

10 Event-history models in social mobility research

HANS-PETER BLOSSFELD, ALFRED HAMERLE AND
KARL ULRICH MAYER

INTRODUCTION

The analysis of social mobility is probably the field of sociology that has been most affected by methodological and theoretical developments in the period since the 1950s (see, e.g., Sørensen, 1986). In classical mobility studies Rogoff (1953), Glass (1954), Svalastoga (1959), and Carlsson (1958) studied social inequality on the basis of *mobility tables*. These tables were obtained from cross-sectional samples of men who were asked about their current position and about their father's position when they grew up. However, as respondents had different ages at the time of the interview and therefore had different historical experiences, these efforts did not provide very interpretable results (Sørensen, 1986).

Mobility tables were sometimes disaggregated by birth cohorts (Featherman & Hauser, 1978; Goldthorpe, 1980), but this does not alleviate the problem because careers take place in the labor force. This means the relevant period is the amount of time spent in the labor force and not the age of the people (Blossfeld, 1986). Members of the same birth cohort are at different career stages because they spent different times in the educational system. Thus, positions reflected by mobility tables of birth cohorts are observations of locations of people at different stages in their career (Sørensen, 1986).

The typical mobility table not only ignores differences in the historical location of cohorts and the amount of labor force experience of people, but also neglects the fact that job mobility is strongly influenced by education and other background variables of the individual. A first solution to this problem was provided by Blau & Duncan (1967) with the introduction of *path analysis* using linear regressions. This research induced a major change in methodology in the late 1960s and early 1970s. At that time, the use of regression analysis and the formulation of causal models in sociology was synonymous with *status attainment research* (Sørensen, 1986). This research was very successful and employed a richer set of variables than work based on classical mobility tables. But the basic design for collecting the data remained the same. This research was based on

cross-sectional surveys. Thus, the defects of the data continued. Status attainment models also ignored the temporal nature of the mobility process. The effects of variables were established whether the current occupation of the respondents was the first or the one they had been holding for 50 years in the labor market (Sørensen, 1986).

Since the late 1970s, *labor market researchers* have increasingly drawn attention to the fact that the observed occupational attainment of an individual at a specific time is the result of a process of change. They have developed continuous-time models to conceptualize individual mobility processes (see, e.g., Sørensen, 1977, 1984; Sørensen & Tuma, 1981; Spilerman, 1977; Tuma, 1985; Carroll & Mayer, 1986; Blossfeld, 1986). The increasing availability of *event history data* on job moves in the 1980s has permitted the application of statistical techniques to estimate such models (Tuma & Hannan, 1984; Blossfeld, Hamerle & Mayer, 1989). Using these methods, substantial progress has been made in understanding the mechanisms of individual career processes, especially in understanding the role played by individual characteristics (e.g., education, labor force experience, etc.) and structural attributes (e.g., occupations, cohorts, etc.) in shaping career trajectories.

In this chapter we give a didactic example of how job mobility can be studied as a continuous process with event-history data. We do this in four steps.

1. We discuss the advantages of the event-oriented collection design for the analysis of job mobility.
2. We describe the German Life History Study which is used as the empirical base of the example.
3. We discuss some statistical aspects of event-history analysis necessary for an adequate understanding of the empirical analysis.
4. We present the results of the analysis of job mobility as a continuous process. The chapter concludes with a short discussion of potentialities and limitations of event-history analysis in social science research.

ADVANTAGES OF EVENT-HISTORY DATA

Compared with cross-sectional or panel data, event-history data have several advantages for studying mobility processes. A simple example may illustrate these advantages. Suppose that data have been collected for an individual with regard to education and occupation with the aid of a cross-sectional sample, a panel, and an event-oriented sample design (Figure 10.1). The individual's career path is differentiated into seven states (training, occupation 1, occupation 2, occupation 3, occupation 4, unemployment, and illness) which the individual may occupy.

First, looking at Figure 10.1 one observes that in a cross-sectional survey the educational and occupational history of a person is only represented by a single point, that being the state at the time of the interview. Somewhat more information is obtained by the four-wave panel in which the circumstances of the respondent can be observed at four different points in time. However, the career between the four waves of the panel remains unclear. It is only in the *event-oriented collection design* that changes in states and their precise times are explored. Such a design allows the educational and occupational career to be reconstructed in detail in its various phases and at any point in time.

This example shows the following.

- As a rule, cross-sectional analysis presupposed a steady state (i.e., the distribution at any given point in time is only informative if the underlying process remains relatively stable over time). In cases of major fluctuations and changes, the 'snapshot' of a cross-section will not be a good picture of the situation because the analysis will depend upon the specific conditions prevailing at the time of survey. In contrast, panel and event-oriented data explicitly take into account change and the dynamics of empirical phenomena.
- Even if empirical conditions are predominantly stable, panel and event-history data are more informative than cross-sections. Cross-sectional data can be regarded as a special case of panel and event-history data because cross-sections can be reconstructed from the latter. Moreover, in cases of empirical application, only the recording of panel or event-history data can demonstrate whether stability really exists over time. Finally, unlike cross-sections, panel and event-history data provide information on prior history which can help to improve the explanatory and prognostic capacity of statistical models.
- Whereas in the panel method the course of events between the individual survey points remains unknown, the event-oriented observation plan permits the reconstruction of the continuous process. The panel method may also be suitable to determine the course of events if the changes take place at clearly defined points in time coinciding with the survey intervals (e.g., the determination of yearly income on a yearly basis) or if a continuous variable (e.g., a person's weight) can only be appropriately observed on the basis of time discrete surveys. Yet, all other changes in qualitative variables that may occur at any point in time can only be fully reconstructed if the states and time of their changes are exactly registered. Therefore, the

event-oriented observation plan proves to be a necessary precondition for the adequate reconstruction of mobility processes.

