A general latent class approach to unobserved heterogeneity in the analysis of event history data

Vermunt, J.K.

Published in:
Applied latent class analysis

Publication date:
2002

Document Version
Peer reviewed version

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Vermunt, J. K. (2002). A general latent class approach to unobserved heterogeneity in the analysis of event history data. In J. Hagenaars, & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 383-407). Cambridge University Press.

To appear in J. Hagenaars and A. McCutcheon (eds.): Applied latent class models, 383-407 Cambridge University Press.

A GENERAL LATENT CLASS APPROACH TO UNOBSERVED HETEROGENEITY IN THE ANALYSIS OF EVENT HISTORY DATA

Jeroen K. Vermunt
Tilburg University

INTRODUCTION

In the context of the analysis of survival and event history data, the problem of unobserved heterogeneity, or the bias caused by not being able to include particular important explanatory variables in the regression model, has received a great deal of attention (see, for instance, Vaupel, Manton, and Stallard 1979; Heckman and Singer 1982, 1984; Trussell and Richards 1985; Chamberlain 1985; Yamaguchi 1986; Mare 1994; Guo and Rodriguez 1994). The reason for this is that this phenomenon has a much larger impact in hazard models than in other types of regression models: Unobserved heterogeneity may introduce, among other things, downwards bias in the time effects, spurious effects of time-varying covariates, spurious time-covariate interaction effects, and dependence among competing risks and repeatable events. This may be true even if the unobserved heterogeneity is uncorrelated with the values of the observed covariates at the start of the process under study.

The models which have been proposed to correct for unobserved heterogeneity differ mainly with respect to the assumptions made about the distribution of the latent variable capturing the unobserved heterogeneity. Heckman and Singer (1982, 1984) proposed a non-parametric random-effects approach which is strongly related to latent class analysis (LCA). Vermunt (1996a, 1997) proposed extending their latent class (LC) approach by specifying simultaneously with the event history model a system of logit models (or causal log-linear model) for the covariates. By explicitly modeling the relationships among the observed and unobserved covariates, it becomes possible to relax and test some of the assumptions which are generally made about the nature of the unobserved heterogeneity. Models can be specified in which the unobserved heterogeneity is related to observed (time-varying) covariates, in which the unobserved heterogeneity itself is time-varying, and in which there are several mutually related latent covariates.

The next section explains event history analysis by means of hazard models paying special attention to situations in which unobserved heterogeneity may distort the results. Subsequently, methods for dealing with unobserved heterogeneity are presented, including the standard and the extended LC approach. The presented LC methodology is illustrated by two empirical examples.

EVENT HISTORY ANALYSIS

The purpose of event history analysis is to explain differences in the time at which individuals experience the events under study. The best way to define an event is as a transition from a particular origin state to a particular destination state. The time variables indicating the duration of nonoccurrence of the events of interest may either be discrete or continuous variables. Although because of space limitations here we deal solely with continuous-time methods, the results can easily be generalized to discrete-time situations (Vermunt 1997). Textbooks on event history and survival analysis are Kalbfleisch and Prentice (1980), Tuma and Hannan (1984), Allison (1984), Lancaster (1990), Yamaguchi (1991), Blossfeld and Rohwer (1995), and Vermunt (1997).

Suppose that we are interested in explaining individual differences in women's timing of the first birth. In that case, the event is having a firth child, which can be defined as the transition from the origin state no children to the destination state one child. This an example of what is called a single non-repeatable event. The term single refers to the fact that the origin state no children can only be left by one type of transition. The term non-repeatable indicates that the event can occur only once. Below, situations in which there are several types of events (multiple risks) and in which events may occur more than once (repeatable events) are presented. In the first birth example, it seems most appropriate to assume the time variable to be a continuous variable although it is, of course, measured discrete, for instance, in days, months, or years after a woman's 15th birthday.

Basic concepts

Suppose T is a continuous random variable indicating the duration of nonoccurrence of the first birth. Let $f(t)$ be the probability density function of T, and $F(t)$ the distribution function of T. As always, the following relationships exist between these two quantities,

$$f(t) = \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t\right)}{\Delta t} = \frac{\partial F(t)}{\partial t}\,,$$

$$F(t) = P(T \leq t) = \int_0^t f(u)d(u)\,.$$

The survival probability or survival function, indicating the probability of nonoccurrence of an event until time t, is defined as

$$S(t) = 1 - F(t) = P(T \geq t) = \int_t^\infty f(u)d(u)\,.$$

Another important concept is the hazard rate or hazard function, $h(t)$, expressing the instantaneous risk of experiencing an event at $T = t$, given that the event did not occur before t. The hazard rate is defined as

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

in which $P\left(t \leq T < t + \Delta t | T \geq t\right)$ indicates the probability that the event will occur during $[t \leq T < t + \Delta t]$, given that the event did not occur before t. The hazard rate is equal to the unconditional instantaneous probability of having an event at $T = t$, $f(t)$, divided by the probability of not having an event before $T = t$, $S(t)$. It should be noted that the hazard rate itself cannot be interpreted as a conditional probability. Although its value is always non-negative, it can take values greater than one. However, for small Δt, the quantity $h(t)\Delta t$ can be interpreted as the approximate conditional probability that the event will occur between t and $t + \Delta t$.

Above $h(t)$ was defined as a function of $f(t)$ and $S(t)$. It is also possible to express $S(t)$ and $f(t)$ completely in terms of $h(t)$, that is,

$$S(t) = \exp\left(-\int_0^t h(u)d(u)\right) ,$$

$$f(t) = h(t)S(t) = h(t)\exp\left(-\int_0^t h(u)d(u)\right) .$$

This shows that the functions $f(t)$, $F(t)$, $S(t)$, and $h(t)$ give mathematically equivalent specifications of the distribution of T.

Log-linear models for the hazard rate

When working within a continuous-time framework, the most appropriate method for regressing the time variable T on a set of covariates is through the hazard rate. This makes is straightforward to assess the effects of time-varying covariates – including the time dependence itself and time-covariate interactions – and to deal with censored observations (Yamaguchi 1991: 10-11; Vermunt 1997: 91-92). Censoring is a form of missing data which is explained in more detail below.

