Event history analysis

Jeroen K. Vermunt
Department of Methodology and Statistics
Tilburg University

1 Introduction

The aim of event history analysis is to explain why certain individuals are at a higher risk than others of experiencing the event(s) of interest. This can be accomplished by using special types of methods which, depending on the field in which they are applied, are called failure-time models, life-time models, survival models, transition-rate models, response-time models, event history models, duration models, or hazard models. Examples of textbooks discussing this class of techniques are Allison (1984), Blossfeld & Rohwer (1995), Kalbfleisch & Prentice (1980), Lancaster (1990), Singer & Willett (2003), Tuma & Hannan (1984), Vermunt (1997), and Yamaguchi (1991). Here, I will use the terms event history, survival, and hazard models interchangeably.

A hazard model is a regression model in which the "risk" of experiencing an event at a certain time point is predicted with a set of covariates. Two special features distinguish hazard models from other types of regression models. The first is that they make it possible to deal with censored observations, which are observations containing only partial information on the timing of the event of interest. Another special feature is that they can deal with covariates that change their values during the observation period, which makes it possible to perform a truly dynamic analysis.

Below I will first explain what is actually analyzed in an event history analysis. Then, I introduce the basic statistical concepts for both continuous- and discrete-time analysis. As far as analysis tools themselves is concerned, I will discuss the Kaplan-Meier estimator, which is a method for describing event history data, as well as regression models for continuous- and discrete-time event history data. I will show that after organizing the data in the appropriate manner, an event history analysis can be performed using standard tools for Poisson and logistic regression analysis. Moreover, I will discuss how multilevel and mixture modeling tools can be used to deal with unobserved heterogeneity.

2 State, event, duration, risk period, and censoring

In order to understand the nature of event history data and the purpose of event history analysis, it is important to understand the following five elementary concepts: state, event, duration, risk period, and censoring (Yamaguchi, 1991). These concepts are illustrated below using an example from the analyzes of marital histories.

The first step in the analysis of event histories is to define the discrete *states* that one wishes to distinguish. States are the categories of the "dependent" variable, the dynamics of which one wishes to explain. At every particular point in time, each person occupies exactly one state. In the analysis of marital histories, four states are generally distinguished: never married, married, divorced, and widowed. The set of possible states is sometimes called the state space.

An *event* is a transition from one state to another, that is, from an origin state to a destination state. In the marital history context, a possible event is "first marriage", which can be defined as the transition from the origin state never married to the destination state married. Other possible events are divorce, becoming a widow(er), and non-first marriage. It is important to note that the states which are distinguished determine the definition of possible events. If only the states married and not married were distinguished, none of the above-mentioned events could have been defined. In that case, the only events that could be defined would be marriage and marriage dissolution.

Another important concept is the *risk period*. Clearly, not all persons can experience each of the events under study at every point in time. To be able to experience a particular event, one must occupy the origin state defining the event, that is, one must be at risk of the event concerned. The period that someone is at risk of a particular event – or exposed to a particular risk – is called the risk period. For example, someone can only experience a divorce when he or she is married. Thus, only married persons are at risk of a divorce. Furthermore, the risk period(s) for a divorce are the period(s) that a subject is married. A strongly related concept is the *risk set*. The risk set at a particular point in time is formed by all subjects who are at risk of experiencing the event concerned at that point in time.

Using these concepts, event history analysis can be defined as the analysis of the *duration of the nonoccurrence of an event* during the risk period.

When the event of interest is "first marriage", the analysis concerns the duration of nonoccurrence of a first marriage, in other words, the time that individuals remained in the state of never being married. In practice, as will be demonstrated below, the dependent variable in event history models is not duration or time itself but a transition rate. Therefore, event history analysis can also be defined as the analysis of rates of occurrence of the event during the risk period. In the first marriage example, an event history model concerns a person's marriage rate during the period that he/she is in the state of never having been married.

An issue that always receives a great amount of attention in discussions on event history analysis is *censoring*. An observation is called censored if it is known that it did not experience the event of interest during a certain amount of time, but the exact time at which it experienced the event is unknown. In fact, censoring is a form of missing data. In the first marriage example, a censored case could be a woman who is 30 years of age at the time of interview (and has no follow-up interview) and is not married. For such a woman, it is known that she did not marry until age 30, but it is not known whether or when she will marry. This is, actually, an example of what is called right censoring. Another type of censoring that is more difficult to deal with is left censoring. Left censoring means that there is no information on the duration of nonoccurrence of the event before the start of the observation period.

3 Why event history analysis?

Why is it necessary to use a special type of technique for analyzing event history data? Why is it impossible to relate the incidence of an event within the period of the study to a set of covariates simply by means of, for instance, a logistic regression model, in which the binary dependent variable indicates whether a particular event occurred within the observation period or not? This is, in fact, what is generally done in the analysis of transition data collected by means of a two-wave panel study. If using such a logistic regression modeling approach were a good strategy, it would not be necessary to use special types of methods for analyzing event history data. However, as will be demonstrated below, such an approach has some significant drawbacks.

Suppose there are data on intra-firm job changes of the employees working at company 'C' which have to be used to explain individual differences with regards to the timing of the first promotion. In other words, the aim of the study is to explain why certain individuals in company 'C' remained in their first job longer than others. A single binary dependent variable could be defined indicating whether a given individual received a promotion within, for instance, the first five years after gaining employment in the company concerned. This dependent variable could be related to a set of covariates, such as age, work experience, job level, educational level, family characteristics, and work-related attitudes by means of a logistic regression model.

Although such a simple approach can be quite valuable, it has four important drawbacks (Yamaguchi, 1991). All of them result from the fact that the choice of the period in which the event may have occurred or not is arbitrary. The first problem is that it leads to a severe loss of information since the information on the timing of a promotion within the five-year period, on the promotions that occur after the five-year period, and on the duration of the nonoccurrence of promotions after the five-year period is not used.