- Finally, if one considers the dynamic analysis of *complex feedback processes* (e.g., between processes in the family and processes in the labor market), the continuous survey of qualitative variables would seem to be the only adequate method to assess empirical change. This is particularly true if the events of parallel processes occur not only at arbitrary points in time, but also have an interactive effect at a later stage.

In sum, the major advantage of an event-oriented observation plan in mobility research is the fact that it permits an adequate representation of changes in careers which may occur at any point in time.

DATA

As an example of an application of event-history analysis in mobility research, we study the process of job change for men in the Federal Republic of Germany. The analysis is based on data from the German Life History Study (GLHS).

The GLHS (Mayer *et al.*, 1987) is useful for this purpose since it provides detailed data about the life histories of 2171 respondents from the birth cohorts 1929–31, 1939–41, and 1949–51, collected in the years 1981–3. These birth cohorts were chosen so that the respondents' phase of transitions from school to work fell in particularly significant periods in history: for the 1930 cohort, this transition phase lies in the immediate postwar period; those born around 1940 left school in a time of large-scale economic growth, and the cohort 1949–51 entered the labor market during a phase marked by the expansion of the welfare state. The underlying hypothesis is that these specific historic conditions at the point of transition had a substantial impact on the respondents' subsequent careers.

The educational and occupational histories of the GLHS were recorded retrospectively in accordance with the event-oriented observation plan. This method is demonstrated by an extract from the questionnaire where respondents were interviewed about their work careers (Figure 10.2). It is characteristic that apart from collecting theoretically interesting information about the area of employment, number of working hours, income, and so on, the exact beginning and end of each job were recorded on a monthly basis. When this information about the sequence of job episodes is combined with records of periods of training and interruption, the educational and occupational history of an individual can be completely reconstructed. Such an event-oriented observation plan provides detailed information