Let $h(t|\mathbf{x}_i)$ be the hazard rate at $T = t$ for an individual with covariate vector $\mathbf{x}_i$. Since the hazard rate can take on values between 0 and infinity, most hazard models are based on a log transformation of the hazard rate, which yields a regression model of the form

$$\log h(t|\mathbf{x}_i) = \log h(t) + \sum_j \beta_j x_{ij} .$$

This hazard model is not only log-linear but also proportional. In proportional hazard models, the time-dependence is multiplicative and independent of an individual's covariate values (Lancaster 1990: 42-43). Below it will be shown how to specify non-proportional log-linear hazard models by including time-covariate interactions.

The various types of log-linear continuous-time hazard models are defined by the functional form that is chosen for the time dependence, that is, for the term $\log h(t)$. In Cox's model, the time dependence is treated in a non-parametric way (Cox, 1972). Exponential models assume the hazard rate to be constant over time, while piecewise exponential model assume the hazard rate to be a step function of T, that is, constant within time periods. Other examples of parametric log-linear hazard models are Weibull, Gompertz, and polynomial models (Blossfeld and Rohwer 1995).

As was demonstrated by several authors (Holford 1980; Laird and Oliver 1981; Clogg and Eliason 1987; Vermunt 1997: 106-117), log-linear hazard models can also be defined as log-linear Poisson models, which are also known as log-rate models. Assume that we have – besides the event history information – two categorical covariates denoted by A and B. In addition, assume that the time axis is divided into a limited number of time-intervals in which the hazard rate is postulated to be constant. In the first birth example, this could be one-year intervals. The discretized time variable is denote by Z. Let h_{abz} denote the constant hazard rate in the zth time interval for an individual with $A = a$ and $B = b$. To see the similarity with standard log-linear models, it should be noted that the log hazard rate can also be written as

$$\log h_{abz} = \log(m_{abz}/E_{abz})$$

where m_{abz} denotes the expected number of events and E_{abz} the total exposure time in cell (a, b, z). Using the notation of hierarchical log-linear models, the saturated log-linear model for the hazard rate h_{abz} can be written as

$$\log h_{abz} = u + u_a^A + u_b^B + u_z^Z + u_{ab}^{AB} + u_{az}^{AZ} + u_{bz}^{BZ} + u_{abz}^{ABZ}, \quad (1)$$

in which the u terms are log-linear parameters which are constrained in the usual way, for instance, by means of ANOVA-like restrictions. It should be noted that this is a non-proportional model as a result of the presence of time-covariate interactions.

As the standard log-linear model, the log-rate model can be formulated in a more general way using a design matrix, i.e.,

$$\log h_{abz} = \sum_j \beta_j x_{abzj}\ . \quad (2)$$

These log-rate models can be estimated using standard programs for log-linear analysis using E_{abz} as a weight vector (Vermunt 1997: 112).

Restricted variants of the (saturated) model described in equation 1 can be obtained by omitting some of the higher-order interaction terms. For example,

$$\log h_{abz} = u + u_a^A + u_b^B + u_z^Z$$

yields a model that is similar to the proportional log-linear hazard model described in equation **??**. Different types of hazard models can be obtained by the specification of the time-dependence. Setting the u_z^Z terms equal to zero yields an exponential model. Unrestricted u_z^Z parameters yield a piecewise exponential model. Other parametric models can be approximated by defining the u_z^Z terms to be some function of Z (or T) (Yamaguchi 1991: 75-77). An approximate Gompertz model, for instance, is obtained by linearly restricting the effect of Z (or T) in a similar way as in log-linear association models (Clogg 1982). And finally, if there are as many time intervals as observed survival times and if the time dependence of the hazard rate is not restricted, one obtains a Cox regression model (Laird Olivier 1981; Vermunt 1997: 113-115).

The presence of unobserved heterogeneity may bias the results obtained from the hazard models discussed so far in various ways (Yamaguchi 1986; Vermunt 1997: 140-141). The best-known phenomenon is the downwards bias of the time dependence that occurs even if – at $T = 0$ – the unobserved factors are uncorrelated with the values of covariates included in the model (Vaupel, Manton, and Stallard 1979; Heckman and Singer 1982, 1984). In such situations, covariate effects will be biased as well since the unobserved variables and the observed variable become correlated after $T = 0$. When the unobserved factors are related with some of the covariates at $T = 0$, that is, when there is some form of selection bias, not only the effects of the covariates concerned are biased, but also spurious time-covariate interactions will be found.

Censoring

A subject that always receives a great amount of attention in discussions on event history analysis is the problem of censoring (Yamaguchi 1991: 3-9; Guo 1993; Vermunt 1997:117-130). An observation is called censored if it is known that it did not experience the event of interest during some time, but it is not known when it experienced the event. In fact, censoring is a specific type of missing data. In the first-birth example, a censored case could be a woman which is 30 years of age at the time of interview (and has no follow-up interview) and does not have children. For such a woman, it is known that she did not have a child until age 30, but be it not known whether nor when she will have her first child.

This is, actually, an example of what is called right censoring. Left censoring means that it is known that the event did not occur, but the exact length of the period that the case concerned did not experience the event is unknown. Left censoring is more difficult to deal with than right censoring (Guo 1993).

As long as it can be assumed that the censoring mechanism is not related to the process under study, dealing with censored observations in maximum likelihood estimation of the parameters of hazard models is straightforward. Let δ_i be a censoring indicator taking the value 0 if observation i is censored and 1 if it is not censored. The contribution of case i to the likelihood function that must be maximized when there are censored observations is

$$\mathcal{L}_i = h(t_i|\mathbf{x}_i)^{\delta_i} S(t_i|\mathbf{x}_i) = h(t_i|\mathbf{x}_i)^{\delta_i} \exp\left(-\int_0^{t_i} h(u|\mathbf{x}_i)du\right) .$$

This likelihood function is, however, only valid if the censoring mechanism can be ignored for likelihood based inference. The presence of unobserved heterogeneity which is shared by the process of interest and the censoring process will lead to a violation of the assumption that censoring is ignorable.

Time-varying covariates

A strong point of hazard models is that one can use time-varying covariates. These are covariates that may change their value over time. Examples of interesting time-varying covariates in the first-birth example are a woman's marital and work status. It should be noted that, in fact, the time variable and interactions between time and time-constant covariates are time-varying covariates as well (Kalbfleish and Prentice 1980: 121-122).

The saturated log-rate model described in equation 1, contains both time effects and time-covariate interaction terms. Inclusion of ordinary time-varying covariates does not change the structure of this hazard model. Suppose, for instance, that covariate B is time varying rather than time constant. This has only implications for the matrix with exposure times E_{abz}. When computing this matrix, it has to be taken into account that individuals can change their value on B (Vermunt 1997: 144-145).

