

# **A Generalized Multiple Durations Proportional Hazard Model With an Application to Activity Behavior During the Evening Work-to-Home Commute**

Chandra R. Bhat

University of Massachusetts at Amherst

## **Abstract**

The model developed in this paper generalizes (in the context of multiple exit states from a duration spell) extant competing risk methods which tie the exit state of duration very tightly with the length of duration. In the current formulation, the exit state is modeled explicitly and jointly with duration models for each potential exit state. The model developed here, however, is much more broad in its applicability than only to the competing risk situation; it is applicable to multiple durations arising from multiple entrance states, multiple exit states, or a combination of entrance and exit states. Multiple entrance states occur frequently in many situations, but have received little attention in the literature. Explicit consideration of the entrance state is important even in single or multiple competing risk models in order to accommodate the sample selection in duration based on the no-entry/entry (to the duration spell) outcome. The generalized multiple durations model developed in the paper is applied to an empirical analysis of activity behavior during the return home from work.

**Keywords:** Proportional hazard, multiple durations, competing risks, activity-type choice, activity durations, work commute.

## 1. Introduction

Hazard-based duration models, which had their roots in biometrics and industrial engineering, are being increasingly used to model duration time in the marketing, economics, and travel demand fields (the reader is referred to Jain and Vilcassim, 1991, Kiefer, 1988, and Hensher and Mannering, 1993 for a review of the applications of duration models in marketing, economics, and travel demand, respectively). Most applications of the hazard model to date have focused on the case where durations end as a result of a single event. For example, the length of unemployment ends when an individual gains employment (Meyer, 1990). A limited number of studies have been directed toward modeling the more interesting and realistic situation of multiple duration-ending outcomes. For example, failure in the context of unemployment duration (*i.e.*, exit from the unemployment spell) can occur either because of a new job, recall to the old job, or withdrawal from the labor force.

Previous research on multiple duration-ending outcomes (*i.e.*, competing risks) have extended the univariate proportional hazard model to the case of two competing risks in one of three ways. The *first method* assumes independence between the two risks (Katz, 1986; Gilbert, 1992). Under such an assumption, estimation proceeds by estimating a separate univariate hazard model for each risk. Unfortunately, the assumption of independence is untenable in most situations and, at the least, should be tested. The *second method* generates a dependence between the two risks by specifying a bivariate parametric distribution for the underlying durations directly. For example, Diamond and Hausman (1985) specify a log bivariate-normal distribution for the durations. This method has the result of placing very strong (and non-testable) parametric restrictions on the form of the baseline cause-specific hazard functions. The *third method* accommodates interdependence between the competing risks by allowing the unobserved components affecting the underlying durations to be correlated. Cox and Oakes (1984) develop a model which generates a positive dependence between the underlying durations based on common dependence on an observed random variable. More recently, Han and Hausman (1990) propose a model which allows an unrestricted correlation in random unobserved components affecting the competing risks. This model permits nonparametric baseline hazard estimation, enables estimation from interval-level data of the type commonly found in econometrics and other fields, and retains an interpretation as an incompletely observed continuous-time hazard model.

A shortcoming of all the extant competing risk methods discussed above is that they tie the exit state of duration very tightly with the length of duration. The exit state of duration is not explicitly modeled in these methods; it is characterized implicitly by the minimum competing duration spell. Such a specification is restrictive, since it assumes that the exit state of duration is unaffected by variables other than those influencing the duration spells and implicitly determines the effects of exogenous variables on exit state status from the coefficients in the duration hazard models (this situation is analogous to the difference between a general endogenous switching regression equation system and the more restrictive disequilibrium market model of demand and supply; see Maddala, (1983).

In this paper, I consider a generalization of the Han and Hausman competing risk specification where the exit state is modeled explicitly and jointly with duration models for each potential exit state. The resulting formulation follows strictly from the proportional hazard specification for the duration spells. This is in contrast to the Han and Hausman specification which uses an approximation to the proportional hazard specification. The model proposed here also extends the Han and Hausman framework to multivariate competing risks.<sup>2</sup> The formulation in this paper does not require placing parametric restrictions on the shapes of hazards within discrete time intervals, as required in the specifications of Han and Hausman, 1990 and Sueyoshi, 1992 (Han and Hausman and Sueyoshi maintain an assumption of a constant hazard within each discrete time-interval in deriving the competing-risk model specification).

A particularly desirable characteristic of the model proposed here is that it is a generalized multiple durations model where the durations can be characterized either by multiple entrance states, multiple exit states or by a combination of entrance and exit states. The focus of econometric literature has been on multiple durations due to multiple exit states (*i.e.*, the competing risk model). However, in many applications, multiple durations may arise because of multiple entrance states. Examples of multiple entrance states include layoffs, being fired, or first-time labor force entry for unemployment duration, activity-type participation choice (shopping, recreation, visiting, *etc.*) for activity duration, and type of initial acquaintance (in college, though personal advertisement, *etc.*)

---

<sup>2</sup> Sueyoshi (1992) has also extended the Han and Hausman framework to the multivariate case. However, like all earlier competing risk models, he characterizes the exit state implicitly based on the duration spells. Further, the Sueyoshi approach becomes cumbersome when dealing with multivariate competing risks since it requires computation of multivariate integrals. In contrast, the approach proposed here requires only the computation of bivariate integrals independent of the number of competing risks.

for marriage durations. Ignoring the entrance state when there are common unobserved factors affecting entrance status and spell duration will lead to biased and inconsistent hazard model parameters due to classic sample selection problems. In this context, information on the absence of a duration spell itself may be valuable; that is, it may be important to consider the “no-entry” state (for example, the “employed” state in unemployment duration modeling, the “home” state in activity duration modeling, or the “unmarried” state in marriage duration modeling) as an explicit entrance state in modeling durations for other entrance states.

In the next section, we develop the model structure and present the estimation procedure for the generalized multiple duration hazard model. Section 3 discusses the data used to model a multiple activity durations model where the multiple durations arise because of multiple entrance states corresponding to participation in different types of activities during the work-to-home trip of individuals. Section 4 presents empirical results from the analysis. The final section provides a summary and identifies important findings.