The second problem of the approach with a single binary dependent variable is that it does not allow the covariate effects to vary with time; in other words, it cannot contain covariate-time interactions. Suppose that the effect of the variable educational level changes with time, or more precisely, that highly-educated employees have a higher probability of being promoted in the first three years that they work at company 'C', while less educated individuals have a higher probability after three years. In that case, the results will heavily depend on the choice of the length of the time interval. If a short time interval is used, a strong positive effect of the educational level will be found, while longer intervals will lead to a smaller positive effect or perhaps even to a negative effect of the same explanatory variable.

The third disadvantage to the logistic regression approach is that it cannot deal with time-varying covariates. An example of a covariate that can change its value during the five-year period is the number of children that someone has. It may be of interest to test whether the number of children a woman has influences the probability of getting promoted. It is clear that in a real dynamic analysis, it must be possible to use covariates which change their value over time.

The last problem of this simple approach is that is cannot deal with observations which are censored within the five-year period. In this case, there may be two types of censored observations: individuals who leave before working five years at the company concerned and before getting a first promotion, and individuals who had worked less than five years at company 'C' and had not yet been promoted at the time that the data were collected. These two types of observations have in common that they provide the information that the event of interest did not occur during a given period of time, but they do not provide information on whether the event does occur during the remaining part of the five-year period. When using the logistic regression approach, it is not clear what should be done with such censored observations. Ignoring the censored observations implies that the information on non-promotion during a given period of time is not used. On the other hand, incorporating the censored observations in the analysis as observations on individuals that did not experience an event adds information, namely, that they would not have experienced an event if they had worked for at least five years at company 'C'.

4 Basic statistical concepts

The manner in which the basic statistical concepts of event history models are defined depends on whether the time variable T, indicating the duration of nonoccurrence of an event, is assumed to be continuous or discrete. Of course, it seems logical to assume T to be a continuous variable. However, in many situations this assumption is not realistic for two reasons. Firstly, in many cases, T is not measured accurately enough to be treated as strictly continuous. An example of this is measuring the duration variable age of the mother in completed years instead of months or days in a study on the timing of the first birth. This will result in many women having the same score on T, which is sometimes also called grouped 'survival' times.

Secondly, the events of interest can sometimes only occur at particular points in time. Such an intrinsically discrete T occurs, for example, in studies on voting behavior. Since elections take place at particular points in time, changes in voting behavior can only occur at particular points in time. Therefore, when analyzing individual changes in voting behavior, the time variable must be treated as a discrete variable. However, if one wishes to explain changes in political preference rather than in voting behavior, one again has a continuous time variable since political preference may change at any point in time.

4.1 Continuous time

Suppose T is a continuous non-negative random variable indicating the duration of nonoccurrence of the event under study, or, equivalently, the time at which the event under study occurred. Let $f(t)$ and $F(t)$ be the density and cumulative distribution function of T, respectively. As always, these are defined as follows,

$$\begin{aligned} f(t) &= \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t\right)}{\Delta t} = \frac{\partial F(t)}{\partial t}, \\ F(t) &= P(T \leq t) = \int_0^t f(u)d(u). \end{aligned}$$

Rather than working with $f(t)$ and $F(t)$, event history analysis typically works with two other quantities: the survival probability $S(t)$ and the hazard rate $h(t)$. The survival function, indicating the probability of nonoccurrence of an event until time t, is defined as

$$S(t) = P(T > t) = 1 - F(t) = 1 - \int_0^t f(u)d(u).$$

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

$$h(t) = \lim_{\Delta t \to 0} \frac{P\left(t \leq T < t + \Delta t | T \geq t\right)}{\Delta t} = \frac{f(t)}{S(t)}, \tag{1}$$

where $P\left(t \leq T < t + \Delta t | T \geq t\right)$ indicates the probability that the event will occur during $[t, t + \Delta t)$ given that it 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 $T = t$, $S(t)$ or $1 - F(t)$. It should be noted that the hazard rate itself cannot be interpreted as a conditional probability. Even though its value is always non-negative, it can take on 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 $t + \Delta t$.

Because the functions $f(t)$, $F(t)$, $S(t)$, and $h(t)$ give mathematically equivalent specifications of the distributions of T, it is possible to express both $S(t)$ and $f(t)$ in terms of $h(t)$. Since $f(t) = -\partial S(t)/\partial t$, equation (1) implies that

$$h(t) = \frac{-\partial \log S(t)}{\partial t}.$$

By integrating and using $S(0) = 1$, that is, no individual experienced an event before $T = 0$, the important relationship

$$S(t) = \exp\left[-\int_0^t h(u)d(u)\right] = \exp\left[-H(t)\right], \qquad (2)$$

is obtained. The term $H(t) = \int_0^t h(u)d(u)$ is usually referred to as the cumulative hazard function. Note also that $H(t) = -\log S(t)$.

From equation 1 it can be seen that $f(t) = h(t)S(t)$, which shows that also $f(t)$ is a function of the hazard rate. The fact that both the survival and the density function of T can be formulated in terms of the hazard function is used in the maximum likelihood estimation of hazard models.

4.2 Discrete time

Suppose T is a discrete random variable indicating the time of occurrence of an event, and t_l is the lth discrete time point, where $0 < t_1 < t_2 < \ldots < t_L$, with L indicating the total number of time points. If the event occurs at t_l, this implies that the event did not occur before t_l, or, in other words, that the duration of nonoccurrence of an event equals t_{l-1}. It should be noted that this is different from the continuous-time situation in which T indicates both the time that an event occurs and the duration of nonoccurrence of an event.