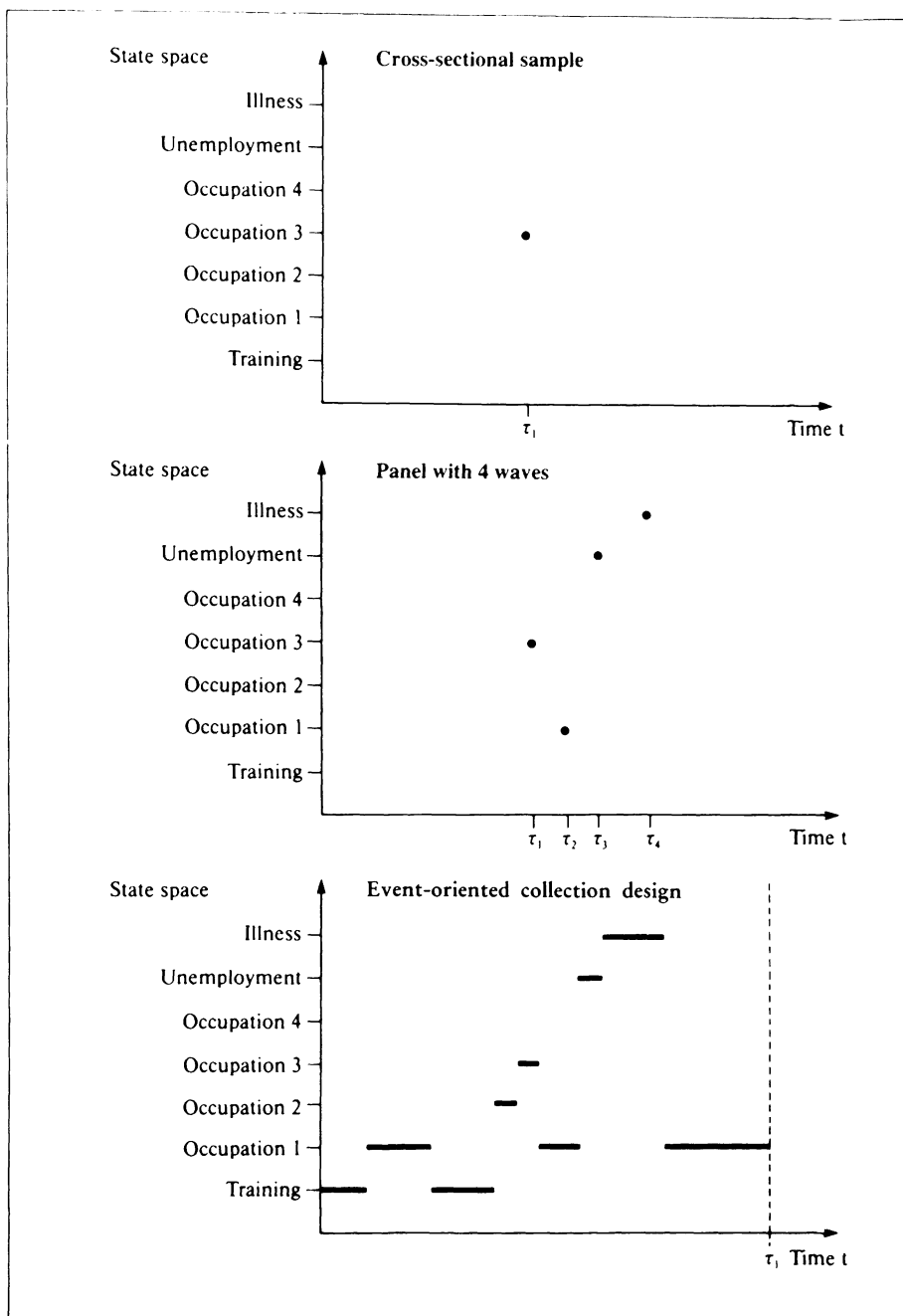


Figure 10.1. Example of an event-oriented observation plan to record work careers.

on the states of a given respondent's career at any point in the period of observation.

A study conducted before the actual drawing of the GLHS sample demonstrated that the reliability of retrospectively recorded data about objective life histories is not systematically affected by a lack of ability to answer questions or deficient memory capacity (Papastefanou, 1980; Tölke, 1980). This study indicated that while the possibility of recall errors was minimal, the form and precision of the survey instrument proved to be of key importance with respect to the quality of the responses. In particular, it was important to divide the life history interview into different spheres of life (education/training, employment, residence, etc.). Lengthy and extensive data editing, data checks, and cross-comparisons also vouched for the quality of the collected information (Mayer & Brückner, 1989). Finally, an examination of the representative quality of the life history data on the basis of census and microcensus surveys shows that the GLHS data provide a reliable picture of sociostructural cross-sections of the past (Blossfeld, 1987).

Because the data cover not only educational and occupational histories, but also provide information on the whole spectrum of the various spheres of life (i.e., information on social background, family history, the spouse's history, residence history, etc.), it is possible to study the effects that events in other parallel processes (e.g., in the case of family history, the event 'marriage') have on occupational careers (e.g., 'stability' of occupational trajectories). Similarly, prior history can be analyzed to examine the extent to which the subsequent career has been predetermined and channelled in certain directions.

A wide range of statistical tools is available today to analyze event-history data. In this chapter we (1) present some fundamental statistical concepts of event history analysis, (2) discuss methods to estimate the unknown parameters, and (3) give an example of the application of event-history analysis to study job changes.

FUNDAMENTAL STATISTICAL CONCEPTS OF EVENT-HISTORY ANALYSIS

The simplest example of an event-history analysis is characterized solely by measurement of the entrance into some initial state until attainment of some final state. The duration of an episode is represented in the statistical model by a nonnegative stochastic variable T . If time is exactly measured, the variable T is a continuous stochastic variable.

Survivor function and hazard rate

The density and distribution function of the duration T ($T \geq 0$) are denoted by $f(t)$ and $F(t)$, respectively. As usual the following relation-