## 2. Model Structure

Let  $s_{qi}$  represent the (continuous) duration time to failure for individual  $q$  corresponding to outcome  $i$ ,  $i = 1, 2, \dots, I$  (the outcomes may represent multiple entrance states, multiple exit states, or a combination of the two). The outcome-specific hazard function for individual  $q$  at some specified time  $T$  on the continuous-time scale,  $\lambda_{qi}(T)$ , is defined using the proportional hazard specification as (see Kiefer, 1988):

$$\lambda_{qi}(T) = \lim_{\delta \rightarrow 0^+} \frac{\text{prob}[T + \delta > s_{qi} \geq T | s_{qi} \geq T]}{\delta} = \lambda_{0i}(T) \exp(-\gamma_i' x_{qi}), \quad (1)$$

where  $\lambda_{0i}(T)$  is the continuous-time baseline hazard at time  $T$  for outcome  $i$ ,  $x_{qi}$  is a column vector of covariates for individual  $q$  and outcome  $i$ , and  $\gamma_i$  is a column vector of parameters specific to outcome  $i$ . We assume in this paper that the covariates do not change with time  $T$ . As indicated by Han and Hausman, one can include time-varying covariates by using the value of  $x_{qi}$  in each discrete interval or its mean during the interval if  $x_{qi}$  changes within intervals. Equation (1) can be written in the equivalent form (Han & Hausman, 1990; Bhat, 1996a),

$$s_{qi}^* = \ln \Lambda_{0i}(s_{qi}) = \ln \int_0^{s_{qi}} \lambda_{0i}(T) dT = \gamma_i' x_{qi} + \eta_{qi}, \quad (2)$$

Where  $\Lambda_{0i}(\cdot)$  is the integrated baseline hazard for outcome  $i$  and  $\eta_{qi}$  takes an extreme value form with distribution function given by:

$$\Pr(\eta_{qi} < \eta) = G(\eta) = 1 - \exp[-\exp(\eta)] \quad (3)$$

The dependent variable in equation (2) is a continuous unobserved variable when data on failures is available only in grouped form (as is the case in most econometric applications). However, we do observe the time interval,  $t_{qi}$ , in which failure occurs for the observed outcome  $i$ . Let the time intervals be represented by an index  $k$  ( $k = 1, 2, 3, \dots, K$ ) with  $k = 1$  if  $T \in [0, T^1]$ ,  $k = 2$  if  $T \in [T^1, T^2]$ , ...,  $k = K$  if  $T \in [T^{K-1}, 4]$  (the duration intervals can be different for different outcomes; however, for ease of notation, we consider them to be the same for all outcomes). Thus,  $t_{qi} = k$  if the duration spell of individual  $q$  ends in time interval  $k$  for observed outcome  $i$ .

Next consider the modeling of the outcome states using a discrete choice model. Define latent variables  $u_{qi}^*$  as follows:

$$u_{qi}^* = \beta_i' z_{qi} + \varepsilon_{qi} \quad (4)$$

Assume that the  $\varepsilon_{qi}$ 's are identically and independently gumbel distributed across outcomes  $i$  and individuals  $q$  with a location parameter equal to 0 and a scale parameter equal to 1. Outcome  $i$  is observed for individual  $q$  if, and only if,

$$u_{qi}^* > \max_{\substack{j=1,2,\dots,I \\ j \neq i}} u_{qj}^* \quad (5)$$

Let  $R_{qi}$  be a dichotomous variable;  $R_{qi} = 1$  if the  $i$ th outcome is observed for the  $q$ th individual and  $R_{qi} = 0$  otherwise. Defining

$$v_{qi} = \left\{ \max_{j=1,2,\dots,I, j \neq i} u_{qj}^* \right\} - \varepsilon_{qi}, \quad (6)$$

and substituting the right side for  $u_{qi}^*$  from equation (4) in equation (5), we can write:

$$R_{qi} = 1 \text{ if and only if } \beta'_i z_{qi} > v_{qi} \quad (7)$$

The implied marginal distribution of  $v_{qi}$  can be obtained from equation (6) and from the distributional assumptions on the  $\varepsilon_{qi}$ 's as (see McFadden, 1973):

$$F_i(v) = \text{Prob}(v_{qi} < v) = \frac{\exp(v)}{\exp(v) + \sum_{j \neq i} \exp(\beta'_j z_{qj})}, \quad i = 1, 2, \dots, I. \quad (8)$$

The overall equation system for the joint outcome and outcome-specific duration hazards can be written from equations (2) through (7) as:

$$\begin{aligned} R_{qi}^* &= \beta'_i z_{qi} - v_{qi}, \quad R_{qi} = 1 \text{ if } R_{qi}^* > 0, \quad R_{qi} = 0 \text{ otherwise} \\ s_{qi}^* &= \gamma'_i x_{qi} + \eta_{qi}, \quad t_{qi} = k \text{ if } \delta_{i,k-1} < s_{qi}^* \leq \delta_{i,k}, \quad t_{qi} \text{ observed only if } R_{qi} = 1, \quad \delta_{i,k} = \ln \Lambda_{0i}(T^k). \end{aligned} \quad (9)$$

Let the correlation between the random components  $(v_{qi}, \eta_{qi})$  for each outcome be  $\rho_i$  (as in the case of the endogenous switching regression system, the correlations among the random components  $\eta_{qi}$  in the different duration equations are not identified). It is the presence of the  $\rho_i$  terms that do not allow separate duration hazard model estimations for each outcome. The key to accommodating this correlation is to transform the non-normal random variables,  $v_{qi}$  and  $\eta_{qi}$ , into standard normal random variables. With the completely specified marginal distributions  $F_i(\cdot)$  and  $G_i(\cdot)$  for  $v_{qi}$  and  $\eta_{qi}$ , respectively, we write:

$$v_{qi}^* = J_i(v_{qi}) = \Phi^{-1}[F_i(v_{qi})] \text{ and } \eta_{qi}^* = H(\eta_{qi}) = \Phi^{-1}[G(\eta_{qi})] \quad (10)$$

where  $\Phi(\cdot)$  is the standard normal distribution function. It then follows from the probability integral transform result (Feller, 1971) that the transformed variables  $v_{qi}^*$  and  $\eta_{qi}^*$  are standard normal. Then, a bivariate distribution  $L_2$  for  $v_{qi}$  and  $\eta_{qi}$  having the marginal distributions  $F(\cdot)$  and  $G(\cdot)$  can be specified as (see Lee, 1983):

$$L_2(v_{qi}, \eta_{qi}, \rho_i) = \Phi_2[J_i(v_{qi}), H(\eta_{qi}), \rho_i] = \Phi_2(v_{qi}^*, \eta_{qi}^*, \rho_i) \quad (11)$$

Also, because  $F(\cdot)$  and  $G(\cdot)$  are absolutely continuous (monotonically increasing) distribution functions, the transformations  $J_i = \Phi^{-1}F_i$  and  $H = \Phi^{-1}G$  are strictly increasing. Thus, we can re-write equation (9) as

$$\begin{aligned} t_{qi} = k & \text{ if } \Phi^{-1}[G(\delta_{i,k-1} - \gamma'_i x_{qi})] < \eta_{qi}^* \leq \Phi^{-1}[G(\delta_{i,k} - \gamma'_i x_{qi})] \text{ and if } v_{qi}^* \leq \Phi^{-1}[F_i(\beta'_i z_{qi})] \\ & = \text{unobserved if } v_{qi}^* > \Phi^{-1}[F_i(\beta'_i z_{qi})] \end{aligned} \quad (12)$$

