	
			2080,969	2 vs 1	6,749	5	0.00525731
			2076,303	3 vs 2	4,666	5	--
	6 Social-Recreation	Trip 3	2039,183			1	
			2039,183	2 vs 1	0	1	--
			2026,901	3 vs 2	12,282	2	<0.0001
EP 6	Trip 3	Activity 3	14281,842			48	
			14233,903	2 vs 1	81,961	54	<0.0001
			14174,128	3 vs 2	46,397	60	<0.0001
EP 7	1 Home	Trip 4	4295,33			5	
			4189,435	2 vs 1	105,895	2	<0.0001
			4189,435	3 vs 2	0	2	--
	2 Work	Trip 4	2132,572			2	
			2115,971	2 vs 1	16,601	4	<0.0001
			2115,971	3 vs 2	0	4	--
	3 School	Trip 4	745,928			2	
			704,971	2 vs 1	40,957	6	<0.0001
			704,971	3 vs 2	0	6	--
	4 Shopping	Trip 4	1641,744			1	
			1641,744	2 vs 1	0	1	--
			1630,417	3 vs 2	11,327	1	--
	5 Personal business	Trip 4	2291,895			4	
			2289,245	2 vs 1	2,65	5	0.06513988
			2285,457	3 vs 2	3,788	6	0.03084262
	6 Social-Recreation	Trip 4	2851,716			3	
			2851,716	2 vs 1	0	3	--
			2839,811	3 vs 2	11,905	4	0.00030054
EP 8	Trip 4	Activity 4	11821,572			48	
			11812,462	2 vs 1	9,11	54	0.08280805
			11743,416	3 vs 2	69,046	60	<0.0001
EP 9	1 Home	Trip 5	8119,097			3	
			7924,799	2 vs 1	194,298	2	<0.0001
			7924,799	3 vs 2	0		--
	2 Work	Trip 5	787,933			2	
			783,447	2 vs 1	4,486	2	--
			783,447	3 vs 2	764,879	3	<0.0001
	3 School	Trip 5	18,568			1	
			10,369	2 vs 1	8,199	2	0.002416242
			10,369	3 vs 2	0	2	--
	4 Shopping	Trip 5	686,955			3	
			686,955	2 vs 1	0	3	--
			678,631	3 vs 2	8,324	2	0.002153825
	5 Personal business	Trip 5	1026,519			1	
			1026,519	2 vs 1	0	1	--
			1026,519	3 vs 2	0	1	--
	6 Social-Recreation	Trip 5	940,273			2	
			940,273	2 vs 1	0	2	--

CONCLUSIONS

This study applies multistate semi-Markov model for daily travel-activity duration estimation. The results of estimates suggest that the duration of activity depends not only on its type but also on the beginning time-of-day and the duration of previous travel/activity conducted. The statistical tests of significance suggest that the incorporation of time-dependent covariates into hazard function is important to estimate the duration of activity. The results of parameter estimations suggests that people spend more time on travel if the duration of activity to participate is longer. Moreover, socio-demographic covariates perform different effects on activity duration during different episode of time. The proposed multistate semi-Markov model allows one to analyze the sequence of travel and activity patterns. It provides also useful insight on the effect of covariates to travel/activity timing and duration analysis. Moreover, it can be used to predict travel-activity patterns or generate activity program of homogeneous socio-demographic classes for activity-based travel demand analysis. The future extensions of current study may incorporate the effect of heterogeneity in hazard function specification.