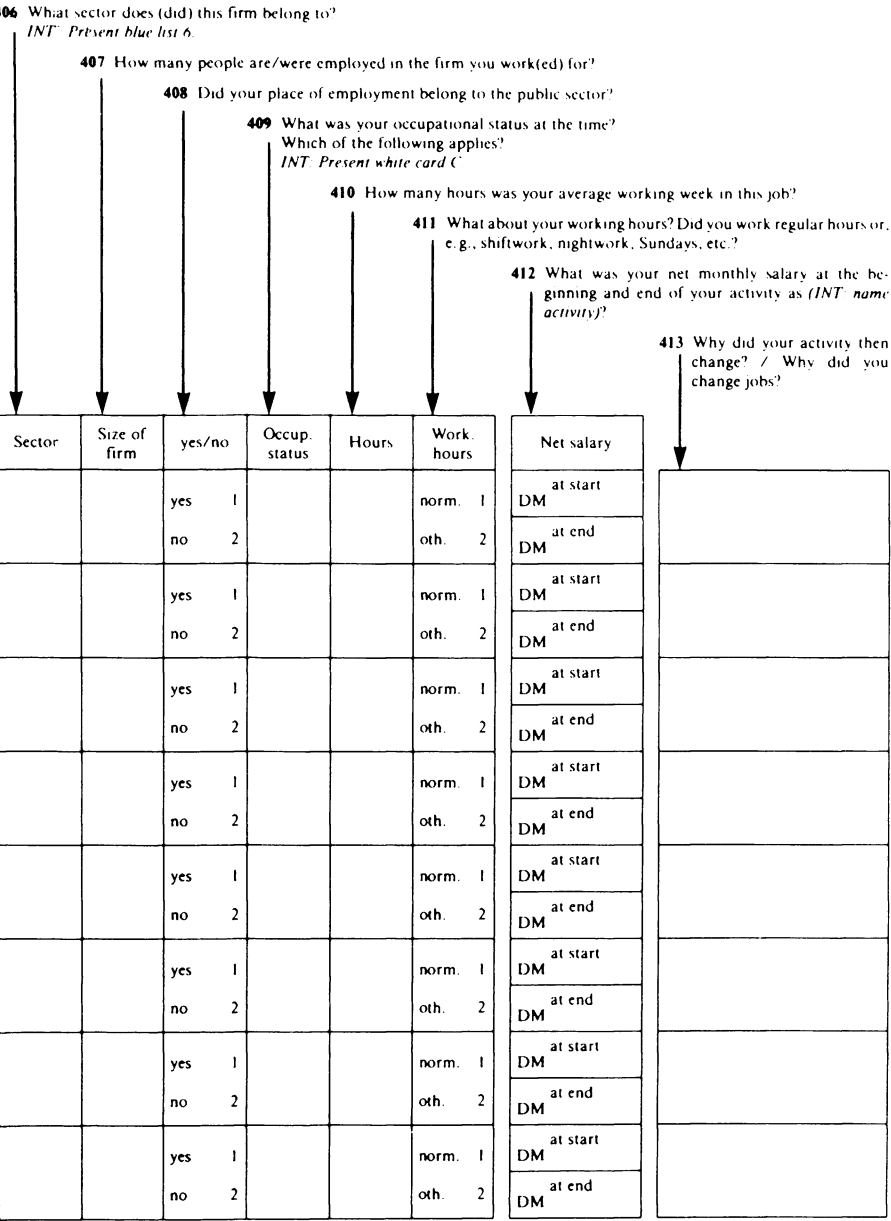


Figure 10.2. Recording of a person's educational and occupational career on the basis of a cross-sectional sample, a panel and an event-oriented collection design.

ship is valid:

$$F(t) = P(T \leq t) = \int_0^t f(u) \, du, \quad (1)$$

and for all points for which $F(t)$ may be differentiated

$$f(t) = F'(t). \quad (2)$$

The *survivor function*

$$S(t) = P(T \geq t) \quad (3)$$

expresses the probability that an individual remains in the state ('survives') until time t , that is, that an event has not yet occurred and the episode is still continuing.

Measuring waiting time continuously, we have

$$S(t) = 1 - F(t). \quad (4)$$

Therefore, the survivor function is a non-increasing function of time, approaching zero as time elapses.

The *hazard rate* (hazard function or failure rate) is defined as

$$\lambda(t) = \lim_{\substack{\Delta t \rightarrow 0 \\ \Delta t > 0}} \frac{1}{\Delta t} P(t \leq T < t + \Delta t \mid T \geq t). \quad (5)$$

The hazard function may be interpreted as the instantaneous probability that episodes in the interval $[t, t + \Delta t]$ are terminating provided that the event has not occurred before the beginning of this interval. Common terminology often found in application of the hazard function are *intensity* or *risk functions*, *transition* or *mortality rates*.

It is important to note that the values of the hazard function themselves are not (conditional) probabilities. Although they are always nonnegative, they may be greater than one. For a small Δt , $\lambda(t)\Delta t$ can be interpreted as an approximation of the conditional probability $P(t \leq T < t + \Delta t \mid T \geq t)$.

The *cumulative hazard function* is represented by the integral

$$\Lambda(t) = \int_0^t \lambda(u) \, du. \quad (6)$$

From definition (5), one immediately obtains the relationship between the hazard rate and the survivor function

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad (7)$$

and since T has been assumed to be continuous we have

$$\lambda(t) = \frac{f(t)}{1 - F(t)}. \quad (8)$$

Inversely, one may derive the relationship between the survivor function and the hazard function by integration of $\lambda(t)$. From (7) and (8), we have

$$\begin{aligned}\int_0^t \lambda(u) du &= \int_0^t \frac{f(u)}{1 - F(u)} du = -\ln(1 - F(u)) \Big|_0^t \\ &= -\ln(1 - F(t)) = -\ln S(t).\end{aligned}\quad (9)$$

This leads to the important relationship

$$S(t) = \exp\left(-\int_0^t \lambda(u) du\right). \quad (10)$$

The density $f(t)$ is obtained from (7) and (10) as a function of the hazard function

$$f(t) = \lambda(t) \cdot S(t) = \lambda(t) \cdot \exp\left(-\int_0^t \lambda(u) du\right). \quad (11)$$

If we consider the relationships (1) to (11) it becomes evident that each of the three quantities $f(t)$, $S(t)$, and $\lambda(t)$ may be used to describe the duration of an episode. If one of these functions is known, the derivation of both the other functions is always possible. In particular, if one knows the hazard rate, the probability law of the process is completely characterized.

But the hazard rate approach does not identify new parameters. As we shall see, for estimation of the unknown parameters the likelihood function can be written equivalently in terms of hazard rates of probability density functions, but it is the same likelihood function in either case.

One of the most commonly applied distributions for the waiting time and lifetime is the exponential distribution. It is characterized by a constant hazard rate

$$\lambda(t) = \lambda, \quad t \geq 0, \lambda > 0.$$

A generalization of the exponential model is the Weibull model which allows for duration dependence. Its hazard function is

$$\lambda(t) = \lambda \alpha (\lambda t)^{\alpha-1} \quad (t > 0). \quad (12)$$

with the parameters $\lambda > 0$ and $\alpha > 0$. In the special case $\alpha = 1$, equation (12) reduces to the hazard rate of the exponential model.

The hazard or transition function of the Weibull distribution increases monotonically if $\alpha > 1$, decreases if $\alpha < 1$, and is constant if $\alpha = 1$. The Weibull model is quite flexible and therefore adaptable to a wide variety of models of durations and lifetimes.

The hazard rate approach can be readily extended to the multistate-multiepisodic case. In this case, the process can be decomposed into two

related processes, the transition process and the duration process. The transition process is the stochastic process which governs the transitions between states. The duration process governs the length of stay in a particular state. Both the duration and the transition process can be simultaneously characterized by a transition-specific or cause-specific hazard rate. We do not give the exact definition here.