From the above equation and the bivariate normal distribution of  $v_{qi}^*$  and  $\eta_{qi}^*$  (equation 11), the joint probability of observing outcome  $i$  and failure in discrete time  $k$  for individual  $q$  is:

$$\begin{aligned} P(R_{qi} = 1, t_{qi} = k) &= \Phi_2[\Phi^{-1}\{F_i(\beta'_i z_{qi})\}, \Phi^{-1}\{G(\delta_{i,k} - \gamma'_i x_{qi})\}, \rho_i] - \\ & \Phi_2[\Phi^{-1}\{F_i(\beta'_i z_{qi})\}, \Phi^{-1}\{G(\delta_{i,k-1} - \gamma'_i x_{qi})\}, \rho_i] \end{aligned} \quad (13)$$

The parameters to be estimated in the multiple duration hazard model are the  $(K-1)$   $\delta_{i,k}$  parameters ( $\delta_{i,0} = -\infty$  and  $\delta_{i,K} = +\infty$ ) and the vectors  $\beta_i$  and  $\gamma_i$  for each possible outcome  $i$ . Defining a set of dummy variables

$$M_{qk} = \begin{cases} 1 & \text{if failure occurs in period } k \text{ for individual } q \text{ (irrespective of outcome } i) \\ & (q = 1, 2, \dots, Q, k = 1, 2, \dots, K) \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

the log likelihood function for the multiple durations model takes the form

$$\log \kappa = \sum_{i=1}^I \sum_{q=1}^Q \left\{ R_{qi} \left( \sum_{k=1}^K M_{qk} \log [P(R_{qi} = 1, t_{qi} = k)] \right) \right\} \quad (15)$$

Accommodation of unobserved heterogeneity in the duration models is conceptually straightforward, but will result in numerical evaluation of integrals in the log-likelihood function above. Also, it should be noted that duration is never observed for the “no-entry” entrance state. Thus, the component of the likelihood function for this outcome becomes

$$\sum_{q=1}^Q R_{qi} \log P(R_{qi} = 1) = \sum_{q=1}^Q R_{qi} \log F_i(\beta'_i z_{qi}). \quad (16)$$

The log-likelihood function in equation (15) does not consider right-censoring. If right-censoring is present in a multiple durations model defined by multiple entrance states, then the entrance state of the censored observation is known and the contribution of the observation (say  $q$ ) to the likelihood function takes the following form (assuming censoring at time-period  $k$ ):

$$\begin{aligned} \log \kappa_q &= \sum_{i=1}^I R_{qi} \log [P(R_{qi} = 1, t_{qi} > k)] \\ &= \sum_{i=1}^I R_{qi} \log \left[ F_i(\beta'_i z_{qi}) - \Phi_2 \left[ \Phi^{-1} \{ F_i(\beta'_i z_{qi}) \}, \Phi^{-1} \{ G(\delta_{i,k} - \gamma'_i x_{qi}) \}, \rho_i \right] \right] \end{aligned} \quad (17)$$

If right-censoring is present when the outcomes are defined by exit states, then the exit state of the censored observation is unknown. The contribution of a censored observation  $q$  to the likelihood function in this case takes the form:

$$\begin{aligned} \log \kappa_q &= \log \left[ \sum_{i=1}^I P(R_{qi} = 1, t_{qi} > k) \right] \\ &= \log \left[ 1 - \sum_{i=1}^I \Phi_2 \left[ \Phi^{-1} \{ F_i(\beta'_i z_{qi}) \}, \Phi^{-1} \{ G(\delta_{i,k} - \gamma'_i x_{qi}) \}, \rho_i \right] \right] \end{aligned} \quad (18)$$

If right-censoring is present when the outcomes are defined by a combination of entrance and exit states, the appropriate contributions of censored observations can be obtained in a manner similar to equations (17) and (18).

All the parameters in the multiple duration model are consistently estimated by maximizing the log-likelihood function. The standard errors of the parameters are obtained as usual from the inverse of the Hessian matrix of the log-likelihood function.

It is easy to see that if  $\rho_i$  is equal to zero for each (and every) outcome  $i$ , then the likelihood in equation (15) partitions into a component corresponding to that of a discrete choice model for outcomes and another component which represents independent univariate ordered logit duration hazard models for each outcome category. The latter component has the form given below (Han and Hausman, 1991):

$$\log \langle \ast = \sum_{i=1}^I \left\{ \sum_{q=1}^Q R_{qi} \sum_{k=1}^K M_{qk} \log [G(\delta_{i,k} - \gamma'_i x_{qi}) - G(\delta_{i,k-1} - \gamma'_i x_{qi})] \right\}. \quad (19)$$

In general, ignoring  $\rho_i$  and estimating independent hazard models for each outcome in a multiple outcome situation will lead to biased estimates of the baseline hazard parameters as well as covariate effects. Thus, it is always preferable to estimate a model which accommodates the sample selection in spell durations based on outcome.

The maximization of the function in equation (15) is achieved using a three-step procedure. In the first step, a separate discrete choice model for outcome  $i$  is estimated along with independent ordered-logit hazard duration models for each outcome. In the second step, the discrete choice model parameters are held fixed and the log-likelihood function in equation (15) is maximized with respect to the duration hazard parameters and the correlation parameters (the hazard parameters from the independent ordered-logit estimations are used as the start parameters, and the correlation terms are initialized to zero). Finally, the parameters from the second step are used as start values for the full-information maximum likelihood estimation of equation (15). The likelihood function at each step is maximized using standard techniques. Maximization is done using the GAUSS matrix programming language. The analytical gradients of the log-likelihood function with respect to the parameters are coded.

The baseline hazard for discrete period  $k$  and outcome  $i$ ,  $\lambda_{0i}^*(k)$ , can be computed using the expression:

$$\lambda_{0i}^*(k) = \frac{G(\delta_{i,k}) - G(\delta_{i,k-1})}{1 - G(\delta_{i,k-1})} \quad (20)$$

where the  $\delta_{i,k}$ 's are estimated from the maximization of equation (15). The continuous-time proportional hazard assumption of equation (1) translates to the following relationship between the discrete-period baseline hazard  $\lambda_{0i}^*(k)$  and the discrete-period hazard  $\lambda_{qi}^*(k)$  at period  $k$  for individual  $q$  and for outcome  $i$ :

$$\lambda_{qi}^*(k) = 1 - [1 - \lambda_{0i}^*(k)]^{\exp(-\beta'x_i)} \quad (21)$$

### 3. Data Source and Sample Description

We apply the generalized multiple durations model proposed in the previous section to examine the duration of participation in shopping and social/recreational activities of workers during the evening work to home commute (in the rest of this paper, we will refer to social/recreational activities together as recreational activities for ease in terminology). Recent studies in the travel demand field have indicated an increasing trend in the number of non-work activity stops made by individuals during the work tour. Gordon *et al.* (1988) report, based on their analysis of the 1990 US National Personal Transportation Survey (NPTS), that non-work travel is the major cause of the evening peak-period congestion and accounts for more than two-thirds of all evening peak-period trips. In an analysis of non-work trips in the northern Virginia suburbs of the Washington, D.C. metropolitan area, Lockwood and Demetsky (1994) find that a significant number of individuals make one or more non-work activity stops during their evening work-to-home trip. These studies emphasize the importance of examining the activity pattern during the evening work-to-home trip and have led to many recent efforts in the travel demand field directed toward this goal (Hamed and Mannering, 1993; Bhat, 1996a). However, none of these previous modeling efforts have attempted to estimate a discrete activity-type choice model jointly with hazard-based activity duration models.