The probability of experiencing an event at $T = t_l$ is given as

$$f(t_l) = P(T = t_l).$$

The survival function, which indicates the probability of having an event neither before nor at $T = t_l$,[1] is

$$S(t_l) = P(T > t_l) = 1 - P(T \leq t_l) = 1 - \sum_{k=1}^{l} f(t_k).$$

An important quantity in the discrete-time situation is the conditional probability that the event occurs at $T = t_l$, given that the event did not occur prior to $T = t_l$. It is defined as

$$\lambda(t_l) = P(T = t_l | T \geq t_l) = \frac{f(t_l)}{S(t_{l-1})}.$$

[1] It should be noted that some authors define the survival probability in discrete-time situations as the probability of not having an event before t_l: $S(t_l) = P(T \geq t_l)$.

Similar to the way $f(t)$ and $S(t)$ are expressed in terms of $h(t)$ in continuous time, $f(t_l)$ and $S(t_l)$ can be expressed in terms of $\lambda(t_l)$. Since $f(t_l) = S(t_{l-1}) - S(t_l)$,

$$\lambda(t_l) = \frac{S(t_{l-1}) - S(t_l)}{S(t_{l-1})} = 1 - \frac{S(t_l)}{S(t_{l-1})}. \tag{3}$$

Rearrangement of this equation results in

$$S(t_l) \quad = \quad S(t_{l-1}) \left[1 - \lambda(t_l)\right].$$

Using $S(0) = 1$ and $f(t_l) = \lambda(t_l) S(t_{l-1})$ leads to the following expressions for $S(t_l)$ and $f(t_l)$:

$$S(t_l) \quad = \quad \prod_{k=1}^{l} \left[1 - \lambda(t_k)\right], \tag{4}$$

$$f(t_l) \quad = \quad \lambda(t_l) \prod_{k=1}^{l-1} \left[1 - \lambda(t_k)\right] = \frac{\lambda(t_l)}{1 - \lambda(t_l)} \prod_{k=1}^{l} \left[1 - \lambda(t_k)\right]. \tag{5}$$

Because $\lambda(t_l)$ is defined in much the same way as the continuous-time hazard rate $h(t)$, it is sometimes called a hazard rate too, which is, however, not correct. To illustrate this, let us have a closer look at the connection between these two quantities. As can be seen from equation (3), the conditional probability of experiencing an event at t_l equals one minus the probability of surviving between t_{l-1} and t_l. Using $h(t)$, this can also be expressed as follows:

$$\lambda(t_l) \quad = \quad 1 - \exp\left[-\int_{t_{l-1}}^{t_l} h(u) d(u)\right]. \tag{6}$$

If the hazard rate is constant in time interval t_l and the length of this time interval equals 1, this expression can be simplified to

$$\lambda(t_l) \quad = \quad 1 - \exp\left[-h(t_l)\right].$$

Rearranging this equation gives the following reversed relationship between the hazard rate and the probability of experiencing the event in time interval t_l:

$$h(t_l) \quad = \quad -\log\left[1 - \lambda(t_l)\right] \tag{7}$$

The quantity $h(t_l)$ could be called a discrete-time hazard rate, or an approximation of the hazard rate in the lth discrete time interval. Note that the relationship between $h(t)$ and $\lambda(t_l)$ as expressed in equation (6) is only meaningful if the event can occur at any point in time, that is, if time is a continuous variable which is measured discretely.

5 Describing event history data

The most popular descriptive tool for event history data is the Kaplan-Meier estimator of the survival function $S(t)$. Kaplan & Meier (1958) provided a method for obtaining a non-parametric estimate of this function when censoring is present in the data. An alternative non-parametric estimator proposed by Nelson (1972) and Aalen (1974) estimates the cumulative hazard rate function $H(t)$, which can however be transformed into an estimate of $S(t)$ using the relationship shown in equation (2).

Let $0 < t_1 < t_2 < t_l < ... < t_L$ be the ordered (continuous) time points at which events occur, n_l is the number of cases at the risk right after t_{l-1} and d_l the number of events at time point t_l. Note that n_l equals n_{l-1} minus the number of events and the number of censored cases in time interval l. The Kaplan-Meier estimator of the survival function is obtained as follows:

$$S_{KM}(t_l) = \prod_{k=1}^{l} \left(1 - \frac{d_k}{n_k}\right).$$

Note that this formula is very similar to the definition of the discrete-time survival function provided in equation (4), where d_k/n_k serves as an estimator for $\lambda(t_l)$. Using the relation in equation (7) and assuming that the hazard rate is constant in the time interval $(t_{l-1}, t_l]$, hazard rate for this interval is obtained by

$$h_{KM}(t_l) = \frac{-\log\left(1 - d_l/n_l\right)}{(t_l - t_{l-1})}.$$

The Nelson-Aalen estimator for the cumulative hazard rate in time equals

$$H_{NA}(t_l) = \sum_{k=1}^{l} \frac{d_k}{n_k}.$$

The corresponding estimator for the survival function is

$$S_{NA}(t_l) = \exp\left(-\sum_{k=1}^{l} \frac{d_k}{n_k}\right),$$

and for the hazard rate in time interval $(t_{l-1}, t_l]$ is

$$h_{NA}(t_l) = \frac{d_l/n_l}{(t_l - t_{l-1})}.$$

The difference between the two estimators is thus that one estimates $h(t_l)(t_l - t_{l-1})$ as d_l/n_l and the other as $-\log(1 - d_l/n_l)$.

[INSERT TABLE 1 ABOUT HERE]

[INSERT FIGURE 1 ABOUT HERE]

Table 1 gives an example of the Kaplan-Meier and Nelson-Aalen computations for the situation in which there are 5 hypothetical observations. Three persons experienced the event of interest at time points 3, 4 and 7, and two were censored at time points 6 and 10. Typically the Kaplan-Meier survival function will be plotted, possibly with its 95% confidence bound. Moreover, one may depict the survival functions for different groups – for example, a treatment and control group in an experiment – in the same graph to see whether groups have different survival probabilities. Figure 1 depicts the estimated survival function for our small example data set.