The presence of unobserved heterogeneity may seriously bias the effects of time-varying covariates. Above, we already mentioned the effect of unobserved risk factors on the time dependence and on time-covariate interactions. The effects of time-varying covariates may be partially spurious as a result of the presence of unobserved risk factors which influence both the covariate process and the dependent process (Yamaguchi 1991: 134-139; Vermunt 1997: 140-141).

Multiple risks

Thus far, only hazard rate models for situations in which there is only one destination state were considered. In many applications it may, however, prove necessary to distinguish between different types of events or risks. In the analysis of the first-union formation, for instance, it may be relevant to make a distinction between marriage and cohabitation. In the analysis of death rates, one may want to distinguish different causes of death. And in the analysis of the length of employment spells, it may be of interest to make a distinction between the events voluntary job change, involuntary job change, redundancy, and leaving the labor force.

The standard method for dealing with situations where – as a result of the fact that there is more than one possible destination state – individuals may experience different types of events is the use of a multiple-risk or competing-risk model. For a discussion of the various types of multiple-risk situations, see Allison (1984:42-44) or Vermunt (1997:146-150). A multiple-risk variant of the general log-rate model described in equation 2 is

$$\log h_{abzd} = \sum_j \beta_{dj} x_{abzdj} \, .$$

Here D is a variable indicating the destination state or the type of event that one has experienced. As can be seen, this variable can be used in the hazard model in the same way as the other variables. Note that the index d is used in the parameters and the elements of the design matrix to indicate that the covariate and time effects may be risk dependent.

Again the presence of unobserved heterogeneity may distort the results obtained from the hazard model. More precisely, if the different types of events have shared unmeasured risks factors, the results for each of the types of events is only valid under the observed hazard rates for the other risks. In fact, the resulting dependence among risks is comparable to what in the field of discrete choice modeling is known as the violation of the assumption of independence of irrelevant alternatives (Hill, Axinn, and Thornton 1993).

Multivariate hazard models

Most events studied in social sciences are repeatable, and most event history data contains information on repeatable events for each individual. This is in contrast to biomedical research, where the event of greatest interest is death. Examples of repeatable events are job changes, having children, arrests, accidents, promotions, and residential moves.

Often events are not only repeatable but also of different types, that is, we have a multiple-state situation. When people can move through a sequence of states, events cannot only be characterized by their destination state, as in competing risks models, but they may also differ with respect to their origin state. An example is an individual's employment history: an individual can move through the states of employment, unemployment, and out of the labor force. In that case, six different kinds of transitions can be distinguished which differ with regard to their origin and destination states. Of course, all types of transitions can occur more than once. Other examples are people's union histories with the states living with parents, living alone, unmarried cohabitation, and married cohabitation, or people's residential histories with different regions as states. Special multiple-state models are the well-known Markov and semi-Markov chain models (Tuma and Hannan 1984: 91-115).

Hazard models for analyzing data on repeatable events and multiple-state data are special cases of the general family of multivariate hazard rate models. Another application of these multivariate hazard models is the simultaneous analysis of different life-course events, or as Willekens (1989) calls it, parallel careers. For instance, it can be of interest to investigate the relationships between women's reproductive, relational, and employment careers, not only by means of the inclusion of time-varying covariates in the hazard model, but also by explicitly modeling their mutual interdependence.

Another application of multivariate hazard models is the analysis of dependent or clustered observations. Observations are clustered, or dependent, when there are observations from individuals belonging to the same group or when there are several similar observations per individual. Examples are the occupational careers of spouses, educational careers of brothers (Mare 1994), child mortality of children in the same family (Guo and Rodriguez 1994), or in medical experiments, measures of the sense of sight of both eyes or measures of the presence of cancer cells in different parts of the body. In fact, data on repeatable events can also be classified under this type of multivariate event history data, since in that case there is more than one observation of the same type for each observational unit as well.

The log-rate model can easily be generalized to situations in which there are several origin and destination states and in which there may be more than one event per observational unit (Vermunt 1997: 169). The only thing we need to do is to define – besides the covariates and the time variables – variables indicating the origin state (O), the destination state (D), and the rank number of the event (M). The general log-rate model for such a situation is of the form

$$\log h_{abzodm} = \sum_j \beta_{odmj} x_{abzodmj} \,. \tag{3}$$

Of course, notation could be simplified by replacing the variables D, O, and M by a single variable indicating the type of event defined by origin, destination, and rank number.

The different types of multivariate event history data have in common that there are dependencies among the observed survival times. These dependencies may take several forms. The occurrence of one event may influence the occurrence of another event. Events may be dependent as a result of common antecedents. And, survival times may be correlated because they are the result of the same causal process, with the same antecedents and the same parameters determining the occurrence or nonoccurrence of an event. If these common risk factors are not observed, the assumption of statistical independence of observation is violated, which may seriously distort the results. This is, actually, the same type of problem that motivated the development of multi-level models (Goldstein 1987; Bryk and Raudenbuch 1992).

DEALING WITH UNOBSERVED HETEROGENEITY

As described in the previous section, unobserved heterogeneity may have different types of consequences in hazard modeling. The best-known phenomenon is the downwards bias of the duration dependence. In addition, it may bias covariate effects, time-covariate interactions, and effects of time-varying covariates. Other possible consequences are dependent censoring, dependent competing risks, and dependent observations.

There are two main types of methods that have been proposed for correcting for unobserved heterogeneity: random-effects and fixed-effects methods. Below these two methods are described and a more general non-parametric random-effects method is presented.

Random-effects approach

The random-effects approach is based on the introduction of a time-constant latent covariate in the hazard model (Vaupel, Manton and Stallard 1979). The latent variable is assumed to have a multiplicative and proportional effect on the hazard rate, i.e.,

$$\log h(t|\mathbf{x}_i, \theta_i) = \log h(t) + \sum_j \beta_j x_{ij} + \log \theta_i$$

Here, θ_i denotes the value of the latent variable for subject i. In the parametric random-effects approach, the latent variable is postulated to have a particular distributional form. The amount of unobserved heterogeneity is determined by the size of the standard deviation of this distribution: The larger the standard deviation of θ, the more unobserved heterogeneity there is.