Introducing covariates: regression models

In addition to the duration or lifetime, generally, various covariates or prognostic factors are collected for each individual or subject, and an important aim of statistical analysis is to ascertain the quantitative influence of these exogenous or endogenous variables on the hazard or transition rate. The covariates may be quantitative or qualitative. Qualitative variables are treated as in analysis of variance by coding categories using dummy variables.

An important difference compared with cross-sectional regression is that in the duration models some covariates may be time dependent. Such is the case, for example, when a specific medical therapy is applied only during a certain time period. The aim of such a study might be to examine the influence of the therapy during the effective application period or to check for possible secondary effects (side-effects). To accomplish this, for example, two dummy variables may be defined, $x_1(t)$ and $x_2(t)$ with (dropping subscript i)

$$x_1(t) = \begin{cases} 1 & \text{during the time period of active participation} \\ & \text{of an individual in a therapy or program,} \\ 0 & \text{otherwise} \end{cases}$$

$$x_2(t) = \begin{cases} 1 & \text{after the individual's termination of the 'medical} \\ & \text{treatment' within a specific therapy or program,} \\ 0 & \text{otherwise.} \end{cases}$$

If the hazard rate depends on the independent variables and given that the corresponding regression coefficients are negatively (positively) significant, a therapy is then effective and lowers (increases) the instantaneous probability of a change of state. Furthermore, if the first coefficient is larger in absolute value than the second, then the effect of the therapy sinks (rises) after termination of the therapy.

One possible method of analyzing the influence of covariates upon the duration times or lifetimes is to formulate a regression model in which the distribution of the duration times or lifetimes are dependent upon the covariates. Designating \mathbf{x} as the covariate vector, one has to specify a model for the duration times or lifetimes T under the given covariate vector \mathbf{x} .

Obviously, an analogous procedure to traditional regression analysis may be chosen and the above introduced distributions, such as the exponential or Weibull distribution, may be generalized such that one or more parameters are assumed to depend on the covariates x .

The values of the quantitative covariates of an individual or study subject i , as well as the codings of all main effects and the interaction effects of the qualitative covariates included in the model are then collected in a design vector x_i .

Since in the analysis of durations or event histories the hazard rate is the mathematically simpler concept, the hazard rate is modeled depending on the covariates.

For the exponential model the hazard rate is given by

$$\lambda(t | x) = \exp(x' \beta). \quad (13)$$

(13) is time invariant provided that the covariates do not depend on time.

The hazard rate of the Weibull regression model is

$$\lambda(t | x) = \delta \lambda_0(\lambda_0 t)^{\delta-1} \exp(x' \beta). \quad (14)$$

The Weibull model belongs to the class of proportional hazards models, since the quotient of the hazard rates for two individuals is time invariant.

An extension is the semi-parametric approach originally proposed by Cox (1972). The hazard rate of the Cox model is

$$\lambda(t | x) = \lambda_0(t) \exp(x' \beta) \quad (15)$$

where $\lambda_0(t)$ is an unspecified 'baseline' hazard rate. Such modeling achieves more flexibility, but different approaches are required for estimation and testing.

ESTIMATION

After the construction of a statistical model for the event history under discussion, the unknown parameters have to be estimated from the data. In applying an estimation method the problem of possibly censored observation has to be solved.

Since in event-history analysis the termination of the entire observation time period is given, a time interval may not be closed. In such a situation we have right censored data. The sample realization t_i of an individual simply states then that the duration of an episode is of at least t_i time units. Usually the sample consists of some t_i values which are complete durations, whereas the remaining t_i values are right censored. This is expressed with the aid of a *censoring indicator* δ_i as

$$\delta_i = \begin{cases} 1 & \text{if } t_i \text{ is not censored} \\ 0 & \text{if } t_i \text{ is censored, } i = 1, \dots, n. \end{cases}$$

The possibility of simply ignoring the censored data and thereby reducing the size of the sample is not recommendable since this generally leads to biased estimates.

The maximum likelihood method offers the possibility of considering explicitly right censored data within the estimation procedure. Accordingly, one must analyze in depth the underlying censoring mechanism of the data and formulate a statistical model.

A widely used approach is to assume the durations T_i and the censoring times C_i are independent random variables and, in addition, that the distribution of the censoring time does not depend upon the parameters determining the duration distribution. Then, the likelihood function is

$$L = c \cdot \prod_{i=1}^n f_i(t_i/x_1)^{\delta_i} S_i(t_i | x_1)^{1-\delta_i}$$

and utilizing the relationship between hazard rate and survivor function, one obtains

$$L = \prod_{i=1}^n \lambda_i(t_i | x_i)^{\delta_i} \exp\left(-\int_0^{t_i} \lambda_i(u | x_i) du\right). \quad (16)$$

Using a special parametric model for the hazard rate, (16) or the log-likelihood function can be maximized with respect to the unknown parameter. Usually this is done by applying an iterative method such as the Newtonian method or a modified Newtonian method.

Hypothesis testing and construction of confidence intervals can be done in the usual way, based on the asymptotic distribution of the maximum likelihood estimator. The asymptotic properties of the maximum likelihood estimator, in particular consistency and asymptotic normality, can be derived elegantly using results from the theory of counting processes (for proportional hazard models). See, for example, Andersen & Gill (1982).