6 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, the instantaneous probability (or "risk") of experiencing the event given that it did not occur before (or given that one belongs to the risk set). This is not only meaningful from a substantive point of view, but it also makes it straightforward to assess the effects of time-varying covariates – including the time dependence itself and time-covariate interactions – and to deal with censored observations.

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

$$\log h(t|\mathbf{x}_i) = \log h_0(t) + \sum_j \beta_j \, x_{ij}, \tag{8}$$

or, equivalently,

$$h(t|\mathbf{x}_i) = h_0(t) \exp\left(\sum_j \beta_j \, x_{ij}\right).$$

Here $h_0(t)$ is the baseline hazard and β_j is the parameter associated with predictor x_{ij}. This model is not only log-linear but also a *proportional hazard model*. In these models, the time-dependence is multiplicative (additive after taking logs) and independent of an individual's covariate values. As a result, the estimated hazard rates of two individuals with different covariates are in the same proportion for all time points. Below it will demonstrated that non-proportional log-linear hazard models can be specified by including time-covariate interactions.

The various types of continuous-time log-linear hazard models are defined by the functional form that is chosen for the time dependence, that is, for the term $h_0(t)$. In Cox's semi-parametric model (Cox, 1972), the time dependence is left unspecified. Exponential models assume the hazard rate to be constant over time, while piecewise exponential models 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.

As long as it can be assumed that the censoring mechanism is not related to the process under study, dealing with right 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

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

As can be seen, the likelihood contribution is the survival probability $S(t_i|\mathbf{x}_i)$ (which is a function of the hazard rate between 0 and t_i) for censored cases, and the density $f(t_i|\mathbf{x}_i)$ [which equals $h(t_i|\mathbf{x}_i)S(t_i|\mathbf{x}_i)$] for non-censored cases. This illustrates that it is rather easy to deal with censoring in the maximum likelihood estimation of the parameters of hazard models.

As was demonstrated by several authors, the most important log-linear hazard models can also be defined as log-linear Poisson regression models because the likelihood functions of the two models are equivalent (Laird &

Oliver, 1981; Vermunt, 1997). To illustrate this, assume that the time scale is divided into L intervals in which the hazard rate is constant, which is the typical set up for a piecewise exponential survival model (a piecewise constant hazard model). The upper limit of the lth time interval equals t_l. Let L_i be the time interval in which case i experienced the event or was censored. For $1 \leq l \leq L_i$, let y_{il} be equal to 1 if individual i experienced the event in interval l – if $l = L_i$ and $\delta_i = 1$ – and 0 otherwise, and let e_{il} be the total time that case i belonged to the risk set in this time interval, which equals $t_l - t_{l-1}$ if $l < L_i$ and $t_i - t_l$ otherwise. Using y_{il} and e_{il}, the likelihood contribution of case i is

$$\mathcal{L}_i = \prod_{l=1}^{L_i} \left\{ h(t_l|\mathbf{x}_i)^{y_{il}} \exp\left[-h(t_l|\mathbf{x}_i) \cdot e_{il}\right] \right\}.$$

Assume now that instead of defining a hazard model, one defines a Poisson regression model using a data file containing L_i records for each case, where y_{il} serves as the dependent variable (number of events) and e_{il} as the exposure variable, and in which the x_{ij}'s and a set of time dummies are used as predictors. The Poisson likelihood contribution of case i is the same as in the above hazard likelihood expect for the multiplicative constant C_i,

$$C_i = \prod_{l=1}^{L_i} \frac{(e_{il})^{y_{il}}}{y_{il}!},$$

which does not depend on the model parameters. This shows that a log-linear hazard model with a constant hazard rate within time periods can be estimated by means of standard Poisson regression tools. The only thing required is that the survival information of each case is split into L_i records. This data handling operation is sometimes referred to as *episode splitting*. Table 2 gives an example of two cases in an episode data file for $L = 3$ and $t_1 = 6$, $t_2 = 12$, and $t_3 = 18$. Person 1 experiences the event of interest at time point 8 and person 2 is censored at time point 16.

[INSERT TABLE 2 ABOUT HERE]

The two extreme choices for L are $L = 1$ and L equal to the total number of different observed survival times at which events occur. The former specification yields the exponential survival or constant hazard model; the latter choice yields the well-known Cox regression model. Rather than including a set of dummy variables for the time categories, one can also model the time dependence using a particular restricted functional model, for instance, with a linear or quadratic function.

The assumption of proportional hazards is needed for the partial likelihood estimation procedure proposed by Cox (1972), as well as for the maximum likelihood estimation of most parametric hazard models. In contrast, when using the Poisson modeling set up presented above, there is no need to make this assumption. *Non-proportional hazard models* can simply be obtained by including time-predictor interactions in the Poisson regression model; that is, by allowing the time effect to depend on predictors or predictor effects to depend on time. This is one of the advantages of the log-linear Poisson approach.

Models with time-varying covariates

One of the reasons for building a regression model for the hazard rate instead of, for instance, the survival probability or the time variable T is that this makes possible to relate the occurrence of the event of interest to predictors that change their values over time. Examples of relevant time-varying covariates in the first marriage example are work status and pregnancy. In this context it should be noted that, in fact, the time variable itself and time-covariate interactions are also time-varying predictors.

With time-varying predictors, a log-linear hazard model becomes

$$\log h(t|\mathbf{x}_{it}) = \log h_0(t) + \sum_j \beta_j \, x_{ijt}.$$

As can be seen, the only change compared to the model in equation (8) is that the predictors have an index t. But, how is such a model estimated in practice? Within the Poisson modeling framework described above, inclusion of time-varying covariates requires only some extra data handling, which is another advantage to this approach. Recall that the time dependence was dealt with by splitting the event history information of each case into L_i episodes, in each of which the categorical time variable is constant. The same trick is used for the time-varying covariates. More specifically, one has to create an episode data set in which the time-varying predictors are constant within episodes. Once this additional episode splitting is done, the same Poisson regression procedure can be used as for models without time-varying covariates.