Vaupel, Manton, and Stallard (1979) proposed using a gamma distribution for θ, with a mean of 1 and a variance of $1/\gamma$, where γ is the unknown parameter to be estimated. Several other authors have proposed incorporating a gamma distributed multiplicative random term in hazard models (Tuma, and Hannan 1984: 177-179; Lancaster 1990: 65-70). According to Vaupel, Manton and Stallard (1979), the gamma distribution was chosen because it is analytically tractable and readily computational. Moreover, it is a flexible distribution that takes on a variety of shapes as the dispersion parameter γ varies: When $\gamma = 1$, it is identical to the well-known exponential distribution; when γ is large, it assumes a bell-shaped form reminiscent of a normal distribution.

Heckman and Singer (1982, 1984) demonstrated by an analysis of one particular data set that the results obtained from continuous-time hazard models can be very sensitive to the choice of the functional form of the mixture distribution. Therefore, they proposed using a non-parametric characterization of the mixing distribution by means of a finite set of so-called mass points, or points of support, whose number, locations, and weights are empirically determined. In this approach, the continuous mixing distribution of the parametric approach is replaced by a discrete density function defined by a set of empirically identifiable mass points which are considered adequate to characterize fully the form of the heterogeneity. Often, two or three points of support suffice (Guo and Rodriguez 1994).

It should be noted that Heckman and Singer's arguments against the use of parametric mixing distributions have been criticized by other authors who claimed that the sensitivity of the results to the choice of the mixture distribution was caused by the fact that Heckman and Singer misspecified the duration dependence in the hazard model they formulated for the data set they used to demonstrate the potentials of their non-parametric approach. Trussell and Richards (1985) demonstrated that the results obtained with Heckman and Singer's non-parametric mixing distribution can severely be affected by a misspecification of the functional form of the distribution of T.

The non-parametric unobserved heterogeneity model proposed by Heckman and Singer (1982, 1984) is strongly related to LCA. As in LCA, the population is assumed to be composed of a finite number of exhaustive and mutually exclusive groups formed by the categories of a latent variable. Suppose W is a categorical latent variable with W^* categories, and w is a particular value of W. The non-parametric hazard model with unobserved heterogeneity can be formulated as follows:

$$\log h(t|\mathbf{x}_i, \theta_w) = \log h(t) + \sum_j \beta_j x_{ij} + \log \theta_w$$

Here, θ_w denotes the (multiplicative) effect on the hazard rate for latent class w. It should noted that this log-linear hazard model can also be written in the form of a log-rate model using the parameter notation from hierarchical log-linear models. This is demonstrated below in the examples.

The contribution of the ith subject to the likelihood function which must be maximized in the case of a single nonrepeatable event is

$$\mathcal{L}_i = \sum_{w=1}^{W^*} \pi_w h(t_i|\mathbf{x}_i, \theta_w)^{\delta_i} S(t_i|\mathbf{x}_i, \theta_w) \,,$$

where π_w is the proportion of the population belonging to latent class w. In the terminology used by Heckman and Singer (1982), the number of latent classes (W^*), the latent proportions (π_w), and the effects of W (θ_w) are called the number of mass points, the weights, and the mass points locations, respectively.

The most important drawback of the parametric and non-parametric random-effects approaches to unobserved heterogeneity is that the mixture distribution is assumed to be independent of the observed covariates. This is, in fact, in contradiction to the omitted variables argument which is often used to motivate the use of these types of mixture models. If one assumes that particular important variables are not included in the model, it is usually implausible to assume that they are completely unrelated to the observed factors. In other words, by assuming independence among unobserved and observed factors, the omitted variable bias, or selection bias, will generally remain (Chamberlain 1985; Yamaguchi 1986, 1991: 132). Below, a non-parametric random-effects approach is presented which overcomes this weak point of the standard approaches.

The use of parametric mixture distributions is relatively simple in models for a single non-repeatable event. However, when there is more than one (latent) survival time per observational unit, that is, when there is a model for competing risks, a model for repeatable events, or another type of multivariate hazard model, this is generally not true anymore. It is not so easy to include several possibly correlated parametric latent variables in a hazard model because that makes it necessary to specify the functional form of the multivariate mixture distribution. Therefore, in such cases, most applications use either several mutually independent cause, spell, or transition-specific latent variables, or one latent variable that may have a different effect on the several cause, spell or transition-specific hazard rates. The former approach was adopted, for instance, by Tuma and Hannan (1984:177-183), and the latter, for instance, by Flinn and Heckmann (1982) in a study in which they used a normal mixture distribution. These two specifications of the unobserved factors have also been used in models with non-parametric unobserved heterogeneity (Heckman and Singer 1985; Moon 1991). The general LC approach which is presented below makes it possible to specify models with several mutually related latent variables without the necessity of specifying a distributional form for their joint distribution. The joint distribution of the latent variables is non-parametric as well and can, if necessary, be restricted by means of a log-linear parameterization of the latent proportions.

Fixed-effects approach

A second method for dealing with unobserved heterogeneity involves adding cluster-specific effects, or incidental parameters, to the model (Chamberlain 1985; Yamaguchi 1986). In fact, a categorical variable is included in the hazard model which indicates to which cluster a particular observation belongs: observations belonging to the same cluster have the same value for this "observed" variable while observations belonging to different clusters have different values. Thus, besides the time and covariate effects, we have to estimate a separate parameter for each cluster.

This approach to unobserved heterogeneity, which is called the fixed-effects approach, can only be applied with multivariate survival data, that is, when there is more than one observation for the largest part of the observational units. Note that actually the unobserved heterogeneity is transformed into a form of observed heterogeneity capturing the similarity among observations belonging to the same cluster.

The advantage of using fixed-effects methods to correct for unobserved heterogeneity is that they circumvent two objections against random-effects methods which were presented above: No functional form needs to be specified for the (joint) distribution of the unobserved heterogeneity and the unobserved heterogeneity is automatically related to both the initial state and the time-constant covariates.

The major limitation of the fixed-effects approach is that since each cluster has its own incidental parameter, no parameter estimates can be obtained for the effects of covariates which have the same value for the different observations belonging to the same cluster: Only the effects of observation-specific or of time-varying covariates can be estimated. Another problem is that the incidental parameters cannot be estimated consistently, since by definition they are based on a limited number of observations regardless of the sample size. This inconsistency may be carried over to the other parameters if the parameters are estimated by means of maximum likelihood methods (Yamaguchi 1986).

