**Activity Duration Analysis: A Discrete-Time Approach for Recurrent Events, Competing Risks, and Multiple States**

Kihong Kim (corresponding author)

Nohad A. Toulan School of Urban Studies and Planning

Portland State University

PO Box 751

Portland, OR 97207-0751

Phone: 971-285-7045

Email: kihong@pdx.edu

Liming Wang, Ph.D.

Nohad A. Toulan School of Urban Studies and Planning

Portland State University

350D Urban Center #751

Portland, OR 97207-0751

Phone: (503) 725-5130

Email: lmwang@pdx.edu

Submitted for

the 94th Annual Meeting of the Transportation Research Board

January 11-15, 2015

Submission Date: August 1st, 2014

Word Count:

ABSTRACT

**INTRODUCTION**

Transportation planners have long developed travel forecasting models to evaluate the impacts of future changes in economics, demographics, land use, and/or transportation on the performance of a regional transportation system. Travel forecasting models can be broadly divided by two components: demand and supply. The demand component estimates travel’s type, amount, timing, location, and mode that collectively represent its demand. Some demand models focus on the travel demand for residents of a region, while other more comprehensive demand models additionally include the travel demand for goods movement as well as special purposes. On the other hand, the supply component measure transportation performance, such as link volume and travel costs (Castiglione, Bradley, and Gliebe, forthcoming).

Both demand and supply components have experienced a significant shift in theory and practice toward a more realistic representation of the model system behavior. Travel demand models have rather slowly evolved from trip-based to activity-based approaches since 1980s, whereas travel supply models have recently rapidly moved from static to dynamic network models. It is important to note that as the two components have been progressed relatively independently, the integration between activity-based demand models and dynamic network supply models becomes more challenging (Lin et al., 2008). Two approaches are available for the integration: sequential integration and dynamic integration (Konduri, 2014). In the sequential integration approach, the demand and supply components are run independently and sequentially in the form of an input-output data flow. In other words, network conditions from the supply model are fed back to the demand model, and the process is repeated until convergence is achieved. The dynamic integration approach adopts an event-based paradigm

In this study, we focus on one aspect of the activity-based travel demand models, namely, activity duration, bearing in mind its dynamic integration with dynamic network models.

Most existing activity-based travel demand models are typically implemented on a tour-based microsimulation framework (Davidson et al., 2007; Vovsha et al., 2005). On the tour-based framework, the basic unit of analysis for modeling travel is not a trip but a tour; in the context of travel demand models, a trip represents a travel unit connecting two locations, while a tour is defined as a chain of trips starting and ending at home (Donnelly et al., 2010). The main advantages of the tour-based structure are to preserve a consistency among multiple trips within a tour in terms of travel mode, destination, and timing. On the microsimulation framework, a full list of households and persons in the synthetic population is simulated during the course of a day. Compared with the “zonal enumeration” approach that is usually used to implement the conventional trip-based four-step models, the microsimulation approach has several advantages. First, the microsimulation may resolve several critical biases that result from demographic, spatial, and temporal aggregations. Second, the microsimulation models are computationally more efficient, virtually allowing an unlimited number of predictors. Third, the microsimulation outcomes look more realistic, being similar to individual activity-travel diaries in travel survey data. Lastly, the microsimulation models are better integrated with the state-of-the-art transport network analysis tools, such as Dynamic Traffic Assignment (DTA) or regional traffic microsimulation models, by providing trip tables or individual trip schedules at a level of compatible temporal resolution (e.g., 30 or 15 minute).

The demand component consists of a series of individual models, which can be either trip-based or activity-based. Trip-based demand models are usually implemented through aggregate zonal enumeration, whereas activity-based demand models typically use a microsimulation technique to simulate individual persons and households in synthetic population.

, visitor travel, and/or goods movement, whereas the supply component measures transportation performance, such as link volume and travel costs (Castiglione, Bradley, and Gliebe, forthcoming). Both model components have experienced a significant shift in theory and practice toward a more realistic representation of the model system behavior. More specifically, since 1980s the travel demand models have rather slowly evolved from trip-based to activity-based approaches. On the other hand, recently the travel supply models, also known as network models, have moves rapidly from static to dynamic approaches. These two domains have been progressed relatively independently; the integration of activity-based demand models and dynamic network supply models is increasingly important.