In the current activity duration analysis setting, the outcomes are characterized by multiple entrance states to the duration spell. The entrance states are: return home directly from work (the “no-entry” state), participation in shopping activity after work, and participation in recreational activity after work.

The data source used in the present study is a household activity survey conducted by the Central Transportation Planning Staff (CTPS) of the Massachusetts Highway Department in the Boston Metropolitan region. The survey was conducted in April 1991 and collected data on socio-demographic characteristics of the household and each individual in the household. The survey also included a 1-day (mid-week working day) activity diary to be filled out by all members of the household above 5 years of age. Each activity pursued by an individual was described by: (a) start time, (b) stop time, (c) location of activity participation, (d) travel time from previous activity, (e) travel mode to activity location, and (f) activity type.

The sample for the current analysis comprises 1950 employed adult individuals who made a work-trip on the diary day. Variable definitions and descriptive statistics are provided in Table 1. About 73% of individuals (1432 individuals) do not participate in any activity after work; they

return home directly. The percentage (number) of individuals who participate in recreational and shopping activities during the return home from work is about 9% (163) and 18% (355), respectively. Table 2 provides descriptive information on recreational activity durations for individuals who participate in recreational activities and Table 3 provides corresponding information on shopping activity durations. The length of the discrete periods is 5 min to about 1 h (except for the first period which has a length of 7.5 min), 10 min between 1 h and 90 min, and 20 min between 90 min and 150 min. Durations are observed until termination for all individuals. Thus, there is no right censoring. The final discrete period is open-ended.

The column labeled “Failures” in Tables 2 and 3 indicates the number of individuals whose activity participations end in discrete time period  $k$ . The column titled “No. at Risk” provides information on the number of individuals who are “at risk” of termination of their activity participation in period  $k$ ; that is, it is the number of individuals whose activity durations have not ended at the beginning of period  $k$ . The discrete-period sample hazard rate associated with each period is computed using the Kaplan-Meier (KM) non-parametric estimator (Kiefer, 1988) as the number of individuals who end their activity participations in period  $k$  divided by the risk set in period  $k$ .

The hazard rates across periods in Table 2 and 3 cannot be compared directly since the length of the discrete time periods vary in the current data. To examine the variation in the hazard rates with time, we plot the equivalent continuous-time sample hazard functions  $\lambda_{0_i}(T)$  under the assumption that the continuous-time hazard remains constant within each period  $k$ ; that is,  $\lambda_{0_i}(T) = \lambda_{0_i}(k)$  for all  $T \in \{T^{k-1}, T^k\}$ . Then, we can write:

$$\hat{\lambda}_{0_i}(k) = \frac{\ln[1 - \hat{\lambda}_{0_i}^*(k)]}{\Delta T^k}, \quad k = 1, 2, \dots, K - 1, \quad (22)$$

where  $\Delta T^k$  is the length of the period  $k$  (in minutes) and  $\hat{\lambda}_{0_i}^*(k)$  is the Kaplan-Meier sample estimate of the hazard in period  $k$ .<sup>3</sup>

---

<sup>3</sup> The assumption of constant hazard within time-intervals is made solely for the purpose of examining the temporal evolution of the hazard function when the time-intervals are of varying length (when all the time-intervals are of equal length, the discrete time interval hazard rates can be compared directly across intervals). This assumption is not required by the model structure itself, as indicated in section 1.

The continuous-time sample hazard functions are shown in Fig. 1. These plots show that the sample hazard rate is substantially lower for recreational activity duration compared to shopping activity duration in the first 30 min or so. This is to be expected since participation in recreational activity, in general, implies a certain minimum time investment. In contrast, brief spells of participation in shopping activity are not uncommon. The sample hazard function for recreational activity exhibits sharp peaks between 30 and 60 min, suggesting that most recreational activity durations last for 30-60 min. The shopping activity duration hazard shows more widespread peaking, indicating a more even distribution (on the time scale) of shopping durations. Finally, the hazard is higher for shopping activity relative to recreational activity in the high time range (greater than 2 hours).

#### **4. Empirical Results**

In this section, we present the results for two models. The first model assumes independence between activity-type choice (outcome) and activity duration. The second model accommodates the potential endogeneity in activity-type choice arising from correlation in unobserved components in the activity-type choice and activity duration decisions. The “Independent Durations Model” (first model) is a restricted version of the “Generalized Multiple Durations Model” (second model). We do not consider unobserved heterogeneity in the analysis since it requires time-consuming numerical evaluation of integrals in the generalized multiple durations model. There is some empirical evidence (for example, see Han and Hausman, 1990) that inclusion of heterogeneity has only a minor effect on results when a nonparametric baseline hazard is used. Thus, our decision to ignore unobserved heterogeneity may not result in very serious mis-specification [however, recent work by Bhat (1996a) cautions against generalizing Han and Hausman's result to other empirical contexts; thus, examining alternative ways to incorporate heterogeneity into the current generalized multiple duration framework is an important direction for future research].

The parameters in the activity-type choice model, the effect of covariates on the duration hazard, and the correlation in unobserved components between activity-type choice and durations are presented in Table 4. We discuss each of these sets of estimates in turn in the next few paragraphs. The log-likelihood value at convergence for the independent durations model is -2750.7 and the corresponding value for the generalized multiple durations model is -2745.8. The log-likelihood value for the independent durations model with only alternative specific constants in the

activity-type choice model and only baseline hazard parameters in the recreational and shopping duration models is -2847.0.