General non-parametric random-effects approach

To overcome the limitations of the fixed- and random-effects approaches, Vermunt (1997:189-256) proposed a more general non-parametric latent variable approach to unobserved heterogeneity. The main difference between this latent variable method and Heckman and Singer's method is that in the former different types of specifications can be used for the joint distribution of the observed covariates, the unobserved covariates, and the initial state. This means that it becomes possible to specify hazard models in which the unobserved factors are related to the observed covariates and to the initial state. A special case is, for instance, the mover-stayer model proposed by Farewell (1982), in which the probability of belonging to the class of stayers is regressed on a set of covariates by means of a logit model. Moreover, when hazard models are specified with several latent covariates, different types of specifications can be used for the relationships among the latent variables, one of which leads to a time-varying latent variable as proposed by Böckenholt and Langeheine (1996). By means of a multivariate hazard model, the latent covariates can also be related to observed time-varying covariates.

The model that is used for dealing with unobserved heterogeneity consists of two parts: a log-linear path model in which the relationships among the time-constant observed covariates, the initial state, and unobserved covariates are specified, and an event history model in which the determinants of the dynamic process under study are specified.

Suppose there is a model with two time-constant observed covariates denoted by A and B and two unobserved covariates denoted by W and Y. In the first part of the model, the relationships between these four variables are specified by means of a log-linear path model (Goodman 1973; Hagenaars 1990; Vermunt 1996a, 1996b, 1997). Let π_{abwy} denote the probability that an individual belongs to cell (a, b, w, y) of the contingency table formed by the variables A, B, W, and Y. Specifying a modified path model for π_{abwy} involves two things, namely, decomposing π_{abwy} into a set of conditional probabilities on the basis of the assumed causal order among A, B, W, and Y, and specifying log-linear or logit models for these conditional probabilities. At least three meaningful specifications for the causal order among A, B, W, and Y are possible, namely, all the variables are of the same order, the latent variables are posterior to the observed variables, and the observed variables are posterior to the latent variables. In the first specification, π_{abwy} is not decomposed in terms of conditional probabilities. The second specification is obtained by

$$\pi_{abwy} = \pi_{ab}\,\pi_{wy|ab}\,,$$

and the third one by

$$\pi_{abwy} = \pi_{wy}\,\pi_{ab|wy}\,.$$

Note that the second type of specification in which the latent variable(s) depend on covariates was used by Clogg (1981) in his LC model with external variables.

Suppose that the second specification is chosen. In that case, π_{ab} and $\pi_{wy|ab}$ can be restricted by means of a non-saturated (multinomial) logit model. A possible specification of the dependence of the unobserved covariates on the observed covariates is

$$\pi_{wy|ab} = \frac{\exp\left(u_w^W + u_y^Y + u_{aw}^{AW} + u_{bw}^{BW}\right)}{\sum_{wy}\exp\left(u_w^W + u_y^Y + u_{aw}^{AW} + u_{bw}^{BW}\right)}\,.$$

Here, W depends on A and B, while Y is assumed to be independent of W and the observed covariates. Other specifications are, for instance,

$$\pi_{wy|ab} = \frac{\exp\left(u_w^W + u_y^Y + u_{wy}^{WY}\right)}{\sum_{wy}\exp\left(u_w^W + u_y^Y + u_{wy}^{WY}\right)} = \pi_{wy}\,,$$

where the joint latent variable is assumed to be independent of the observed variables, and

$$\pi_{wy|ab} = \frac{\exp\left(u_w^W + u_y^Y\right)}{\sum_{wy}\exp\left(u_w^W + u_y^Y\right)} = \pi_w\pi_y\,,$$

in which the two latent variables are mutually independent and independent of the observed variables.

The second part of the model with non-parametric unobserved heterogeneity consists of an event history model for the dependent process to be studied. The hazard model which is used here is a log-rate model which, as was shown in equation 3, in its most general form involves specifying a regression model for the hazard rate $h_{abzwyodm}$. As before, z serves as index for the time variable, o for the origin state, d for the destination state, and m for the rank number of the event.

When obtaining maximum likelihood estimates of the parameters of a hazard model with the observed covariates A and B and latent covariates W and Y, the contribution of case i to the likelihood function is

$$\mathcal{L}_i \;=\; \sum_{wy}\pi_{abwy}\mathcal{L}_i^*(h)\,,$$

in which $\mathcal{L}_i^*(h)$ denotes the contribution of person i to the complete data likelihood function for the hazard part of the model, and a and b are the observed values of A and B for person i. Vermunt (1997:186-188) proposed estimating the parameters of a general class of event history models with missing data by means of the EM algorithm. A computer program called ℓ_{EM} (Vermunt 1993, 1997) has been developed in which this non-parametric random-effects approach is implemented.

EXAMPLES

This section presents two applications of the general random-effects approach presented above. The first application concerns a single non-repeatable event. In the second example, there is information on the occurrence of four events for each observational unit.

Example I: first interfirm job change

In his textbook on event history analysis, Yamaguchi (1991) presented an example on employees' first interfirm job change to illustrate the use of log-rate models. He reported the data taken from the 1975 Social Stratification and Mobility Survey in Japan in Table 4.1. Here, the same data set is reanalyzed to demonstrate a number of possible specifications of unobserved heterogeneity.

The event of interest is the first interfirm job separation experienced by the sample subjects. The time variable is measured in years. In the analysis, the last one-year time intervals are grouped together in the same way as Yamaguchi did, which results in 19 time intervals. It should be noted that contrary to Yamaguchi, we do not apply a special formula for the computation of the exposure times for the first time interval.

Besides the time variable denoted by Z, there is information on one observed covariate: firm size (F). The first five categories range from small firm (1) to large firm (5). Level 6 indicates government employees. The most general log-rate model that will be used is of the form

$$\log h_{fzw} = u_f^F + u_z^Z + u_{zf}^{ZF} + u_w^W \,,$$

where W is the label of the latent variable assumed to capture the unobserved heterogeneity.

The log-likelihood values, the number of parameters, as well as the BIC values for the estimated models are reported in Table 1.

[INSERT TABLE 1 ABOUT HERE]

We will first describe the results for the models without unobserved heterogeneity. Models A and B serve as reference points. Model A postulates that the hazard rate does not depend on Z and F. Model B is a saturated log-rate model: It includes the main effects of Z and F and the ZF interaction term. The comparison of these two models shows that the log-likelihood is 305 points higher in Model B using 113 additional parameters. A Cox-like model is obtained by omitting the two-variable interaction between Z and F (Model C). With a Cox-like model we mean a model that contains a separate parameter for each time point and that does not contain time-covariate interactions.This model with 90 parameters less than Model B does not perform so badly. The much lower log-likelihood values of models D and E compared to Model C indicate that neither the time effect nor the effect of firm size can be omitted from Model C.