Pinjari, A. R., & Bhat, C. R. (1970). Activity-Based Travel Demand Analysis. In A. de Palma, R. Lindsey, E. Quinet, & R. Vickerman (Eds.), *Handbook of Transport Economics* (pp. 1–36). Edward Elgar Publishing.

In the context of the tour-based microsimulation framework, there are two different approaches to describing people’s daily decision-making processes on activity and travel (Gliebe and Kim, 2010). The first approach is that people pre-plan the number of tours and trips for a day, set the duration of activities, and calculate the remaining time windows under their space-time constraints. If things do not fit well, they re-schedule the day by adjusting activity timing, activity locations and/or travel modes. This approach is sometimes referred to as “pre-planning” or “rubber-banding” scheduling. The second approach assumes that as the day goes by, people decide what to do next, where to do it, and how to get there. These sequential decisions depend on previous activities, time windows, business hours, intra-household interactions, and so on. This approach is also known as “sequential” or “growing” scheduling.

The pre-planning scheduling microsimulation is a popular choice because of the plausible assumption that individuals’ daily activities are planned in advance, following a fixed hierarchy of activity types. The activity hierarchy typically contains solo mandatory, joint maintenance, joint discretionary, allocated maintenance, and solo discretionary activities, in order from most important to least important. However, there are little empirical evidences for such a rigid structure of activity priorities (Doherty and Mohammadian, 2011). The authors demonstrate that more than 50% of mandatory activities are not planned first in forming home-based tours. Especially, other activity types barely display any particular planning sequences of activities.

On the other hand, the sequential scheduling microsimulation is more flexible because there is no need to pre-determine activity priorities and trip/tour frequencies. Instead, this approach focuses more on how individuals’ activity-travel decisions change over time (Gliebe and Kim, 2010). The authors make a novel proposition that “utility for daily activity-travel alternatives is updated rather than accumulated.” Such a sequential microsimulation model requires to model activity, location, and mode choices that are time dependent. Hazard-based duration models are a promising statistical method to incorporate such time dependencies. For example, the choice of a next activity is one of the important dimensions to be modeled for the sequential microsimulation. The simplest model for next activity choice would be a multinomial logit regression. However, the decision on whether to stay in a current activity or move to a next activity strongly depends on the duration of the current activity. This duration dependence cannot be captured in the static discrete choice model. Instead, a competing-risks hazard-based duration model is proposed to incorporate duration dependence into the activity choice modeling (Ettema et al., 1995).

It is worthwhile to mention that the growing scheduling approach accounts for the activity pre-planning behavior in a different way. Instead of predicting individuals’ number of tours and stops a priori, a very specific household role can be assigned to each household member in the beginning of the simulation day to capture idiosyncratic patterns of pre-planning behavior. The household roles include, for example, working outside home with childcare responsibilities, working outside home while attending college classes, working outside home with planned joint activities with other household adults, and so on (Gliebe and Kim, 2010). Alternatively, sequence alignment methods can be applied to segment people into similar groups in terms of their activity sequences (Kim, forthcoming). Activity diaries are first transformed into sequences of characters representing activity types on fixed time intervals, say, 5-min time intervals. Then, the dissimilarities between all activity sequences are measured through sequence alignment. The resultant pairwise dissimilarity matrix is combined with ANOVA-like analysis to find out significant covariates affecting variations in the activity sequence patterns. An induction tree is also introduced to display how activity sequences vary with the covariates.

Hazard-based duration models deals with time to an event, using a hazard function that represents the conditional probability of an event occurring at a time period, given that the event did not occur before the time period. Duration analysis is often referred to as event history analysis in social science, survival analysis in medical science, and failure time analysis in industrial engineering. There is a wide range of hazard-based duration models. One broad distinction between the models can be made according to whether duration times are measured in continuous or discrete time (Steel, 2005). Most existing duration models in transportation research belong to the continuous-time approach. Bhat (2008) lists the recent applications of the continuous-time duration models to activity participation and scheduling studies.