For the Cox model another estimation procedure is necessary, since the likelihood function depends on the unknown 'unisance' function $\lambda_0(t)$. Cox (1972, 1975) proposed a partial likelihood method. For a detailed description of this estimation procedure see, for example, Kalbfleisch and Prentice (1980, Ch. 4), Lawless (1982, Ch. 7), or Blossfeld, Hamerle & Mayer (1989, Section 3.6).

Regressor variables whose values change over the course of time intervals are conceptually straightforward to handle in the hazard rate framework. No analogue in cross-sectional regression is available. Suppose the regressor x is a function of time $x(t)$, where t is measured from the beginning of the time interval. Write the hazard rate as $\lambda(t | x(t))$. Using our integration formulas we can write the integrated

hazard, survivor, and density functions. These will in general depend on the entire time path of the regressor up to t . Owing to the linearity of the integral, the calculation is simplified if the regressors are step functions over time. An application of this simpler case is presented in the next section.

There are many extensions and generalizations, for example, to the inclusion of several states (competing risks) or multiple time intervals of an individual, or with regard to residual analysis and specification checking. For further details see, for example, Kalbfleisch & Prentice (1980), Cox & Oakes (1974), Blossfeld, Hamerle & Mayer (1989), Anderson & Borgan (1985), Clayton (1988), or Manton & Stallard (1988).

A problem to which recently a lot of attention has been given is unobserved heterogeneity. On this point it has always been assumed that the measured covariates completely determine the hazard rate. But there may be omitted variables which are important and influence the hazard rate as well. In the duration model the unobserved heterogeneity can be represented by a random variable v . A convenient approach is

$$\lambda(t \mid x, y) = \lambda(t \mid x) \cdot v.$$

Since the hazard rate is not negative, v must be limited to positive values. The 'deviation' v varies from individual to individual and is not observable. In demographic and medical studies these models are referred to as 'frailty' models (see Vaupel, Mauton & Stallard, 1979). If a parametric form for the distribution of unobserved heterogeneity is assumed, one can estimate the unknown parameters of this distribution along with the unknown parameters of $\lambda(t \mid x)$ using a 'marginal likelihood' integrating out the heterogeneity. For a more detailed treatment see, for example, Blossfeld, Hamerle & Mayer (1989), Aalen (1988), or Heckman & Singer (1984).

RESULTS

The German Life History Data can be used to estimate the job change among men as being dependent on education (EDU), prestige (PRES), number of previously held jobs (NOJ) labor force experience at the beginnings of each job (LFX), as well as dummy variables for the birth cohorts 1939–41 (COHO2) and 1949–51 (COHO3) (see Appendix) with an exponential model:

$$\lambda(t \mid x_k) = \exp(x' \beta), \quad t = \text{duration}.$$

The exponential model rests on the assumption of a constant event risk and thus implies the assumption of proportional risks over the duration. This model lends itself to easy interpretation and is commonly

applied as the basic or reference model to which estimates of more complex distribution models are compared. The maximum likelihood estimation for this model is calculated using the program RATE (developed by Nancy Tuma, 1985). The results of this estimation are reported as Model 1 in Table 10.1.

First of all, if one compares this exponential model with the exponential model without covariates, then based upon the likelihood ratio test a chi-square value of 705.90 with six degrees of freedom is obtained. The included covariates thus contribute to explaining the risk of men changing jobs, and the null hypothesis which states that none of the introduced β coefficients is different from zero must be rejected.

A significance test may be implemented for the individual regression parameters by dividing the coefficients $\hat{\beta}_i$ by their estimated asymptotic standard errors $s(\hat{\beta}_i)$. Assuming the hypothesis $H_0: \beta_i = 0$, these test values are approximately characterized by a standard normal distribution. If one uses a 0.05 significance level and a two-sided test, then the covariates have a significant effect if the absolute values of their standardized coefficients satisfy

$$\left| \frac{\hat{\beta}_i}{s(\hat{\beta}_i)} \right| > 1.96.$$

This is the case for the constant β_0 (CONST) as well as for the variables PRES, NOJ, LFX, COHO2, and COHO3. Only education (EDU) does not have a significant influence upon the rate of job change among men.

One can easily interpret the influence of a covariate x_i , given the constancy of the remaining covariates $\mathbf{x}'\boldsymbol{\beta}$, by demonstrating the percentage change in the rate given an increase in the covariate x_i by a specific value Δx_i . For example, an increase in the number of previously held jobs (NOJ) of 1 unit results in an increase of the rate equivalent to about 19% $[(\exp(0.171426) - 1) \cdot 100\% = 18.70\%]$. On the other hand, an increase in prestige (M59) of 20 units leads to a decline in the inclination to change jobs of about 10% $[(\exp(-0.00521)^{20} - 1) \cdot 100\% = -9.9\%]$. A simultaneous change of NOJ by 1 unit and prestige (M59) by 20 units, which would represent a career advance, raises the rate by only about 7% $[(\exp(0.171426)^1 \cdot \exp(-0.00521)^{20} - 1) \cdot 100\% = 6.95\%]$ and not of the magnitude of about 9% $[18.70\% - 9.9\% = 8.80\%]$.