The estimates of time parameters of Model C showed that the hazard rate rises in the first time periods and subsequently starts decreasing slowly. Models F and G were estimated to test whether it is possible simplify the time dependence of the hazard rate on the basis of this information. Model F contains only time parameters for the first and second time point, which means that the hazard rate is assumed to be constant from time point 3 to 19. Model G is the same as Model F except for that it contains a linear term to describe the negative time dependence after the second time point. The comparison between Models F and G shows that this linear time dependence of the log hazard rate is extremely important: The log-likelihood increases 97 points using only one additional parameter. Comparison of Model G with the less restricted Model C and the more restricted Model D shows that Model G captures the most important part of the time dependence. Although according to the likelihood-ratio statistic the difference between Models C and G significant, Model G is the preferred model according to the BIC criterion.

The observed negative time dependence after the second time point may be caused by the presence of unobserved heterogeneity: Individuals with a lower rate of job changing remain with their first employer and therefore the observed hazard rate declines. To test this explanation of the negative time dependence, a model is specified with an unobserved heterogeneity component. Model H contains a two-class latent variable which is assumed to be independent of the observed covariate. This is, in fact, Heckman and Singer's non-parametric model. The other part of the specification is the same as in Model F. As can be seen from the test results, the log-likelihood value of Model H is 72 points higher that of Model F using only 2 additional parameters. This provides evidence that the negative time dependence is at least partially the result of the selection mechanism resulting from unobserved heterogeneity. The comparison of the log-likelihood values of Models G an H indicates that the two-class latent variable W does not fully describe the observed negative time dependence. Inclusion a third latent class does, however, not lead to an improvement of the log-likelihood.

Model I relaxes one of the main assumptions that is generally made in models with unobserved heterogeneity: the unobserved covariate W is allowed to be related to the observed covariate F. The relationship between W and F is modeled treating W as posterior to F, which implies that a logit model is specified for $\pi_{w|f}$. From a substantive point of view, this means that it is assumed that firm size determines the (unobserved) individual characteristics of employees. As can be seen from the increase of the log-likelihood of 24 points using 5 additional parameters, including the direct of effect F on W improves the model a great deal. In Model J, the direct effect of F on the hazard rate is omitted. The fact that Model J performs almost as well as Model I with 5 fewer parameters indicates that – given the postulated causal order between F and W – there is only and indirect effect of firm size on the hazard rate via the latent variable W.

It should be noted that, in this case, the same fit is obtained irrespective of the assumed causal order between W and F. The interpretation of the results is, however, quite different. If F would be assumed to be posterior to W, the results from Model J would indicate that the effect of F on the hazard rate is spurious.

And finally, to simplify the model even more, in Model K the log-linear parameters describing the effect of firm size on W is restricted to be linear for levels 1 to 5 of F. In fact, the log odds of belonging to the high-risk class is assumed to linearly decrease with the size of the firm. Because the result of Model J showed that level 6 (government) is situated between levels 4 and 5, a separate parameter is included for this category. The resulting Model K performs very well.

The parameter estimates of Model K indicate that class one – containing more than 60 percent of the sample – has an eleven times higher rate of an interfirm job change than class two. The probability of belonging to class one is higher for employees of small firms than for employees of large firms: The membership probabilities of class one for the first five levels of the variable firm size are 0.87, 0.79, 0.67, 0.53, and 0.38, respectively. Government employees take a position in between the two largest firm sizes (0.46).

The main substantive conclusion is that there are two types of employees: stayers and movers. When controlling for this unobserved heterogeneity, the association between firm size and the risk of a job change disappears. However, there is a strong association between firm size and being a mover or a stayer. Two plausible explanations for this phenomenon are: 1] firms of different sizes select and contract different types of employees, or 2] firms of different sizes offer different job conditions making employees stayers or movers. These explanation assume that the causal effect goes from firm size to type of employee. The effect could, however, also be reversed: people belonging to the stayer type prefer working at larger firms or governement while movers prefer working at smaller firms.

Another substantive conclusion is that the declining risk of a job change after the second year of employment can be attributed to unobserved heterogeneity. This is a good illustration of the well-known phenomenon of spurious negative time dependence.

Example II: first experience with relationships

The second example concerns adolescents' first experiences with relationships. The data are taken from a small scale two-wave panel survey of 145 cases (Vinken 1998). In this application, we use the information on the age at which the sample members experienced for the first time the events "sleeping with someone", "having a steady friend", "being very much in love", and "going out". A substantial part of the sample is censored, that is, did not experience one or more of these events. The model that is used for each of the hazard rates has the form of a Cox model:

$$\log h_{zwm} = u^{ZM}_{zm} + u^{WM}_{wm} .$$

Here, M is the label for the type of event, Z for the time variable, and W for the latent variable. As can be seen, both the time dependence and the effect of W is assumed to be event specific.

The purpose of the analysis is to use the dependence between the four events to construct a latent typology. This example can either be seen as the application of LCA as a clustering method, where the complication is that the indicators are censored survival times, or as a method for dealing with correlated survival times. In the former case, we could use the label LC cluster analysis and in the latter LC regression analysis (see Vermunt and Magidson 2000; Vermunt and Magidson this volume; and Wedel and DeSarbo this volume). In a second step of the analysis, it is checked whether it is possible to explain differences in class membership by means of three observed covariates.

[INSERT TABLE 2 ABOUT HERE]

Table 2 reports the log-likelihood values and the number of parameters for the one, two, three, and four class models. Although the inclusion of the second class gives the largest improvement in log-likelihood (46 points), also the inclusion of a third and a fourth class improves the log-likelihood. According to the BIC criterion, the two and three class models perform equally well.

[INSERT TABLE 3 ABOUT HERE]

The results for the three-class model are presented in Table 3. The log-linear hazard parameters show that the three identified latent classes differ clearly with respect to the timing of the four events of interest. Class one has the lowest hazard rates for each of the four events while class three has the highest hazard rates. This implies that the first class, which contains 17 percent of the sample, experiences the events later than classes two and three, while class three, which contains 41 percent of the sample, experiences the events earlier than classes one and two. Table 3 also reports the estimates of the median ages at which the three types experienced the events of interest. These indicate that the largest differences between the three types occur in the timing of events 1 (sleeping with someone) and 2 (having a steady relationship).