Table 3 illustrates episode splitting for time-varying covariates by expanding the example of Table 2. A second covariate is added which is time varying: it changes its value at time point 2 for person 1 and at time point 7 for person 2. This means that for person 1 the first record in Table 2 is split into two separate episodes and for person 2 the second record.

[INSERT TABLE 3 ABOUT HERE]

Competing-risk models

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 model that is usually used when individuals may leave the origin state to different destination states is the competing-risk model. This multiple-risk variant of the hazard rate model described in equation (8) can be defined as follows:

$$\log h_d(t|\mathbf{x}_{it}) = \log h_{0d}(t) + \sum_j \beta_{jd}\, x_{ijt}.$$

Here, the index d indicates one of the D destination states or event types. As can be seen, the only change in the hazard model compared to the single type of event situation is that a separate set of time and covariate effects is included for each type of event. As far as the maximum likelihood estimation of the parameters of competing-risk models is concerned, it is important to note that a person experiencing event d is treated as a censored case for the other $D-1$ risks.

Also the competing-risk models can be easily set up using the Poisson modeling framework described earlier. The data sets should contain D sets of episode records, one for each of the competing risks. The dependent variable takes on the value one if the person experienced that event in the time interval concerned, and is equal to zero otherwise. A variable "event type" can be added to the data file to allow the time dependence and predictor effects to be risk specific. Again data restructuring and creating the right set of predictors will do the full job. Table 4 modifies the example of Table 3 for the situation in which $D = 2$ and person 1 experiences the second type of event and person 2 is censored.

[INSERT TABLE 4 ABOUT HERE]

Models for multivariate event histories

Many events studied in the social and behavioral sciences are repeatable. 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, residential moves, curing from a mental illness, and moving to a next developmental stage.

Often events are not only repeatable but also of different types, yielding a situation in which transitions may occur across multiple states. 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.

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. 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, child mortality of children in the same family, 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 hazard 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. The only thing that changes is that indices are needed for the origin state (o), the destination state (d), and the rank number of the event (m). A log-linear hazard rate model for such a situation is

$$\log h_{od}^{m}(t|\mathbf{x}_{it}) = \log h_{0od}^{m}(t) + \sum_{j} \beta_{jod}^{m} x_{ijt}.$$

Also this model can be specified as a Poisson regression model after organizing the data in the right way. The most important difference with the previous specifications is that the dependent variable may be equal to 1 more that ones.

The various 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. Hence, unobserved heterogeneity should be taken into account (see below).

7 Discrete time models

When the time variable is measured rather crudely, which typically leads to many ties in the recorded event times, or when the process under study is intrinsically discrete, it is more appropriate to use a discrete-time event history model. These models involve regressing the conditional probability of occurrence of an event in the lth time interval given that the event did not occur before this period, denoted by $\lambda(t_l)$, on a set of covariates. It must be noted that when these probabilities are relatively small for all time and covariate combinations, the parameters of discrete-time models and continuous-time models are very similar. The reason for this is that the hazard rate $h(t)$ and $\lambda(t_l)$ have almost the same value if the hazard rate is small. On the basis of the relationship between $h(t)$ and $\lambda(t_l)$ given in equation (7), it can be derived that values of .1, .2, and .5 for $\lambda(t_l)$ correspond with values of .105, .223, and .693 for $h(t)$. This means that if all $\lambda(t_l)$ are about .1 or smaller, discrete-time methods provide good approximations of continuous-time methods.

There are several ways to parameterize the dependence of the conditional probability of experiencing an event on time and on covariates. The most popular choice is the logistic regression function (Cox, 1972; Allison, 1982; Singer & Willett, 2003)

$$\lambda(t_l|\mathbf{x}_{it_l}) \;=\; \frac{\exp\left(\alpha_l + \sum_j \beta_j \, x_{ijt_l}\right)}{1+\exp\left(\alpha_l + \sum_j \beta_j \, x_{ijt_l}\right)},$$

which leads to the well-known discrete-time logit model

$$\log\left[\frac{\lambda(t_l|\mathbf{x}_{it_l})}{1-\lambda(t_l|\mathbf{x}_{it_l})}\right] \;=\; \alpha_l + \sum_j \beta_j \, x_{ijt_l}.$$

Although the logistic regression model is a somewhat arbitrary choice, it has the advantages that it constrains $\lambda(t_l|\mathbf{x})$ to between 0 and 1 and that it can be estimated with generally available software (as is shown below).

On the other hand, if one assumes that the data are generated by a continuous-time proportional hazard model, it is more appropriate to use the complementary log-log transformation for $\lambda(t_l)$ (Allison, 1982). As can be derived from equation (6), the conditional probability of experiencing an event in time interval l can be written in terms of the hazard rate as

$$\lambda(t_l|\mathbf{x}_{it_l}) \;=\; 1-\exp\left(-\int_{t_{l-1}}^{t_l} h(u|\mathbf{x}_{it_l})d(u)\right).$$

If there is no information on the variation of the hazard rate within time intervals, it seems reasonable to assume that the hazard rate is constant within intervals, or that

$$\lambda(t_l|\mathbf{x}_{it_l}) \;=\; 1-\exp\left(-h(t_l|\mathbf{x}_{it_l})\cdot e_l\right), \tag{9}$$

in which e_l denotes the length of the lth time interval. This amounts to assuming exponential survival within every particular time interval. Suppose the following log-linear and proportional hazard model is postulated:

$$h(t_l|\mathbf{x}_{it_l}) \cdot e_l \;\; = \;\; \exp\left(\alpha_l + \sum_j \beta_j \, x_{ijt_l}\right). \qquad (10)$$