Although activity duration data are continuous in nature, in this study we rely on the discrete-time approach for several reasons. First, activity diaries are retrospectively collected from household travel surveys. Therefore, respondents are more likely to approximate their activity arrival and departure times to multiples of 5 minutes (e.g., 5, 10, 15, 30, 60 minutes, etc.). Bhat (1996) suggests that activity duration data should be treated as being discrete. Second, because activity durations are discretely measured, the data may contain numerous ties at those discrete time intervals. While discrete-time models can easily handle the tied observations, serious biases can occur in the use of Cox proportional hazards model that is one of the most popular continuous-time models (Steel et al., 1996). In the context of activity-based modeling, such tied observations are common if one focuses on joint activity participations of household members. Third, it is also straightforward to include time-varying covariates into a discrete-time model (Steel, 2005). The status of household members, traffic path information provided by DTA, and transit operating hours are varying over time, which are important variables in advanced activity-based models, even if they are not considered in this study.

A discrete-time model for activity duration can be extended for more complex situations. First, many persons participate in multiple activities for a day. In other words, an event that terminates an origin activity is recurrent to each person. In addition, multiple types of destination activity are competing to end an origin activity. Further, there are multiple states of origin activity. Each origin activity state may have different types of destination activity. The aim of this paper is to introduce discrete-time duration models for recurrent events, competing risks, and multiple states, and to illustrate their application to activity duration data. In the next section we will review the discrete-time duration analysis. Then, we describe data used for this study, together with data transformation required to perform the discrete-time method and with the model development process. Next, the model estimation results will be summarized for both random and fixed effects. In conclusion, some limitations of this study will be presented.

**BACKGROUND**

Since discrete-time hazard-based duration models are rare in transportation research, in this section we briefly overview the methodology based on Steele’s multiple works (2004; 2005; 2008; 2011). For more detailed information, an excellent textbook is available (Singer and Willett, 2003). We begin the overview with a simplest case in which a single non-recurrent event is concerned. Then we add other complexities associated with recurrent events, competing risks, and multiple states in order.

**A Discrete-Time Duration Model for a Single Event**

Suppose that for each episode *i*, we observe duration *yi* accounting for time to a single target event (e.g., leaving a current activity episode). Suppose also that the duration *yi* is measured in discrete time intervals indexed by *t* (*t* = 1, 2, 3, …, *K*), which is either fully observed if the event occurs or right-censored if not . The first step of a discrete-time analysis is to convert the *individual-episode* file to an *individual-episode-period* file; for each time period *t*, we define a binary response *yti* that indicates whether or not the event occurred during the time interval as follows:

In the *individual-episode* file, for example, means that an episode *i* experiences an event during the fourth time interval. In the *individual-episode-period* file, the time to an event or the duration is converted into the four binary responses, namely, . If the episode is right-censored at the same time interval, then the binary responses are coded as .

Now we define a discrete-time hazard for interval *t*, that represents the conditional probability of an event occurring during interval *t*, given that the event did not occur before *t*, as follows:

The next step is to model how the discrete-time hazard function depends on duration and covariates. Note that in the transformed data set (i.e., the *individual-episode-period* file), the dependent variable of interest is binary, indicating the occurrence of an event. A popular solution to analyze binary responses is to perform a logit transformation of the hazard function. As a result, the log odds of the discrete-time hazard is modeled as a linear combination of two sets of predictors, which is given as follows:

where is a function of time period *t* to incorporate the duration effect, namely, the dependence of the hazard function on *t*, which is referred to as the baseline logit-hazard. There are two different ways of specifying the baseline logit-hazard: non-parametric and parametric. The non-parametric specification includes a sequence of temporal dummy variables. Since there are *K* time intervals in the transformed data set, the baseline logit-hazards can be specified with the *K* temporal dummies, . The resultant multiple intercepts represent the baseline logit-hazard for each time period. Although this non-parametric approach is attractive because of its flexibility, there is a practical drawback. In case that the number of time periods in a data set is large, the model needs a substantially large number of dummy variables, which is unwieldy. To be more “parsimonious”, one can parameterize the duration effect. Depending on a plot of the observed hazards over time, a variety of forms of are possible, such as a linear function , a quadratic form or a log function , where represents an overall intercept term.