The results of the activity-type choice model are almost the same in both the independent durations model and the generalized multiple durations model. In Table 4, we adopt the notation “Var (RA)” to indicate a recreational activity-specific variable and “Var (SA)” to represent a shopping activity-specific variable. Among the socio-demographic variables, the effect of age indicates that older people are less likely to engage in recreational activities and more likely to participate in shopping activities relative to heading directly home after work. The sex of the individual appears to be a strong determinant of participation in shopping and recreational activities; specifically, males are more likely to engage in recreational activities and less likely to engage in shopping activities compared to women [see Chapin (1974) similar results]. The presence of children has been known to influence the activity participation behavior of adults in the household (Jones *et al.*, 1990). The results indicate that individuals with one or more children under 11 years of age are likely to go home directly after work rather than make any shopping or recreational stops. Between recreational and shopping stops, they are less likely to make recreational stops. This may be a reflection of the more discretionary nature of recreational stops relative to shopping stops so that when faced with the responsibility of taking care of young children at home, parents are likely to forego discretionary activities more so than the maintenance-oriented shopping activities. A high income leads to a greater probability of participating in recreational activity. As the number of unemployed individuals in a household increases, it places less burden on the employed individuals to fulfill household maintenance activities and thus is likely to lead to lesser shopping stops on the way back from work, as the results suggest. The coefficients on the work-related characteristics indicate a higher propensity to participate in shopping activities (relative to participation in recreational activities or going home directly) for individuals who drive alone to work (probably because driving alone makes it more convenient to make stops and to transport shopping items back home) and for individuals who leave work before 4 p.m. (presumably because of more time availability). A higher work duration leads to a greater likelihood of going home directly after work.

The estimates of the covariate effects in the hazard duration models for shopping and recreational activities show larger differences (compared to the activity-type choice parameters) between the independent durations model and the generalized multiple durations model. A positive coefficient on a covariate implies that the covariate increases activity duration (equation 2) or,

equivalently, the covariate lowers the hazard rate (equation 1). As to the magnitude of the effect of covariates, a unit change in a covariate results in a  $[\exp(-\gamma_i) - 1] \times 100$  percent change in the continuous-time hazard for activity-type  $i$ . The results indicate that men and individuals whose households have high income tend to have larger durations for recreational activity compared to women and individuals whose households have low income, respectively; however, these effects are smaller for the generalized multiple duration model compared to the independent durations model (the effect of being a male reduces the recreational activity (continuous-time) hazard by 12% in the independent durations model and by 7.4% in the generalized multiple durations model; the corresponding numbers for an increase in household income by \$1,000 are 0.12% and 0.10%, respectively; these results indicate that the effect of income, though statistically significant at the 0.05 level, is rather small). Returning young adults have larger durations in both recreational and shopping activity (no significant differences in the effect of this variable were found between the two activity-types). This is intuitively reasonable since such individuals have less familial responsibilities and, therefore, may have less constraints on their time. The magnitude of the effect of this variable is high relative to the effect of other covariates, and is underestimated in the independent duration models (the continuous-time hazard associated with an individual who is a returning young adult is estimated to be 53% smaller in the independent durations model and 60% smaller in the multiple durations model relative to the hazard of other individuals). Work-related characteristics have a very significant impact on activity duration. Individuals who drive alone to work have shorter shopping activity durations compared to individuals who rideshare or use transit. Departure from work before 4:00 p.m. provides more time and opportunity to participate in shopping activity after work and therefore appears to increase shopping duration. The effect of the “departure from work before 4:00 p.m.” variable on activity duration is overestimated in the independent durations model, while the effect of the “individual drives alone” variable is underestimated. These variables did not have any significant impact on recreational activity duration. Finally, a longer duration at work leaves less time for participation in other activities after work and, therefore, decreases activity duration (increases the activity duration hazard). The effect of this variable was not significantly different between the shopping and recreational duration hazards. The independent durations model overestimates the effect of work duration on activity duration. The percentage shifts in the hazard due to changes in work-related characteristics can be computed in the usual way.

There are two correlation terms in the generalized multiple durations model. These correspond to the correlation in unobserved components affecting recreational activity choice and recreational activity duration and the corresponding correlation for shopping activity and shopping duration. If both of these correlation terms are statistically insignificant, then (and only then) does the multiple durations model collapse to the independent durations model. In the current estimation, both these correlation terms are positive and significantly different from 0 at about the 0.05 level or better (the likelihood ratio test statistic for the test of the hypothesis that both correlation terms are zero is 9.8, which is larger than the chi-squared with 2 degrees of freedom at any reasonable level of significance). Thus, there is sample selection in activity duration based on activity-type choice and the generalized durations model is to be preferred over the independent durations model (note that the correlation terms may be insignificant in certain empirical contexts; however, estimating a generalized multiple durations model is to be always preferred over imposing an *a priori* assumption of independence). From equation (9), note that the positive correlations imply that unobserved factors, that increase the propensity to participate in recreational or shopping activity, also decrease the duration of participation in that activity. The positive correlations may suggest that individuals who by nature tend to participate for shorter durations in a particular activity type are most likely to participate in that activity type, presumably since participation in that activity type presents less time constraints to such individuals than to other individuals. On the other hand, the positive correlations may also be a reflection of the trade-off between frequency of participation in an out-of-home activity-type and activity duration: longer-duration activities are likely to be engaged in less often, making them rarer in the sample.<sup>4</sup>

We now proceed to a discussion of the baseline hazard estimates in the independent durations and generalized multiple durations model. For both models, the nonparametric baseline hazard clearly rejected parametric specifications such as the Weibull distribution. As in the case of the sample hazard rates, we plot the equivalent continuous-time hazard functions under the assumption that the continuous-time hazard remains constant within each discrete-time period. The continuous-time hazard plots are shown in Figs. 2 and 3 for recreational activity and shopping activity, respectively. The temporal trend of the hazard function is similar for each outcome between the independent and generalized multiple durations models (these trends are also similar to the

---

<sup>4</sup> The author would like to thank an anonymous referee for suggesting this alternative explanation for the positive correlations in unobserved components between activity-type choice and activity durations.

corresponding sample hazards plotted in Fig. 1). However, the hazard rate values are lower in the generalized multiple durations model compared to the independent durations model for both outcome categories. The difference in hazard rates is particularly significant for shopping activity. Thus, ignoring the negative correlation between activity type choice and activity duration leads to an overestimation of the hazard rates for each outcome in the current empirical context. This is because the low durations for individuals who participate in out-of-home activities manifested as high hazard rates in the independent durations model. The baseline hazards for the two different outcomes in the generalized multiple durations model indicate the same pattern of differences as discussed earlier with respect to the sample hazard functions.

## 5. Summary and Conclusions

This paper has developed a methodological framework to estimate a joint model of outcome and outcome-specific hazard durations. The outcome-specific hazard models assume a proportional hazard form for incorporating the effects of covariates and are nonparametric with respect to the baseline hazard. The model accounts for the discrete nature of duration data frequently encountered in practice.

The multiple durations model developed in this paper generalizes extant competing risk methods which tie the exit state of duration very tightly with the length of duration. This generalization is achieved by modeling the exit state (or outcome) explicitly and jointly with duration models for each potential exit state. The resulting model follows strictly from the proportional hazard specification for the duration spells. This is in contrast to the Han and Hausman specification which uses an approximation to the proportional hazard specification. The model proposed here lends itself easily to the analysis of multivariate competing risks since it requires only the evaluation of bivariate integrals regardless of the number of competing risks. Current methods, on the other hand, become cumbersome very quickly as the number of competing risks increases, since they require multivariate integration of the same order as the number of risks. Our formulation also does not require placing parametric restrictions on the shapes of hazards within discrete time intervals.