Substitution of equation (10) into equation 9 yields

$$\lambda(t_l|\mathbf{x}_{it_l}) \;\; = \;\; 1 - \exp\left[-\exp\left(\alpha_l + \sum_j \beta_j \, x_{ijt_l}\right)\right].$$

Rearrangement of this equation yields what is known as the complementary log-log transformation of the conditional probability of experiencing an event at t_l,

$$\log\left\{-\log\left[1 - \lambda(t_l|\mathbf{x}_{it_l})\right]\right\} \;\; = \;\; \alpha_l + \sum_j \beta_j \, x_{ijt_l}.$$

The β parameters can now be interpreted as the covariate effects on the hazard rate under the assumption that $h(t_l)$ is constant within each of the L time intervals. Since $h(t_l|\mathbf{x}_{it_l})\dot{e}_l$ appears at the left-hand side of equation (10) instead of $h(t_l|\mathbf{x}_{it_l})$, the estimates for the baseline hazard rates or the time parameters must be corrected for the interval lengths e_l: $\log h_0(t_l) = \alpha_l - \ln(e_l)$.

If the model is a proportional hazard model, that is, if there are no time-covariate interactions, the β parameters of a complementary log-log model are not affected by the choice of the interval lengths since e_l is completely absorbed into α_l. This is the main advantage of this approach compared to the discrete-time logit model, which is not only sensitive to the choice of the length of the intervals, but also requires that the intervals be of equal length (Allison, 1982). The reason for this is that the interval length influences the probability that an event will occur in the interval concerned, and therefore also the logit of $\lambda(t_l)$. Although the complementary log-log model can handle unequal interval lengths in proportional hazard models with one parameter for each time interval, unequal time intervals are problematic when the time dependence is parameterized or when the model is nonproportional (Allison, 1982).

Discrete-time models are typically estimated by means of maximum likelihood methods. Just as in continuous-time models, $f(t_l|\mathbf{x}_i)$ is the contribution to the likelihood function for an individual who experienced an event and $S(t_l|\mathbf{x}_i)$ for an individual who was censored. Letting l_i denote the time interval in which the ith person experienced an event or was censored and using the definitions in equations (4) and (5), its likelihood contribution is

$$\mathcal{L}_i \;=\; \left(\frac{\lambda(t_{l_i}|\mathbf{x}_{it_l})}{1-\lambda(t_{l_i}|\mathbf{x}_{it_l})}\right)^{\delta_i} \prod_{k=1}^{l_i} \left(1-\lambda(t_k|\mathbf{x}_{it_k})\right).$$

Let y_{il}, for $1 \leq l \leq l_i$, be a variable taking on the value 1 if person i experienced an event in t_l – that is, if $l_i = l$ and $\delta_i = 1$ – and 0 otherwise. Using this vector of indicator variables, the likelihood contribution of case i becomes

$$\mathcal{L}_i \;=\; \prod_{k=1}^{l_i} \left\{\lambda(t_k|\mathbf{x}_{it_k})^{y_{ik}} \left[1-\lambda(t_k|\mathbf{x}_{it_k})\right]^{(1-y_{ik})}\right\},$$

which is, in fact, the likelihood contribution of l_i observations in a regression model for a binary response variable. This shows that a discrete-time logit model can be estimated by means of standard software for logistic regression analysis. The data file should contain one record for every time unit that an individual belongs to the risk set. Such a file is sometimes called person-period records. The complementary log-log model is available in generalized linear modeling (GLM) routines.

8 Unobserved heterogeneity

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. This is not surprising because this phenomenon, which is also referred to as selectivity or frailty, may have a much larger impact in hazard models than in other types of regression models.

[INSERT TABLE 5 ABOUT HERE]

With a small hypothetical example, I will illustrate some of the biasing effects that unobserved heterogeneity may have on the parameter estimates of hazard models. Suppose that in the population under study there are two dichotomous factors, X_1 and X_2, that affect the hazard rate. The baseline hazard $h_0(t)$ for the group with $X_1 = 0$ and $X_2 = 0$ is 0.01 (constant over time). Controlling for X_2, the hazard rate for $X_1 = 1$ is two times larger than for $X_1 = 0$, and controlling for X_1, the hazard rate for $X_2 = 1$ is five times larger for than $X_2 = 0$. In addition, assume that at $T = 0$ each combination of X_1 and X_2 contains 25% of the population. Table 5 shows the resulting hazard rates for each of the possible combinations of X_1 and X_2 at four time points. As can be seen, the true hazard rates are constant over time within levels of X_1 and X_2. The hazard rates in the columns labeled "marginal" show what happens when X_2 is not observed; that is, after marginalizing over X_2. The first thing that can be seen is that despite of the true rates being time constant for both $X_1 = 0$ and $X_1 = 1$ the marginal hazard rates decline over time. This is an illustration of the fact that unobserved heterogeneity biases the estimated time dependence in a negative direction. Furthermore, whereas the marginal hazard ratio between $X_1 = 1$ and $X_1 = 0$ equals the true value 2.00 at $t = 0$, it declines over time (see last column). Thus, when estimating a hazard model with these marginal hazard rates, a smaller effect of X_1 than the true value of (log) 2.00 will be found. Finally, modeling the changing (declining) effect X_1 over time or, equivalently, the smaller (negative) time effect for $X_1 = 0$ than for $X_1 = 1$ requires the inclusion of a time-X_1 interaction in the hazard model.

Unobserved heterogeneity may have different types of consequences in hazard modeling. The best-known phenomenon is the downwards bias of the duration dependence illustrated with the hypothetical example. In addition, as could also be seen, it may bias covariate effects and time-covariate interactions. Other possible consequences are dependent or informative censoring, dependent competing risks, and dependent multivariate observations (Vermunt, 1997). The common way to deal with unobserved heterogeneity is to include random effects in the hazard model of interest (Heckman & Singer, 1982; Vaupel, Manton, & Stallard, 1979).