REFERENCES

1. Chapin, F. S. The Study of Urban Activity Systems. *Urban Land Use Planning*, University of Illinois Press, Urbana, 1965.
2. Hemmens, G. C. Analysis and simulation of urban activity patterns. *Socio-Economic Planning Sciences*, Vol. 4, No. 1, 1970, pp. 53-66.
3. Arentze, T., and H. J. P. Timmermans. A learning-based transportation oriented simulation system. *Transportation Research Part B*, Vol. 38, No. 7, 2004, pp.613-633.
4. Bhat, C. R., S. Srinivasan., and K. W. Axhausen. An analysis of multiple interepisode durations using a unifying multivariate hazard model. *Transportation Research Part B* 39, 2005, pp.797–823.
5. Lee, Y., M., Hickman, and S. Washington. Household type and structure, time-use pattern, and trip-chaining behavior. *Transportation Research Part A*, Vol. 41, 2007, pp.1004–1020.
6. Recker, W. W., M. G. McNally, and G. S. Root. A model of complex travel behavior: Part 1: Theoretical development. *Transportation Research A*, Vol. 20, 1986, pp.307–318.
7. Recker, W. W., M. G. McNally, and G. S. Root. A model of complex travel behavior: Part 2: An operational model. *Transportation Research A*, Vol. 20, 1986, pp.319–330.
8. Ettema, D., A. Borgers, and H.J.P. Timmermans. 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., 1995, pp.101–109.
9. Bhat, C. R. A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. *Transportation Research Part B*, Vol. 30, No. 3, 1996, pp.189–207.
10. Lee, B., and H.J.P. Timmermans. A latent class accelerated hazard model of activity

- episode durations. *Transportation Research Part B*, Vol. 41, No. 4, 2007, pp.426-447.
11. Bhat, C.R. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. *Transportation Research Part B*, Vol. 30, No. 6, 1996, pp. 465–480.
 12. Srinivasan, K. K., and G. Zhiyong. Analysis of trip and stop duration for shopping activities: Simultaneous hazard duration model system. In *Transportation Research Record: Journal of the Transportation Research Board*, No 1854, Transportation Research Board of the National Academies, Washington, D.C., Paper No. 03-4499, 2007.
 13. Pendyala, R. M., and C. R. Bhat. An exploration of the relationship between timing and duration of maintenance activities. *Transportation*, Vol. 31, 2004. pp. 429–456.
 14. Popkowski Leszczyc, P. T. L., and H. Timmermans, Unconditional and conditional competing risk models of activity duration and activity sequencing decisions: An empirical comparison. *Journal of Geographical Systems*, Vol. 4, 2002, pp. 157–170.
 15. Lee, B., H. J. P. Timmermans, Z. Junyi, and A. Fujiwara. A generalized multi-state hidden markov extension model for multi-state activity episode duration. *Proceeding of the 11th international conference on travel behavior research*, Kyoto, 2006.
 16. Hougaard P. Multi-state Models: A Review. *Lifetime Data Analysis*, Vol. 5, No. 3, 1999, pp. 239-264.
 17. Dabrowska, D. M., G. W., Sun, and M. M. Horowitz. Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association*, Vol. 89, No. 427, 1994, pp. 867–877.
 18. Putter, H., J. van der Hage, G. H. de Bock, R. Elgalta, and C.J.H., van de Velde. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal*. Vol. 48, 2006, pp. 366–380.
 19. Cox, D. R. Regression Models and Life Tables. *Journal of the Royal Statistical Society B*, Vol. 34, 1972, pp. 187-200.
 20. Cox, D. R. The statistical analysis of dependencies in point processes. In: *Symposium on Point Processes*, ed. PAW Lewis, New York, John Wiley, 1973.
 21. Kalbfleisch, J. D., and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. New York, John Wiley, 1980.
 22. Kalbfleisch, J. D. and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics, 2002.