On the other hand, in Equation (1), is a set of covariates to detect observed heterogeneity in hazard across episodes. The covariates are either time constant or time varying. In discrete-time models, it is straightforward to add time-varying covariates by placing their different values in time periods. In addition, the assumption of proportionality that the effects of covariates are constant over time is common in continuous-time models. In discrete-time models, however, the proportionality assumption can be easily relaxed by including the interactions between *x* and *t* as an additional explanatory variable.

**A Discrete-Time Duration Model for Recurrent Events**

An event may occur one or more times to an individual during a given observation period. For example, in activity survey data, most individuals participate in multiple out-of-home activities during a day. In other words, an individual experiences an event that terminates an activity to carry out another activity several times for a day. In a discrete-time data set with recurrent events, we can define a series of binary response {*ytij*} in which *ytij* indicates whether an event has occurred in time interval *t* for episode *i* of individual *j*. Then, the corresponding discrete-time hazard function can be written as

If recurrent events are observed to an individual, it cannot be assumed that the durations of episodes are independent within the same individual. There may be unobserved heterogeneity, also known as shared frailty, between individuals, which results from time-constant omitted variables. Such unobservables can be captured using multilevel modeling techniques. Note that recurrent events occur in a two-level hierarchical structure where episodes in the first level are nested within individuals in the second level. Hence, in discrete-time analysis, we can add a random intercept term to Equation 1, which is written as

where *uj* is a random effect associated with the *j*th individual. The values of *uj* are typically assumed to follow a normal distribution with zero mean and variance . In such a two-level random intercept model, the log-odds of an event in interval *t* is only shifted by *uj* for the *j*th individual. In a more complex random coefficient model, the coefficients s can be specified to be random across individuals.

**Modeling Competing Risks and Multiple States**

Another common extension of duration models is to take into account more than one kind of event. So far, it is assumed that a single type of event occurs to an individual. However, in many situations, multiple types of event, also referred to as competing risks, may be competing to end an episode. If competing risks arise, the dependent variable in the discrete-time data is no longer binary and it becomes multinomial. Therefore, Equation 2 can be generalized to a multinomial logit model with a random intercept.

Suppose that there are *R* types of event. In competing-risks analysis, a categorical response *ytij* is defined for each interval *t* of episode *i* of individual *j*. If an event of type r occurs in interval *t*, then for *r* = 1, …, *R*, while if no event occurs, . The event-specific discrete-time hazards are defined as follows:

and Equation 2 for the single event analysis becomes

The above random intercept multinomial logit model consists of *R* equations contrasting the risk of an event of type *r* with the risk of no event as the reference category. The form of baseline logit-hazard and the set of covariates may be specified differently for each type of event. In addition, the effects of duration and covariates may vary across event types. Further, the random effect associated with individual *j* may vary by event type, even if the random effects (i.e., ) are assumed to follow a multivariate normal distribution with a covariance matrix.

In competing-risk models, we focus on transitions from one origin state. However, there may be multiple origin states from which multiple types of event are competing to end the origin state. A simple way of handling multiple origin states simultaneously is to add dummy variables indicating which state is occupied during interval *t* to the competing-risk model of Equation 3 as predictor variables. A more general discrete-time duration model for recurrent events, competing risks, and multiple states is described by Steel (2004).

**APPLICATION TO ACTIVITY DURATION ANALYSIS**

The main goal of this study is to develop a hazard-based duration model for daily in-home and out-of-home activity episodes on a discrete-time framework. In this study, an activity episode is defined as a continuous period during which an individual is at risk of experiencing an event that terminates an origin or current activity and then moves to a destination or next activity. The duration (i.e., the time to an event) of activity episode is measured in discrete time intervals. This study takes into account a complex situation where there are recurrent events, competing risks, and multiple states.

The complexity comes from the fact that there are multiple states for the origin activity and multiple kinds of event for the destination activity. An additional complexity is that activity episodes are recurrent within individuals.

the timing of the occurrence of an event.

In this section, we describe data transformation and model structure.