Specification of a random-effects hazard model involves the introduction of a time-constant latent covariate in the model. The latent variable is typically assumed to have a multiplicative and proportional effect on the hazard rate, i.e.,

$$\log h(t|\mathbf{x}_{it}, \theta_i) = \log h(t) + \sum_j \beta_j \, x_{ijt} + \theta_i,$$

where 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. When working with the Poisson modeling set up, these types of models can be estimated with random-effects Poisson regression software.

Heckman & Singer (1982) showed that the results obtained from a random-effects continuous-time hazard model can be sensitive to the choice of the functional form of the mixture distribution. They, therefore, proposed using a non-parametric characterization of the mixing distribution by means of a finite set of so-called mass points, or latent classes, whose number, locations, and weights are empirically determined. For a more extended discussion, see Vermunt (1997) and Vermunt (2002). This non-parametric approach, as well as the parametric approach with normally distributed random effects, is implemented in the Latent GOLD software (Vermunt & Magidson, 2005).

9 An empirical example

A small example of hazard modeling is now presented in which a data set is used from the 1975 Social Stratification and Mobility Survey in Japan reported in Yamaguchi's textbook on event history analysis (Yamaguchi, 1991). The event of interest is the first interfirm job separation experienced by the sample subjects. In other words, one is interested in explaining the duration of stay with the first employer, where duration is measured in years.[2] The time-constant categorical predictor that is used in the analysis is "firm size". The first five categories of this predictor range from small firm (1) to large firm (5), and the sixth category refers to government. The two main questions to be answered are: 1) what is the time-dependence of the job separation rate – are individuals more likely to leave during the first years or after say five years – and 2) does the job separation rate depend on the size of the firm in which an individual is employed.

Let us recall the main advantages of using a hazard regression model

[2]In the analysis, the last 18 of the 31 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, I do not apply a special formula for the computation of the exposure times for the first time interval.

instead of a simple binary logistic regression model for the occurrence of the event within a predefined time period, for example, within a period of 5 years. By taking into account that a job change may occur at any of the 19 time intervals, no information is lost and the time dependence of the event can be studied. Two other advantages are that it is possible to deal with censored observations and to investigate whether the covariate effect changes over time. The last advantage – the possibility to include time-varying covariates in the model – is not exploited in the presented application.

[INSERT TABLE 6 ABOUT HERE]

[INSERT Figure 2 ABOUT HERE]

The log-likelihood values, the number of parameters, as well as the BIC[3] values for the estimated hazard models are reported in Table 6.[4] Model 1 postulates that the hazard rate does neither depend on time nor firm size and Model 2 is an exponential survival model with firm size as a categorical predictor. The large difference in the log-likelihood values of these two models shows that the effect of firm size on the rate of job change is significant. A Cox proportional hazard model is obtained by adding an unrestricted time effect (Model 3). This model performs much better than Model 2, which indicates that there is a strong time dependence. Inspection of the estimated time dependence of Model 3 shows that the hazard rate rises in the first time periods and subsequently starts decreasing slowly (see Figure 2). Models 4 and 5 were estimated to test whether it is possible to simplify the time dependence of the hazard rate on the basis of this information. Model 4 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 5 is the same as Model 4 except for that it contains a linear term to describe the negative time dependence after the second time point. The comparison between Models 4 and 5 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 5 with the less restricted Model 3 and the more restricted Model 2 shows that Model 5

[3]BIC is defined as minus twice the log-likelihood plus $\ln(N)$ times the number of parameters, where N is the sample size (here 1782).

[4]A more extended analysis in which also models with unobserved heterogeneity are estimated is presented in Vermunt (2002).

captures the most important part of the time dependence. Though according to the likelihood-ratio statistic the difference between Models 3 and 5 is significant, Model 5 is the preferred model according to the BIC criterion. Figure 2 shows how Model 5 smooths the time dependence compared to Model 3.

The log-linear hazard parameter estimates for firm size obtained with Model 5 are 0.51, 0.28, 0.03, -0.01, -0.48, and -0.34, respectively.[5] These show that there is a strong effect of firm size on the rate of a first job change: The smaller the firm the more likely an employee is to leave the firm or, in other words, the shorter he will stay. The hazard ratio comparing a small firm (category 1) with a large firm (category 5) equals $\exp[0.51-(-0.34)] = 2.34$, which means that the hazard rate is more than two times larger for the former category. Government employees (category 6) have a slightly higher (less low) hazard rate than employees of large firm (category 5).

10 Final remarks

This chapter discussed the most important concepts and statistical methods for event history analysis in continuous and discrete time. It was stressed that these methods have important advantages compared to alternative approaches when the aim of a study is to determine the factors affecting the duration of non-occurrence of a particular event. I also demonstrated that – after some restructuring of the data – the most important regression models for event history data can be estimated using standard Poisson or logistic regression analysis software.

Two topics that were not discussed in detail are left censoring, which is somewhat more difficult to deal with than right censoring, and more extended models for discrete-time data, such as models for competing risks and multiple events, which can be estimated using multinomial and multilevel logistic regression analysis software, respectively. Other more advanced topics that have received attention in the recent statistical literature on event history analysis are models for unobserved heterogeneity that is correlated with the observed covariates, for missing data on covariates, for covariates containing measurement error, for states measured with errors, and for the simultaneous analysis of event and covariate processes.

[5] Very similar estimates are obtained with Model 3. Moreover, note that I used effect coding (these parameters sum to 0).

References

Aalen, O. O. (1974). Non parametric inference for a family of counting processes. *Annals of Statistics*, 6, 701-726.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. S. Leinhardt (ed.), *Sociological Methodology 1982*, 61-98. San Francisco: Jossey-Bass.

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

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

Breslow, N. (1972). Contribution to the discussion on paper of Cox (1972). *Journal of the Royal Statistical Society B*, 34, 187-220.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89-100.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society B*, 34, 187-220.