Finally, a three class model was estimated in which class membership was regressed on three dichotomous covariates by means of multinomial logit model. These covariates are youth centrism (Y), sex (S), and educational level (E). Youth centrism is one of the central concepts in youth studies indicating the extent to which young people perceive their peers as a positively valued ingroup and perceive adults as a negatively valued outgroup. The dichotomous youth-centrism scale which is use here was construct by Vinken (1998) using LCA.

The covariate part of the model is a logit model of the form

$$\pi_{w|yse} = \frac{\exp\left(u_w^W + u_{yw}^{YW} + u_{sw}^{SW} + u_{ew}^{EW}\right)}{\sum_w \exp\left(u_w^W + u_{yw}^{YW} + u_{sw}^{BW} + u_{ew}^{EW}\right)}.$$

The other part of the model consist of the same multivariate hazard model as we used above.

The covariate effects are reported in Table 3, where the omitted level is used as reference category. It can be seen that youth-centristic adolescents, girls, and lower educated belong less often to the late type and more often to the early types. In addition, sex has a much stronger effect on class membership than the other two covariates.

In this example, we showed that the information on the timing of four life events can be summerized into an easy to interpret typology. The typology identified subgroups that experience the relational events at different (median) ages. We also showed how the latent typology can be related to covariates. It turns out that girls experience the events of interest earlier than boys, low educated earlier than high educated, and youth-centristic earlier than non-youth-centristic adolencents.

FINAL REMARKS

This paper described a general non-parametric random-effects approach for dealing with unobserved heterogeneity in the analysis of event histories. This approach, which is based on specifying simultaneously with the hazard model a log-linear path model for the observed and unobserved covariates, overcomes most of the weak points of the standard random- and fixed-effects approaches. Two application were presented to demonstrate the potentials of the new method. Many other applications can be found in Vermunt (1996, 1997), for instance, on models for dependent competing risks, multiple-state models, models with several latent variables, and models with spurious effects of time-varying covariates.

It must, however, be stressed that substantive arguments must determine the specification of the nature of the unobserved heterogeneity. In the job change example, the latent covariate was postulated to be posterior and related to the observed covariate, while in the relationship example, the latent variable was assumed to describe the observed correlations between four survival times. The obtained results make only sense if these are the correct specifications.

The LCA approach described here is a special case of a more general framework for dealing with missing data problems in event history analysis proposed by Vermunt (1996, 1997). Other forms of missing data that can be dealt with are measurement error in covariate values, measurement error in occupied states, partially missing information in covariate values, and partially missing information in occupied states.

REFERENCES

Allison, P.D. 1984. *Event history analysis: regression for longitudinal event data.* Beverly Hills, London: Sage Publications.

Blossfeld, H.P., and Rohwer, G. 1995. *Techniques of event history modeling.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Bryk, A.S., and Raudenbush, S.W. 1992. *Hierarchical linear models: applications and data analysis methods.* Newbury Park: Sage Publications.

Böckenholt, U., and Langeheine, R. 1996. "Latent change in recurrent choice data." *Psychometrika* 61: 285-302.

Chamberlain, G. 1985. "Heterogeneity, omitted variable bias, duration dependence." In *Longitudinal analysis of labor market data*, edited by J.J. Heckman and B. Singer. Cambridge: Cambridge University Press.

Clogg, C.C. 1981. "New developments in latent structure analysis." Pp. 215-245, in *Factor analysis and measurement in sociological research*, edited by D.J. Jackson and E.F. Borgotta. Beverly Hills: Sage Publications.

Clogg, C.C. 1982. "Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*" 77: 803-815.

Clogg, C.C., and Eliason, S.R. 1987. "Some common problems in log-linear analysis." *Sociological Methods and Research* 16: 8-14.

Cox, D.R. 1972. "Regression models and life tables." *Journal of the Royal Statistical Society B* 34: 187-203.

Farewell, V.T. 1982. "The use of mixture models for the analysis of survival data with long-term survivors." *Biometrics* 38: 1041-1046.

Flinn, C.J., and Heckman, J.J. 1982. "New methods for analysing individual event histories." Pp. 99-140, in *Sociological Methodology 1982*, edited by S. Leinhardt.

Goldstein, H. 1987. *Multilevel models in educational and social research.* London: Charles Griffin & Compagny Ltd..

Goodman, L.A. 1973. "The analysis of multidimensional contingenty tables when some variables are posterior to others: a modified path analysis approach." *Biometrika* 60: 179-192.

Guo, G. 1993. "Event-history analysis for left-truncated data." Pp. 217-243, in *Sociological Methodology 1993*, edited by P.V. Marsden. Oxford: Basil Blackwell.

Guo, G., and Rodriguez, G. 1994. "Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala." *Journal of the American Statistical Association* 87: 969-976.

Hagenaars, J.A. 1990. *Categorical longitudinal data - loglinear analysis of panel, trend and cohort data.*. Newbury Park: Sage.

Heckman, J.J., and Singer, B. 1982. "Population heterogeneity in demographic models." In *Multidimensional mathematical demography*, edited by K. Land and A. Rogers. New York: Academic Press.

Heckman, J.J., and Singer, B. 1984. "The identifiability of the proportional hazard model." *Review of Economic Studies* 51: 231-241.

Heckman, J.J., and Singer, B. 1985. “Social sciences duration analysis.” Pp. 38-110, in *Longitudinal analysis of labour market data*, edited by J.J. Heckman and B. Singer. Cambridge: Cambridge University Press.

Hill, D.H., Axinn, W.G., and Thornton, A. 1993. “Competing hazards with shared unmeasured risk factors.” Pp. 245-277, in *Sociological Methodology 1993*, edited by P.V. Marsden. Oxford: Basil Blackwell.

Holford, T.R. 1980. “The analysis of rates and survivorship using log-linear models.” *Biometrics* 32: 299-306.

Kalbfleisch, J.D., and Prentice, R.L. 1980. *The statistical analysis of failure time data*. New York: Wiley.

Laird, N., and Oliver, D. 1981. “Covariance analysis of censored survival data using log-linear analysis techniques.” *Journal of the American Statistical Association* 76: 231-240.

Lancaster, T. 1990. *The economic analysis of transition data*. Cambridge: Cambridge University Press.