**Sample Data Definition**

In this study we use the 2011 Oregon Travel and Activity Survey (OTAS). The Portland metropolitan portion of the 2011 OTAS records daily in-home and out-of-home activities of about 11,000 persons of nearly 4,800 households. Among them, we select those aged 65 and over who do not work in order to reduce computational burden. Another reason to study this particular group of people is that elderly travel behavior is an interesting research topic among transportation planners as the Baby Boomers just started retiring a few years ago. For simplicity, we also remove persons with any type of censored observations, including left-censored observations (i.e., out-of-home activities in the beginning of the day), right-censored observations (i.e., out-of-home activities in the end of the day), left-right-censored observations (i.e., stay at home or travel all day). As a result, this study examines only retirees who participated in at least one out-of-home activity, starting and ending the day at home. Table 1 compares the frequency and average duration of aggregated activity types between all persons and retirees. Note that in-home activities are subdivided into three levels: home in the beginning of the day, home returning temporarily in the middle of the day, and home in the end of the day. As expected, retirees stayed longer at home than all persons (429.7 vs. 343.5 min.). In addition, retirees’ average duration minute of all out-of-home activities was one half of that of all persons (61.0 vs. 122.5 min.). When excluding subsistence activity types, such as work, work-related, and school activities, however, the average durations for each type of out-of-home activity were similar each other between retirees and all persons.

TABLE 1 Frequency and Average Duration of Activity Episodes for Retirees



**Data Transformation**

The original survey data set is released in an *individual-episode* file format where a two-level hierarchical structure is revealed with activity episodes (level 1) nested within individuals (level 2). The key attributes of activity episodes are obtained from the survey data, including origin activity type, destination activity type, activity duration, activity location, travel mode, and so on. The dependent variable of this study is the duration of activity episode, which is defined as time to an event terminating an origin activity state. Although the duration variable is continuous in nature, in this study it is treated as an interval variable because more than 50% of the duration times are recorded to the nearest multiples of 5 (e.g., 5, 10, 15, 30 min, etc.). On the other hand, the activity diaries involve both the multiple states of origin activity and the competing risks of destination activity. Table 2 shows the transitions of activity episodes from an origin activity to a destination activity. It is important to note that different destination types are allowed for each origin state. For example, it is not allowed to move from an in-home origin (i.e., H1 and H2) to an in-home destination (i.e., H2 and H3). For all the out-of-home origin activity types, it is allowed to move to any type of destination, including the same activity type.

TABLE 2 Transitions of 2,480 Activity Episodes from Origin to Destination



Now we convert the *individual-episode* file to an *individual-episode-period* file, which is illustrated in Figure 1. Initially, the duration times of 2,480 activity episodes of 588 retirees were grouped into 5-min intervals, yielding 82,160 observations in the converted data set, which caused a prohibitive computational burden. Therefore, we decided to reduce the file size by increasing the width of the time interval to 15 minutes, which produced 28,189 observations in the *individual-episode-period* file. Note that broadening the time interval does not change the number of episodes. If there are several episodes within a 15-min interval, all the episodes are retained in the smaller data set because a duration time of one 15-min interval is recorded for each of them. The top panel of Figure 1 illustrates an individual who experienced three events of terminating a current activity episode and moving to a next activity episode for a day: from H1 to RE, from RE to SH, and from SH to H3. The duration of the first activity H1 was 420 minutes, which accounts for 28 intervals in 15-min time slots. This can be said that the first event of ending H1 and transitioning to RE occurred during the 28th 15-min time interval. As shown in the bottom panel of Figure 1, the one record of the event occurrence (rij = RE) with the duration (yij = 28) is converted to a sequence of 28 multinomial responses (y1ij, y2ij, …, y27ij, y28ij) = (0, 0, …, 0, RE). Similarly, the second activity episode ended in the 5th time interval and the third episode terminated during the 1st interval. Thus, each of the activity durations is coded as a series of 5 multinomial responses and a single multinomial response, respectively. In the converted data set, the response of zero represents no event occurrence during that time interval, which will be used as the reference category for the multinomial logit model.



Figure 1 Converting from Individual-Episode to Individual-Episode-Period

**Model Development**

For the discrete-time duration variable, namely, *ytij*, we develop a two-level multinomial logit model using Equation 3, which is re-written here:

The first two terms of Equation 4 form the baseline logit-hazard. Instead of using temporal dummies, we use the parametric specification to make the model more parsimonious. Two parametric forms for the baseline logit-hazard were compared – quadratic and log, and it turned out that the model fit with the log form was better in this study. Both the overall intercept (i.e., ) and the duration effect (i.e., ) are specified to vary by event type *r*.