In addition to its application to competing risks, the model proposed here is also applicable to the case where multiple durations arise from multiple entrance states. Such a situation occurs frequently in many practical situations, but has received little (if any) attention in the literature.

Explicit consideration of the entrance state is important, even in single or multiple competing risk models in order to accommodate the sample selection in duration based on the no-entry/entry outcome. The current model formulation can accommodate the sample selection discussed above; more generally, the formulation corresponds to a generalized multiple durations model where the multiple durations can be characterized by multiple entrance states or by multiple exit states or by a combination of entrance and exit states.

The model is applied to an empirical analysis of activity behavior during the return home from work. The outcomes in this empirical context correspond to the multiple entrance states of “go home directly”, “participate in recreational activity”, and “participate in shopping activity”. The results indicate that there is significant sample selection in activity duration based on the entrance state. Ignoring this “endogeneity” in activity-type choice and modeling durations for each outcome independently leads to biases in both baseline hazard estimates and covariate effects. The empirical results also indicate that the activity-type choice and activity durations of individuals during the return home from work are affected in important (and intuitively reasonable) ways by socio-demographic and work-related characteristics.

In closing, it should be noted that the model structure developed in this paper has far wider applicability than to duration hazard models alone. In particular, the formulation corresponds to a joint unordered-ordered choice model in discrete choice literature. The author has recently applied this formulation to analyze mode choice to work (unordered choice) and number of stops during the work tour (ordered choice) in a travel demand context (Bhat, 1996b) and is currently applying it to estimate a model of brand choice (unordered choice) and number of units of purchase (ordered choice) in a marketing context.

### **Acknowledgement**

The author would like to thank Mr. Ian Harrington at the Central Planning Transportation Staff Office in Boston for making the Boston household activity survey data available for this research. Prof. Frank Haight and two anonymous referees provided useful comments on an earlier draft of the paper.

## References

- Bhat, C.R. (1996a). A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity, *Transpn Res.* **30B**, 189-207.
- Bhat, C.R. (1996b). A joint unordered-ordered discrete model of work mode choice and number of stops during the work commute, accepted for publication in *Transportation Research*.
- Chapin, F.S. (1974) *Human Activity Patterns in the City*. John Wiley, New York.
- Cox, D.R. and Oakes D. (1984) *Analysis of Survival Data*, Chapman and Hall, London.
- Diamond, P. and Hausman J. (1984) The retirement and unemployment behavior of older men. In *Retirement and Economic Behavior* (Aaron H. and Burtless G., Eds), pp. 97-134. Brookings Institute, Washington, DC.
- Feller, W. (1971) *An Introduction to Probability Theory and its Applications*. John Wiley, New York.
- Gilbert, C.C.M. (1992) A duration model of automobile ownership. *Transpn Res.* **26B**, 97-114.
- Gordon, P., Kumar A. and Richardson H.W. (1988) Beyond the journey to work, *Transpn Res.* **22A**, 419-426.
- Hamed, M.M and Mannering F.L. (1993) Modeling travelers' postwork activity involvement: toward a new methodology. *Transpn Sci.* **27**, 381-394.
- Han, A. and Hausman J.A. (1990) Flexible parametric estimation of duration and competing risk models, *J. Appl. Econometrics.* **5**, 1-28.
- Hensher, D.A. and Mannering F.L. (1994) Hazard-based duration models and their application to transport analysis. *Trans. Rev.* **14**, 63-82.
- Jain, D.C. and Vilcassim N.J. (1991) Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* **10**, 1-23.
- Jones, P. M., Koppelman F.S. and Orfeuil J.P. (1990) Activity analysis: state of the art and future directions. In *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, pp. 34-55, Gower, Aldershot, U.K.
- Katz, L.F. (1986) Layoffs, recall and the duration of unemployment. NBER working paper no. 1825, University of California, Berkeley.
- Kiefer, N.M. (1988) Economic duration data and hazard functions. *J. Econ.Lit.* **27**, 646-679.
- Lee, L.F. (1983). Generalized econometric models with selectivity. *Econometrica* **51**, 507-512.
- Lockwood, P.B. and Demetsky M.J. (1994) Nonwork travel—a study of changing behavior. Presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, DC, January.
- McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (Zarembka G.S., Ed.) Academic Press, New York.
- Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.
- Meyer, B.D. (1990) Unemployment insurance and unemployment spells. *Econometrica* **58**, 757-782.
- Sueyoshi, G.T. (1992) Semiparametric proportional hazards estimation of competing risks models with time varying covariates. *J.Econometrics* **51**, 25-58.

**List of Figures**

Figure 1. Continuous-time sample hazard functions.

Figure 2. Continuous-time hazard for recreational activity duration.

Figure 3. Continuous-time hazard for shopping activity duration.

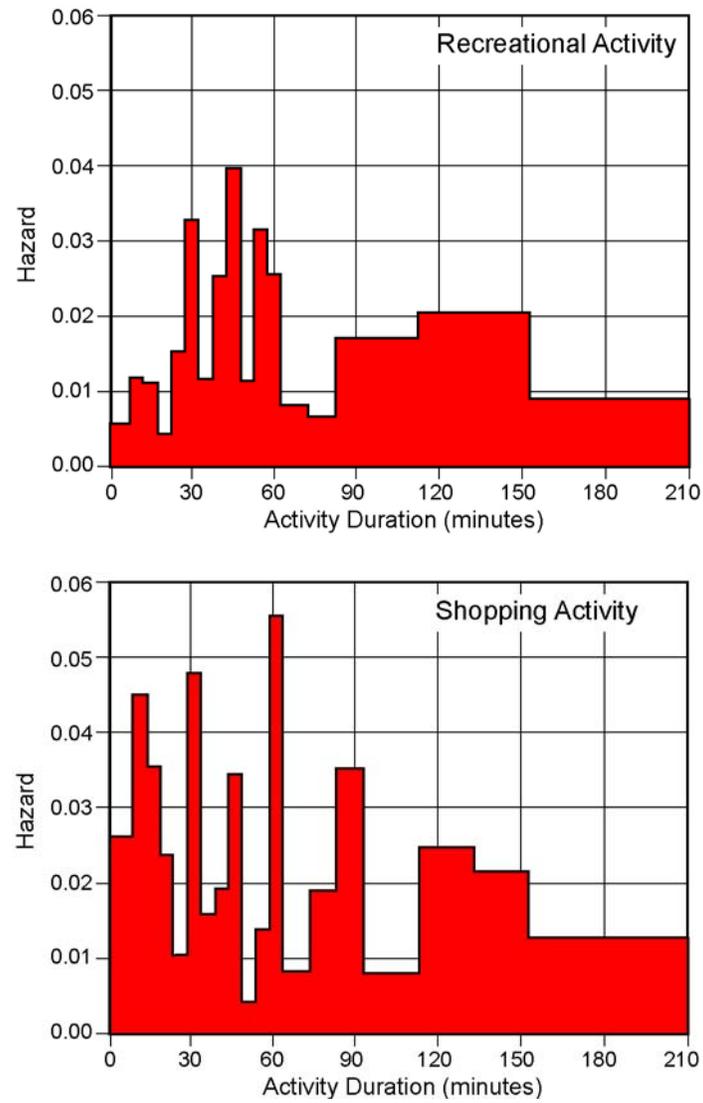
**List of Tables**

Table 1. Variable Definitions and Descriptive Statistics (n = 1950)