Applying the relationship

$$E(T | \mathbf{x}) = \frac{1}{\lambda(\mathbf{x})} = \frac{1}{\exp(\mathbf{x}'\boldsymbol{\beta})},$$

given an exponential distribution, one may explicitly state, how, when all other covariates are constant, the average duration $E(T | \mathbf{x})$ changes if

one increases the value of the independent variable x_i by the amount Δx_i :

$$\delta_{\Delta x_i} = \frac{\frac{1}{\exp(x'\beta + \beta_i(x_i + \Delta x_i))} - \frac{1}{\exp(x'\beta + \beta_i x_i)}}{\frac{1}{\exp(x'\beta + \beta_i x_i)}} \cdot 100\%$$

$$\delta_{\Delta x_i} = \left(\frac{1}{\exp(\beta_i \Delta x_i)} - 1 \right) \cdot 100\%.$$

Increasing the number of previously held jobs (NOJ) by 1 accordingly results in a decrease in average employment duration by about 16% $[(1/\exp(0.171426) - 1) \cdot 100\% = -15.75\%]$. On the other hand, an increase in prestige (M59) by 20 units results in an increase in employment duration by about 11% $[(1/\exp(-0.00521)^{20} - 1) \cdot 100\% = 10.98\%]$. A simultaneous change of NOJ by 1 unit and prestige (M59) by 20 units, which again would represent a career jump, lowers the average duration by about 7% $[(1/(\exp(0.171426)^1 \exp(-0.00521)^{20}) - 1) \cdot 100\% = -6.5\%]$ rather than by about 5% $[10.98\% - 15.75\% = -4.77\%]$.

For selected subgroups, prognoses may also be made concerning the average duration, the median of the duration, the average number of events occurring in a given time period, and the probability of remaining in the same state at a given point in time.

For example, if one observes a man from the cohort 1939–41 (COHO2 = 1 and COHO3 = 0) who is employed in an occupation valued at the 50-point prestige level (M59 = 50) and who previously worked in 10 jobs (NOJ = 10) as well as having collected 100 months of labor force experience (LFX = 100), the following forecast equation for the rate of job change is derived:

$$\begin{aligned}\hat{\lambda} &= \exp(-4.338 - 0.005 \cdot 50 + 0.171 \cdot 10 - 0.009 \cdot 100 + 0.180 \cdot 1) \\ &= 0.0274.\end{aligned}$$

Consequently, this person is characterized by an average duration of about 37 months $[1/\hat{\lambda} = 1/0.0274 = 36.496]$, which is well below the average of about 98 months. Furthermore, one can forecast the median duration of employment as $\hat{M}^* = 0.6934 \cdot 36.5 = 25.31$ months and in one year expect an average of $\hat{\lambda}v = 0.0274 \cdot 12 = 0.3288$ job changes. Finally, the probability that this individual is still employed in the same occupation about eight years is about 7% $[\hat{S}(96) = \exp(-0.0274 \cdot 96) = 0.072]$, whereas the average for all men is about 38%.

By means of further prognoses for additional subgroups, one obtains

Table 10.1. *Estimates of models for the risk of job change*

	Estimates for model:		
	1	2	3
CONST	-4.338*	-4.283*	-3.492*
EDU	0.013	0.025	0.007
PRES	-0.005*	-0.004*	-0.004*
NOJ	0.171*	0.173*	0.160*
LFX	-0.009*	-0.007*	-0.008*
COHO2	0.179*	0.159*	0.124*
COHO3	0.486*	0.415*	0.341*
MAR		-0.714*	
TDEP			0.8266
χ^2	705.9	969.9	868.7
d.f.	6	7	7

* Statistically significant at 0.05 level.

a well-differentiated picture of job changing behavior among men as well as an indication of the significance of varying factors of influence.

Model 1 in Table 10.1 reported the influence of time-independent variables. One speaks of time-independent covariates when they are measured at the beginning of episode k and their values remain unchanged over the duration $v_k = t - t_{k-1}$ (see Figure 10.3c). Time-dependent covariates may, however, change their values within episode k . Given discrete time-dependent covariates, the values remain constant over certain subintervals v_{k_i} , $v_k = \sum_{i=1}^s v_{k_i}$ (see Figure 10.3b), whereas continuous time-dependent variables may change continuously.

Time-dependent covariates are especially interesting because one may more realistically formulate the influence of the covariates on hazard rates and combine two or more parallel processes directly with one another (Blossfeld, 1986). It is not uncommon that a duration dependency arises because the observed states are aggregates of unobserved states. This is due to the simple fact that time-dependent covariates are dealt with as time-independent covariates in the model.

In general, economics and the social sciences deal with variables that do not change continuously over time. Such variables are characterized by a step function and influence duration by changing the rate within a given episode. For example, if one assumes that for men the event of marriage has a stabilizing effect on the process of job mobility (see Figure 10.3), then this relationship may be studied by introducing a time-dependent covariate.