Dummy variables for origin activity are included to take into account multiple states from which an event occurs. Since some transitions are not allowed, different dummy coding schemes are used for each origin activity state. For the in-home destination types, six dummies for origin state (i.e., EO, ES, OM, RE, SH, and SO) are defined with HC as the reference category. For the out-of-home destinations, we create eight origin state dummies (i.e., H2, EO, ES, HC, OM, RE, SH, and SO), taking H1 as the reference case. As the coefficient of each origin state dummy is allowed to vary across event type *r*, it is expected to capture all the chaining effects from an origin from a destination.

Time-of-day effects on the risk of an event are an important factor. Transitions to a particular type of new activity strongly depend on hours of day. For example, leaving a present activity for eating out usually occurs between 11:00 AM and 1:00 PM for lunch and between 5:00 PM and 7:00 PM for dinner. People tend to participate in social activities at certain times of day, say, in the evening. An easy way is to include dummy variables indicating hours of day. However, given that many state dummies are already specified in the model, adding a larger number of time-of-day dummies is not a good choice. Also, since the time-of-day dummies may be severely correlated, especially for adjacent hours of day, including them as predictors may lead to unstable models. Alternatively, hours of day can be treated as a circular or periodic variable. To obtain a periodic variable of hour of day measured in degrees, hour of day (H) is multiplied by , where *P* represents the known period of the periodic phenomenon (in our case 24 hours). The sine and cosine of the periodic variable are then inserted as explanatory variables. Therefore, the time-of-day effects are included as trigonometric predictors as follows:

The coefficients and may vary across competing risk *r*. Higher order trigonometric polynomials of the periodic variable can be explored as possible independent variables for each event type *r*.

Lastly, is the level-two random effect for the contrast of response category *r* with the reference category 0. The random effects allow the response probabilities to vary across individuals. If , an activity episode in individual *j* has an above-average change of being in response category *r* rather than 0. In other words, for a given value of the covariates, the ratio of to is expected to be higher than average for activity episodes in individual *j*. In the most general specification, we assume that 8 random effects, one for each destination type, follow a multivariate normal distribution with mean **0** and covariance matrix

on the diagonal of is the between-individual variance in the log-odds of a response outcome r versus the reference outcome “stay.” Non-zero variances suggest the presence of unobserved individual-level influences on the odds of being in a response category *r* rather than the reference category 0. On the other hand, on the off-diagonal of is the individual-level covariance while non-zero covariances (i.e., the off-diagonal terms in ) capture correlation between the unobserved individual-specific influences of the different response categories. It should be noted that estimation of these correlations is only possible if the multinomial contrasts are estimated simultaneously.

we add nine random effects, one for each type of nine destinations. It is assume that the random effects are multivariate normally distributed. We also allow for the correlations between the random effects to see if there are shared uno with a covariance matrix. As the correlation between random effects is allowed, it is expected to find out shared unobserved risk factors across competing risks.

The multinomial logit model consists of 9 equations, each contrasting the risk of one type of 9 destination activity with the risk of no event

**RESULTS**

This section describes the model estimation results. This description does not represent a “final” model specification. Since the aim of this paper is to demonstrate the potential of a discrete-time approach to modeling activity duration, we simplify the model by including only a few fixed effects (duration, state, and time-of-day effects) together with random effects. As shown in Table 3, the log-likelihood statistics were good, particularly for a model with a large number of alternatives.

TABLE 3 Model Fit Statistics



> lrtest(mnl.sl, mnl.ml.cov)

Likelihood ratio test

Model 1: chosen ~ thisH2\_EO + thisH2\_ES + thisH2\_HC + thisH2\_PB + thisH2\_SH +

thisH2\_SR + thisHC\_EO + thisHC\_ES + thisHC\_HC + thisHC\_PB +

thisHC\_SH + thisHC\_SR | thisEO + thisES + thisPB + thisSH +

thisSR + log(actDur15) + hSIN1 + hCOS1 | 0

Model 2: chosen ~ thisH2\_EO + thisH2\_ES + thisH2\_HC + thisH2\_PB + thisH2\_SH +

thisH2\_SR + thisHC\_EO + thisHC\_ES + thisHC\_HC + thisHC\_PB +

thisHC\_SH + thisHC\_SR | thisEO + thisES + thisPB + thisSH +

thisSR + log(actDur15) + hSIN1 + hCOS1 | 0

#Df LogLik Df Chisq Pr(>Chisq)