Mare, R.D. 1994. “Discrete-time bivariate hazards with unobserved heterogeneity: a partially observed contingency table approach.” Pp. 341-385, in *Sociological Methodology 1994*, edited by P.V. Marsden. Oxford: Basil Blackwell.

Moon, C.G. 1991. “A grouped data semiparametric competing risks model with nonparametric unobserved heteogeneity and mover-stayer structure.” *Economics Letters* 37: 279-285.

Trussell, J., and Richards, T. 1985. “Correcting for unobserved heterogeneity in hazard models using the Heckman-Singer procedure.” Pp. 242-276, in *Sociological Methodology 1985*, edited by N. Tuma. San Francisco: Jossey-Bass.

Tuma, N.B., and Hannan, M.T. 1984. *Social dynamics: models and methods*. New York: Academic Press.

Vaupel, J.W., Manton, K.G., and Stallard, E. 1979. “The impact of heterogeneity in individual frailty on the dynamics of mortality.” *Demography* 16: 439-454.

Vermunt, J.K. 1996a. *Log-linear event history analysis: a general approach with missing data, unobserved heterogeneity, and latent variables*. Tilburg: Tilburg University Press.

Vermunt, J.K. 1996b. “Causal log-linear modeling with latent variables and missing data.” Pp. 35-60, in *Analysis of change: advanced techniques in panel data analysis*, edited by U.Engel and J. Reinecke. Berlin/New York: Walter de Gruyter.

Vermunt, J.K. 1997. “Log-linear models for event history histories.” In *Advanced Quantitative Techniques in the Social Sciences Series*, Volume 8. Thousand Oakes: Sage Publications.

Vermunt, J.K. 1997. *LEM: A general program for the analysis of categorical data. User’s manual*. Tilburg University, The Netherlands.

Vermunt, J.K., and Magidson, J. 2000. *Latent GOLD’s User’s Guide*. Boston: Statistical Innovations Inc..

Vinken, H. 1998. *Political values and youth centrism. Theoretical and empirical perspectives on the political value distinctiveness of Dutch youth centrists*. Tilburg: Tilburg University Press.

Willekens, F.J. 1991. “Understanding the interdependence between parallel careers.” Pp. 2-31, in *Female labour market behavior and fertility, a rational choice approach*, edited by J.J. Siegers, J. De Jong-Gierveld and E. Van Imhoff. The Hague: NIDI.

Yamaguchi, K. 1986. "Alternative approaches to unobserved heterogeneity in the analysis of repeatable events." Pp. 213-249, in *Sociological Methodology 1986*, edited by B. Tuma. Washington, DC.: American Sociological Association.

Yamaguchi, K. 1991. "Event history analysis." In *Applied Social Research Methods*, Volume 28. Newbury Park: Sage Publications.

Table 1: Test results for example I

Model	log-likelihood	# parameters	BIC[1]
without unobserved heterogeneity			
A. $\{\}$	-7261	1	14530
B. $\{ZF\}$	-6956	114	14665
C. $\{Z, F\}$	-7012	24	14203
D. $\{F\}$	-7183	6	14410
E. $\{Z\}$	-7072	19	14286
F. $\{Z_1, Z_2, F\}$	-7123	8	14306
G. $\{Z_1, Z_2, Z_{lin}, F\}$	-7030	9	14129
with unobserved heterogeneity			
H. $\{Z_1, Z_2, F, W\}$	-7051	10	14215
I. $\{Z_1, Z_2, F, W\}$+$\{FW\}$	-7027	15	14166
J. $\{Z_1, Z_2, W\}$+$\{FW\}$	-7028	10	14132
K. $\{Z_1, Z_2, W\}$+$\{F_{lin}W, F_6W\}$	-7031	7	14114

1. Note that the BIC we use here is computed with the log-likelihood value and the number of parameters rather than with the L^2 and the degrees of freedom.

Table 2: Test results for example II

Model		log-likelihood	# parameters	BIC[1]
A.	1 class	-1279	4	2995
B.	2 classes	-1233	9	2929
C.	3 classes	-1220	15	2929
D.	4 classes	-1209	20	2930

1. Note that the BIC we use here is computed with the log-likelihood value and the number of parameters rather than with the L^2 and the degrees of freedom.

Table 3: Parameter estimates for example II: three-class model

Parameter	class 1	class 2	class 3
class size	0.17	0.42	0.41
log-linear hazard parameters			
event 1	-1.79	-0.24	2.03
event 2	-2.14	0.20	1.92
event 3	-0.87	0.24	0.63
event 4	-0.45	-0.16	0.61
median ages			
event 1	23	19	16
event 2	23	18	15
event 3	19	15	14
event 4	15	14	14
log-linear covariate effects			
youth centristic	-0.33	0.18	0.15
boy	1.03	-0.34	-0.68
low education	-0.62	0.44	0.18

SOFTWARE

The kinds of analyses described in this chapter can quite easily be performed with the ℓEM program (Vermunt, 1997). This is the only software for LCA that recognizes the special structure of event history data.

Another option is to use the new LCA program Latent GOLD (Vermunt and Magidson, 2000), which is a fully Windows-based program that is easier to use than ℓEM. In Latent GOLD, an exponential survival model with unobserved heterogeneity should be specified as a LC regression model for a Poisson count. With Latent GOLD's repeated observations option, it is possible to specify piece-wise exponential models.

SYMBOLS

T	continuous time variable
t	value of the time variable
δ	censoring indicator
i	index to denote a particular case
$f(..)$	density function
$F(..)$	distribution function
$S(..)$	survival function
$h(..)$ or h	hazard rate
$\mathbf{x}$	covariate vector
β	model parameter
j	index to denote a particular parameter
Z	discretized time variable
z	index for discretized time variable
A, B, C	categorical covariates
a, b, c	indices for categorical covariates
O	variable indicating the origin state
o	index for the value of origin state
D	variable indicating destination state
d	index for the value of destination state
M	variable indicating type of event
m	index for the value of type of event
E	variable denoting the exposure time
u	log-linear parameter
θ	random effect
π	probability
W, X, Y	latent variables
W^*, X^*, Y^*	number of categories of latent variables
w, x, y	indices for latent variables

FURTHER READING

For further reading on event hisory analysis in general see Kalbfleisch and Prentice (1980), Tuma and Hannan (1984), Allison (1984), Lancaster (1990), Yamaguchi (1991), or Blossfeld and Rohwer (1995). For futher reading on the specific topic of this chapter see Vermunt (1997).