Heckman, J. J., and Singer, B. (1982). Population heterogeneity in demographic models. In: Land K and Rogers A. (eds.), *Multidimensional mathematical demography*. New York, NY: Academic Press.

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

Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.

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 econometric analysis of transition data*. Cambridge: Cambridge University Press.

Nelson W., (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14, 945-966.

Singer J. D. & Willet J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford: Oxford University Press.

Tuma, N. B., and Hannan, M. T. (1984). *Social dynamics: models and methods.* New York, NY: 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. (1997). Log-linear models for event history histories. *Advanced Quantitative Techniques in the Social Sciences Series, Volume 8.* Thousand Oakes, CA: Sage.

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

Vermunt, J. K., and Magidson, J. (2005). *Latent GOLD 4.0 User's Guide.* Belmont, MA: Statistical Innovations.

Yamaguchi, K. (1991). Event history analysis. *Applied Social Research Methods, Volume 28.* Newbury Park, CA: Sage.

Table 1: Example of Kaplan-Meier and Nelson-Aalen computations

l	t_l	n_l	d_l	d_l/n_l	$S_{KM}(t_l)$	$H_{KM}(t_l)$	$h_{KM}(t_l)$	$S_{NA}(t_l)$	$H_{NA}(t_l)$	$h_{NA}(t_l)$
1	3	5	1	0.20	0.80	0.22	0.07	0.82	0.20	0.07
2	4	4	1	0.25	0.60	0.51	0.29	0.64	0.45	0.25
3	7	2	1	0.50	0.30	1.20	0.23	0.39	0.95	0.17

Table 2: Illustration of data organization in Poisson regression based hazard modeling

id	t_i	δ_i	t_{l-1}	t_l	y_{il}	e_l	l	x_i
1	8	1	0	6	0	6	1	0
1	8	1	6	12	1	2	2	0
2	16	0	0	6	0	6	1	1
2	16	0	6	12	0	6	2	1
2	16	0	12	18	0	4	3	1

Table 3: Illustration of data organization in Poisson regression based hazard modeling with time-varying covariates

id	t_i	δ_i	t_{l-1}	t_l	y_{il}	e_l	l	x_{1i}	x_{2it}
1	8	1	0	2	0	2	1	0	0
1	8	1	2	6	0	4	1	0	1
1	8	1	6	12	1	2	2	0	1
2	16	0	0	6	0	6	1	1	1
2	16	0	6	7	0	1	2	1	1
2	16	0	7	12	0	5	2	1	0
2	16	0	12	18	0	4	3	1	0

Table 4: Illustration of data organization in Poisson regression based hazard modeling with time-varying covariates and competing risks

id	t_i	δ_i	t_{l-1}	t_l	y_{il}	e_l	l	d	x_{1i}	x_{2it}
1	8	0	0	2	0	2	1	1	0	0
1	8	0	2	6	0	4	1	1	0	1
1	8	0	6	12	0	2	2	1	0	1
1	8	1	0	2	0	2	1	2	0	0
1	8	1	2	6	0	4	1	2	0	1
1	8	1	6	12	1	2	2	2	0	1
2	16	0	0	6	0	6	1	1	1	1
2	16	0	6	7	0	1	2	1	1	1
2	16	0	7	12	0	5	2	1	1	0
2	16	0	12	18	0	4	3	1	1	0
2	16	0	0	6	0	6	1	2	1	1
2	16	0	6	7	0	1	2	2	1	1
2	16	0	7	12	0	5	2	2	1	0
2	16	0	12	18	0	4	3	2	1	0

Table 5: Hazard rates illustrating the effect of unobserved heterogeneity

time	$X_1 = 0$			$X_1 = 1$			ratio between
point	$X_2 = 0$	$X_2 = 1$	marginal	$X_2 = 0$	$X_2 = 1$	marginal	$X_1 = 1$ and $X_1 = 0$
0	.010	.050	.030	.020	.100	.060	2.00
10	.010	.050	.026	.020	.100	.045	1.73
20	.010	.050	.023	.020	.100	.034	1.50
30	.010	.050	.019	.020	.100	.027	1.39

Table 6: Test results for the job change example (T=time and F=firm size)

Model	log-likelihood	# parameters	BIC
1. $\{\}$	-3284	1	6576
2. $\{F\}$	-3205	6	6456
3. $\{T, F\}$	-3024	24	6249
4. $\{T_1, T_2, F\}$	-3205	8	6471
5. $\{T_1, T_2, T_{lin}, F\}$	-3053	9	6174

Figure 1: Kaplan-Meier survival function

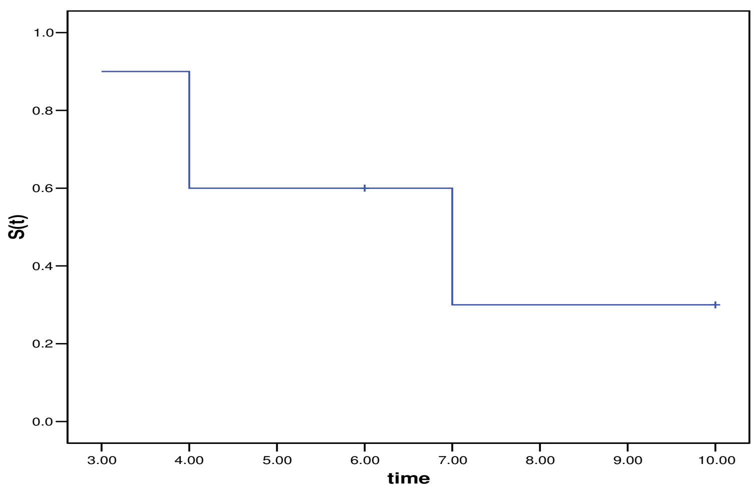

Figure 2: Time dependence according to Model 3 and Model 5