1 84 -10627

2 120 -10557 36 140.67 2.641e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Random Effects**

For comparison, we fitted a single-level multinomial logit model, which is not shown in this paper, that ignores the possible correlation between durations of activity episodes for the same individual. Overall, we found little differences in the parameter estimates between single-level and multilevel models, even if the standard errors in the single-level model were mostly underestimated as expected. We also conducted a likelihood ratio test to test whether the multilevel model performed statistically better than the single-level model. The chi-square statistic of 140.67 with 36 degrees of freedom indicated as a whole the significance of the random-effect parameters, var(*ur*) and cov(*ur,uq*) for *r, q* {H2,H3,EO,ES,HC,PB,SH,SR} and *r q*.

Table 4 shows the covariance matrix of the estimated random effects. There is evidence of unobserved individual-specific influences on the hazard of event type *r* rather than no event, indicated by non-zero variances. The random effect was statistically significant for the risk of returning home temporarily (H2), returning home in the end (H3), and transitioning to an escort activity (ES), but not for other out-of-home activity purposes (i.e., EO, HC, PB, SH, and SR).

between retirees in the hazards of transitions to all types of destination activity.

The correlations between the random effects for transitions to

In other words, the results from the multilevel model, however, show that a significant amount of variation between retirees remains unexplained by the covariates in the model. Later in the paper we illustrate the impact of these unobserved factors on the estimated probabilities of transitions to a new activity

TABLE 4 Random Effects Covariance Matrix



**Fixed Effects**

Table 5 shows the estimated coefficients and standard errors for the fixed part of the full model. We begin by examining the effect of duration of a current activity episode on transition from the current activity to a next activity. The results indicate that the duration effects were significantly positive for all event types of destination activity after controlling for the origin activity states and the times of day. This is as expected because the risk of terminating a current activity and then moving to a next activity increases as the time interval of the current activity episode increases. For example, the odds of returning to home temporarily (H2) versus staying in a current activity increased by with respect to a natural log of one 15-min time interval increment in a current activity episode, for example,

Home activity episodes in the beginning of the day have the lowest rate of any type of transitioning to an out-of-home activity

One could sensibly fit a separate model for each of 8 origin activity states, thus obtaining a state specific random effect. We did not do this here because

The state dummy variables were statistically significant for most destination activity types.

The odds of moving to eat out rather than staying was exp(2.99) = times higher when people are at home in the middle of day (H2) than in the beginning of day (H1).

For example, the predicted probability of moving to eat out is

Time of day effects were significant.

TABLE 5 Estimated Coefficients



* The results show a significant positive effect of activity duration on the propensity of undertaking a new activity
* As a person accumulates time throughout the day, the probability of returning to work becomes much higher
* significant positive transition relationships for returning to home temporarily from escort, personal business, and shopping.

CONCLUSION

* This paper illustrates the use of a discrete-time competing-risks duration model with random effects, with an application to activity participations.
* By including the multilevel random effect, one can control both for unobserved heterogeneity and for possible correlation between durations of activities for individuals who contribute multiple spells of activity.

Lastly, discrete-time duration modeling is essentially the same as discrete choice modeling. For more complex situations, such as repeated events, multiple states, competing risks, and multiple processes, compared with the continuous-time models, it is relatively easy to extend the discrete-time models by using existing procedures for discrete choice models (Steel, 2008).

Activity duration is one of key dimensions needed to develop most activity-based travel demand models, together with activity timing, activity location, travel mode, activity sequencing, and so on. The duration of an activity can be defined as the time to an event occurrence that terminates a current activity state and moves to a next activity state via a trip. The most appropriate analytical tool for modeling duration data is a hazard-based analysis, which is also known as survival analysis in medical science, failure time analysis in industrial engineering, and event history analysis in social science. In the hazard-based approach, the conditional probability

REFERENCES