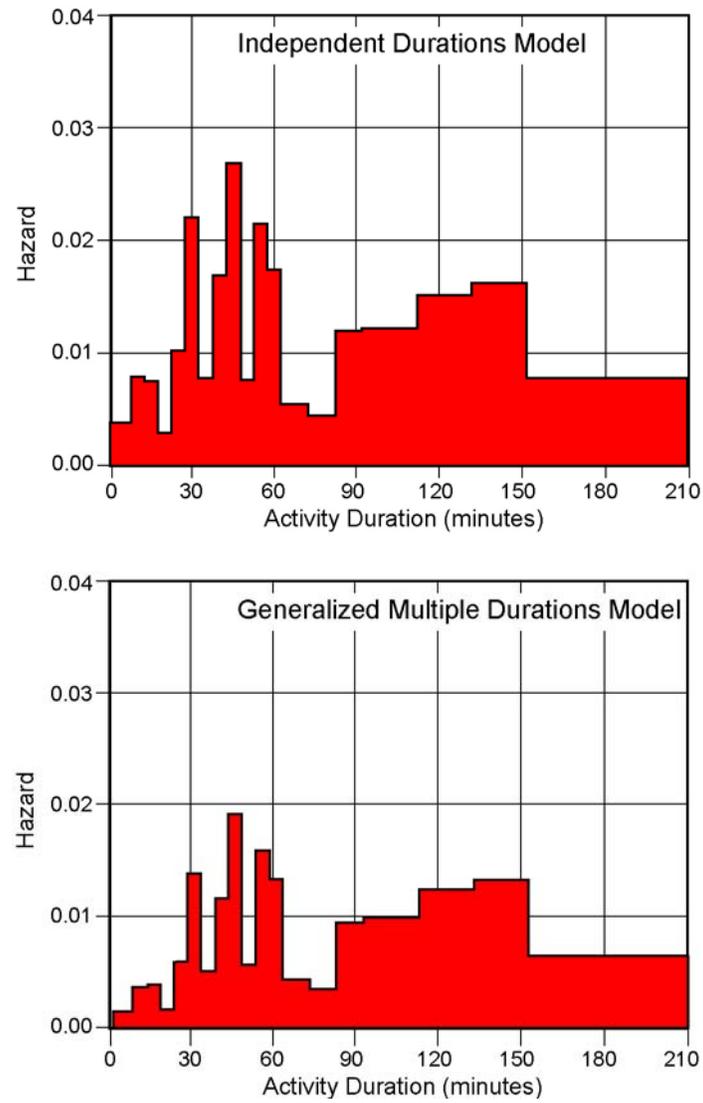
Table 2. Recreational Activity Durations and the Discrete Period Sample Hazard

Table 3. Shopping Activity Durations and the Discrete Period Sample Hazard

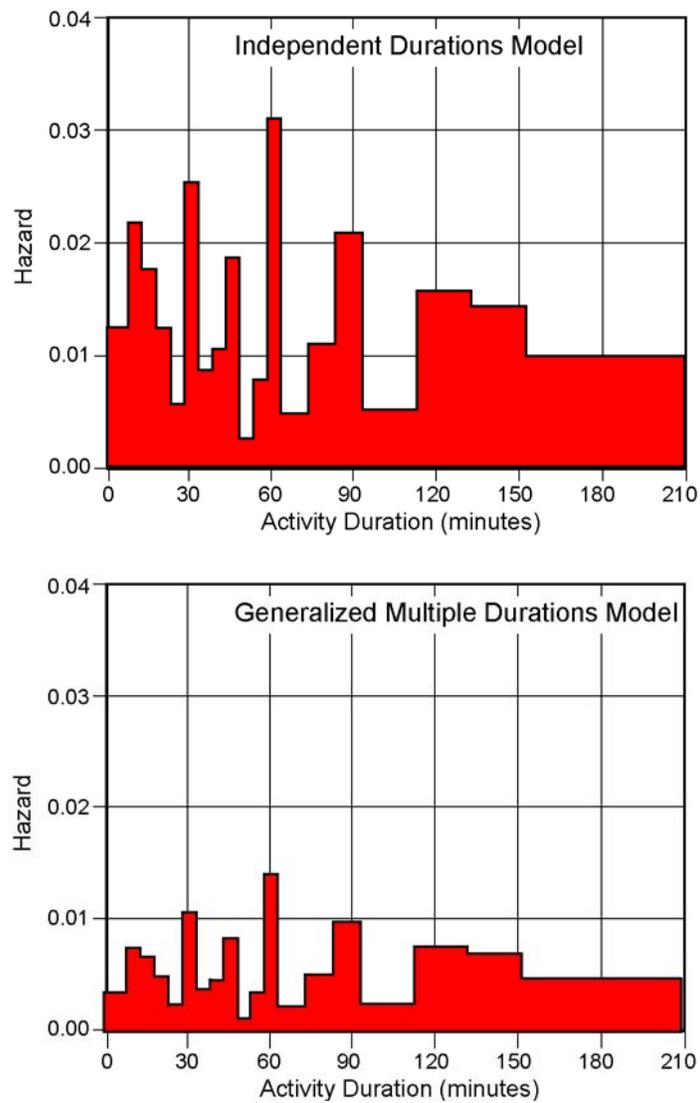
Table 4. Parameter Estimates - Independent Durations and Generalized Multiple Durations



**Figure 1. Continuous-time sample hazard functions.**



**Figure 2. Continuous-time hazard for recreational activity duration.**



**Figure 3. Continuous-time hazard for shopping activity duration.**

**Table 1. Variable Definitions and Descriptive Statistics (n = 1950)**

Variable	Definition	Minimum	Maximum	Mean	Standard Deviation
<b>Socio-demographic</b>					
Age	Age of individual in years	18.000	83.000	41.460	12.518
Male	1 if male	0.000	1.000	0.535	0.499
Presence of kids < 11 yrs	1 if there are one or more children less than 11 years of age in individual's household	0.000	1.000	0.200	0.400
Household income x 10 <sup>-3</sup>	Household income in thousands of dollars per year	15.000	115.000	60.025	27.566
No. of unemployed in HH	Number of unemployed adults in individual's household	0.000	4.000	0.324	0.567
Returning young adult	1 if individual is an employed adult living with one or both parents	0.000	1.000	0.080	0.271
<b>Work-related</b>					
Drive alone to work	1 if individual drives alone to work	0.000	1.000	0.698	0.459
Depart from work before 4 pm	1 if individual leaves work before 4 pm	0.000	1.000	0.294	0.456
Work duration	Time between arrival at work in the morning to departure from work at the evening (in minutes)	60.000	900.000	515.560	106.377