TABLE 4 Covariates parameter estimates of multistate semi-Markov model

Episode	Current state	Next state	Sex	Age	Prof	H_type	H_location	N_children	N_Car	PT_proximity	log(t+1)	Log (duration of travel/activity conducted in previous episode)
EP 1	Home	Trip 1	0.15728	-0.01116	0.44004	-0.10035	-0.08038	0.03885				
EP 2	Trip 1	1 Home										
		2 Work	-0.16524			-0.20807					0.56289	
		3 School	0.16167	-0.07994	2.25086				0.16197		2.63689	
		4 Shopping					0.27953	0.09688				
		5 Personal business							0.12817		-0.30164	
		6 Social-Recreation									0.54416	
EP 3	1 Home	Trip 2					0.80145				1.49741	-0.30303
	2 Work	Trip 2	-0.23827	0.00845			0.27487	0.07657	0.10258	0.04669	1.48287	-0.15902
	3 School	Trip 2	-0.12262		0.61514	-0.51865					2.15880	-0.22992
	4 Shopping	Trip 2			0.18045	-0.24806		-0.09735				-0.40401
	5 Personal business	Trip 2		-0.00662	0.22429		0.16377	0.08653	0.16210	-0.08144		-0.21772
	6 Social-Recreation	Trip 2			0.24758	-0.33423	0.28583				0.59542	-0.24370
EP 4	Trip 2	1 Home									-0.50475	-0.18716
	Trip 2	2 Work										
	Trip 2	3 School	0.21348	-0.03789	0.67749	0.28447	0.16901	0.11709	0.23621			-0.32192
	Trip 2	4 Shopping	-0.19752				0.18367	0.07933				
	Trip 2	5 Personal business										
	Trip 2	6 Social-Recreation							0.18067			-0.17553
EP 5	1 Home	Trip 3		-0.00332	0.26485				0.08423		2.53244	-0.07857
	2 Work	Trip 3	0.24977				-0.19845		0.14557		1.77627	-0.11300
	3 School	Trip 3				0.92518					2.07160	-0.62397
	4 Shopping	Trip 3			0.25786	0.39682						-0.44919
	5 Personal business	Trip 3	-0.31422			-0.31344				-0.08418	-0.72340	-0.17792
	6 Social-Recreation	Trip 3			0.34265							-0.23866

TABLE 4 Covariates parameter estimates of multistate semi-Markov model (continue)

Episode	Current state	Next state	Sex	Age	Prof	H_type	H_location	N_children	N_Car	PT_proximity	log(t+1)	log(duration of travel/activity conducted in previous episode)
EP 6	Trip 3	1 Home									-0.82040	-0.23879
	Trip 3	2 Work	-0.18622							-0.07114		-0.28295
	Trip 3	3 School		-0.02036								
	Trip 3	4 Shopping		-0.00617	-0.24164			0.11529				-0.10429
	Trip 3	5 Personal business		-0.00600				0.08428				-0.18567
	Trip 3	6 Social-Recreation					-0.51591				0.96701	
EP 7	1 Home	Trip 4	0.14633								3.85019	
	2 Work	Trip 4			-1.35655			0.09858		0.12244	1.43956	
	3 School	Trip 4	-0.38712	-0.03848	-1.19506		0.50171			0.22072	6.82285	
	4 Shopping	Trip 4										-0.35293
	5 Personal business	Trip 4	-0.28514					0.19155	0.31469	-0.16686	-0.43567	-0.14446
	6 Social-Recreation	Trip 4	-0.25921	-0.00735					0.16109			-0.21401
EP 8	Trip 4	1 Home	-1.03168	-0.04186					1.67869			-0.22575
	Trip 4	2 Work		-0.00930		-0.32675		-0.10478				
	Trip 4	3 School	0.32832	-0.02150							11.54575	-1.73542
	Trip 4	4 Shopping								0.10017		
	Trip 4	5 Personal business										-0.23612
	Trip 4	6 Social-Recreation			-0.43332							-0.26345
EP 9	1 Home	Trip 5			0.21338						5.10553	
	2 Work	Trip 5				-0.55196		-0.21510			1.23155	
	3 School	Trip 5								-1.87511	47.76890	
	4 Shopping	Trip 5										-0.47778
	5 Personal business	Trip 5							0.20990			
	6 Social-Recreation	Trip 5		-0.00806						0.16774		

Remark: t is the entering time of current state