**Table 2. Recreational Activity Durations and the Discrete Period Sample Hazard**

Period $k$	Time interval (mins.)	Failures $F_k^1$	No. at Risk $R_k^2$	Discrete-period $H_k = F_k/R_k$	Std. error of $H_k$
1	0.0 - 7.5	7	163	0.0429	0.0159
2	7.5 - 12.5	9	156	0.0577	0.0187
3	12.5 - 17.5	8	147	0.0544	0.0187
4	17.5 - 22.5	3	139	0.0216	0.0123
5	22.5 - 27.5	10	136	0.0735	0.0224
6	27.5 - 32.5	19	126	0.1508	0.0319
7	32.5 - 37.5	6	107	0.0561	0.0222
8	37.5 - 42.5	12	101	0.1188	0.0322
9	42.5 - 47.5	16	89	0.1798	0.0407
10	47.5 - 52.5	4	73	0.0548	0.0266
11	52.5 - 57.5	10	69	0.1449	0.0424
12	57.5 - 62.5	7	59	0.1186	0.0421
13	62.5 - 72.5	4	52	0.0769	0.0370
14	72.5 - 82.5	3	48	0.0625	0.0349
15	82.5 - 92.5	7	45	0.1556	0.0540
16	92.5 - 112.5	11	38	0.2895	0.0736
17	112.5 - 132.5	9	27	0.3333	0.0907
18	132.5 - 152.5	6	18	0.3333	0.1111
19	152.5 - 212.5	5	12	0.4167	0.1423
20	> 212.5	7	7	1.0000	-

<sup>1</sup> Failures,  $F_k$ , represents the number of individuals whose recreational participation end in period  $k$ .

<sup>2</sup> The number at risk,  $R_k$ , is the number of individuals who are “at risk” of terminating their recreational activity participation in period  $k$ .

**Table 3. Shopping Activity Durations and the Discrete Period Sample Hazard**

Period $k$	Time interval (mins.)	Failures $F_k^1$	No. at Risk $R_k^2$	Discrete-period hazard $H_k = F_k/R_k$	Std. error of $H_k$
1	0.0 - 7.5	64	355	0.1803	0.0204
2	7.5 - 12.5	59	291	0.2027	0.0236
3	12.5 - 17.5	38	232	0.1638	0.0243
4	17.5 - 22.5	22	194	0.1134	0.0228
5	22.5 - 27.5	9	172	0.0523	0.0170
6	27.5 - 32.5	35	163	0.2147	0.0322
7	32.5 - 37.5	10	128	0.0781	0.0237
8	37.5 - 42.5	11	118	0.0932	0.0268
9	42.5 - 47.5	17	107	0.1589	0.0353
10	47.5 - 52.5	2	90	0.0222	0.0155
11	52.5 - 57.5	6	88	0.0682	0.0269
12	57.5 - 62.5	20	82	0.2439	0.0474
13	62.5 - 72.5	5	62	0.0806	0.0346
14	72.5 - 82.5	10	57	0.1754	0.0504
15	82.5 - 92.5	14	47	0.2979	0.0667
16	92.5 - 112.5	5	33	0.1515	0.0624
17	112.5 - 132.5	11	28	0.3929	0.0923
18	132.5 - 152.5	6	17	0.3529	0.1159
19	152.5 - 212.5	6	11	0.5455	0.1501
20	> 212.5	5	5	1.0000	-

<sup>1</sup> Failures,  $F_k$ , represents the number of individuals whose shopping participation end in period  $k$ .

<sup>2</sup> The number at risk,  $R_k$ , is the number of individuals who are “at risk” of terminating their shopping participation in period  $k$ .

**Table 4. Parameter Estimates - Independent Durations and Generalized Multiple Durations**

Model	Variable	Independent Durations Model		Generalized Multiple Duration Model						
		Parameter	t-statistic	Recreational Duration		Shopping Duration				
		Parameter	t-stat.	Parameter	t-stat.	Parameter	t-stat.	Parameter	t-stat.	
Activity- Type Choice	Constant (RA)	-0.5368	-1.23	-0.5599	-1.30	-0.6109	-1.62			
	Constant (SA)	-0.5720	-1.50							
	<b>Socio-demographic characteristics</b>									
	Age (RA)	0.0171	2.45	0.0171	2.49					
	Age (SA)	0.0108	2.24	0.0113	2.39					
	Male (RA)	0.4832	2.74	0.4713	2.66					
	Male (SA)	-0.3946	-3.13	-0.4159	-3.43					
	Presence of kids < 11 yrs (RA)	-0.8378	-3.26	-0.8102	-3.10					
	Presence of kids < 11 yrs (SA)	-0.3145	-1.81	-0.3289	-1.97					
	Household Income x 10 <sup>-3</sup> (RA)	0.0034	1.68	0.0033	1.58					
	No. of unemployed in HH (SA)	-0.3495	-2.73	-0.3166	-2.52					
	<b>Work-related characteristics</b>									
	Drive alone to work (SA)	0.2153	1.60	0.2012	1.50					
	Depart from work before 4 pm (SA)	0.5100	3.84	0.5210	3.97					
Work duration in minutes x 10 <sup>-2</sup> (Home)	0.2502	4.62	0.2455	4.59						
Hazard Duration	Variable	Recreational Duration		Shopping Duration		Recreational Duration		Shopping Duration		
		Parameter	t-stat.	Paramete	t-stat.	Parameter	t-stat.	Parameter	t-stat.	
	<b>Socio-demographic characteristics</b>									
	Male	0.1281	2.91	-	-	0.0770	1.98	-	-	
	Household income x 10 <sup>-3</sup>	0.0012	1.96	-	-	0.0010	1.73	-	-	
	Returning young adult	0.7505	3.33	0.7505	3.33	0.9070	3.37	0.9070	3.37	
	<b>Work-related characteristics</b>									
Drive alone to work	-	-	-0.4626	-3.78	-	-	-0.5757	-3.88		
Departure from work before 4 pm	-	-	0.3406	2.72	-	-	0.1453	1.86		
Work duration in minutes x 10 <sup>-2</sup>	-0.1175	-2.97	-0.1175	-2.97	-0.0712	-2.01	-0.0712	-2.01		
Error correlation between activity-type choice and duration		Assumed to be zero				0.3116	1.62	0.6169	4.75	

Note: Constant (RA) and Constant (SA) are alternative-specific constants for recreational and shopping activities, respectively. Other variables are to be interpreted similarly