If for an individual i , $t_{i,k-1}$ is designated as the beginning of an

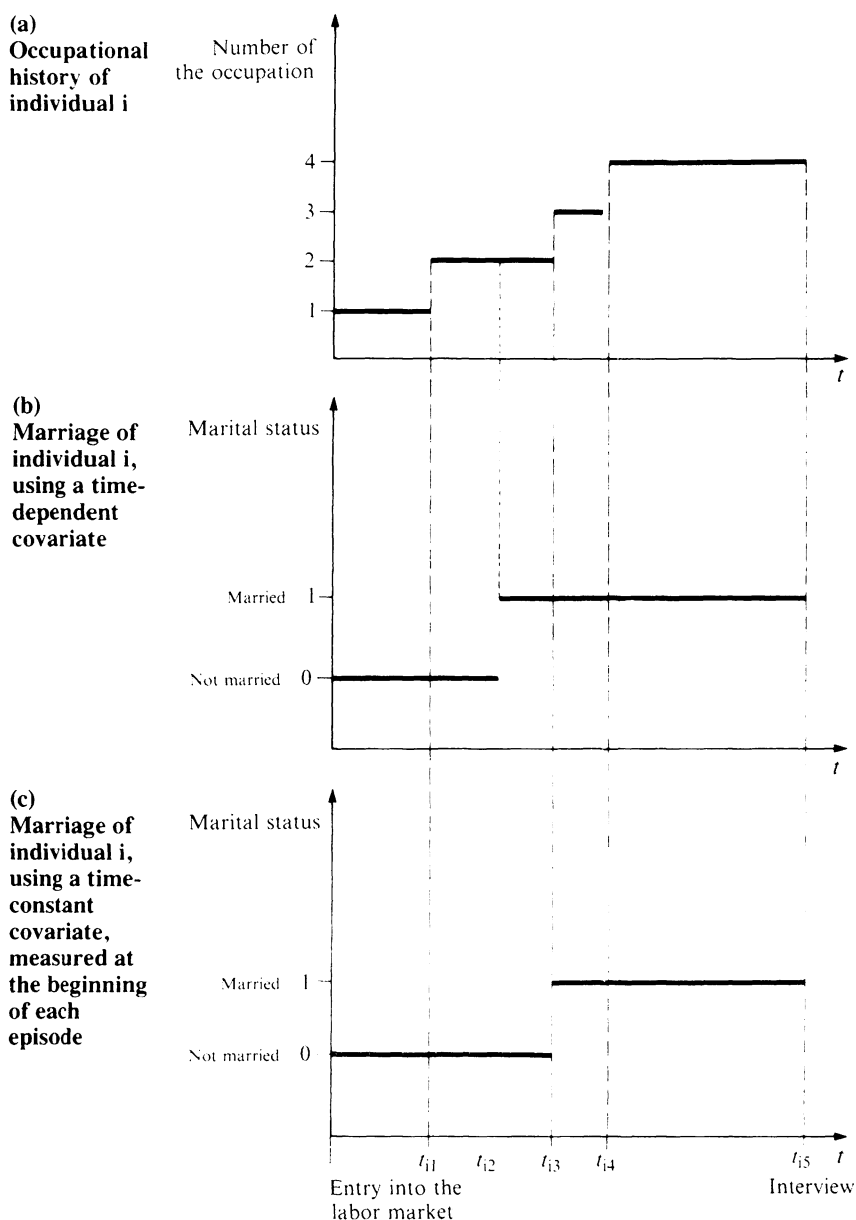


Figure 10.3. Modeling the influence of marriage (a) on the career path, (b) as a time-dependent covariate, and (c) as a time-dependent covariate measured at the beginning of each episode.

occupational episode k , and if v represents the duration of episode k , and t_i^H is the time of marriage of individual i , then the value of the time-dependent dummy variable marriage $x_{ik}^H(v)$ is characterized by:

$$x_{ik}^H(v) = \begin{cases} 0 & \text{for } t_i^H - t_{i,k-1} \geq v \\ 1 & \text{for } t_i^H - t_{i,k-1} < v. \end{cases}$$

Including discrete time-dependent covariates within a maximum likelihood estimation procedure is relatively easy. These follow a step function over time and are constant over subepisodes (see Figure 10.3b). Designating $t_0 < t_1 < \dots < t_s$ as the time points of change of the covariate vector within the duration interval $[0, t)$ and given $t_{s+1} = t$, then the cumulative hazard rate may be decomposed into an integral sum. The probability that no event occurs until the time point t is obtained from the product of the survivor functions of the subepisodes in which the covariate vector remains unchanged:

$$S(t | x(t)) = \prod_{r=1}^{s+1} S(t_r | t_{r-1}, x(t_{r-1})).$$

The maximum likelihood estimation is gained by splitting the observed duration t_i based upon the s_i time points of change into $s_i + 1$ independent subepisodes. The hazard rate is estimated as in the case of time-constant covariates. Thus, for each of these subepisodes in which the covariate vector remains unchanged, the newly created data record contains the following information:

1. the values of the covariates at the beginning of each subepisode;
2. the duration at the beginning and end of a subepisode (the duration as such is sufficient only for the exponential model);
3. censoring information indicating whether the subepisode ends with an event (CEN = 1) or not (CEN = 0).

The newly created event-history data set which has split episodes according to the time of marriage, may now be treated in the same way as for time-constant covariates. The data set is accessed by the program RATE to estimate the exponential model with the time-dependent marriage variable

$$\lambda(t | x(t)) = \exp(x'(t)\beta).$$

The results are presented in model 2 of Table 10.1. The $\hat{\beta}$ coefficient of the time-dependent covariate MAR is significant and negative, which means that the rate of changing occupations after marriage is greatly reduced. Compared with the mobility rate of unmarried men, the mobility rate of married men decreases by about 51% $[(0.4897 - 1) \cdot 100\% = -51.03\%]$.

Whereas discrete time-dependent covariates can very simply be included in parametric rate models by splitting the original durations into subepisodes characterized by constancy, this simple way does not

exist given continuous time-dependent covariates. If the continuous time-dependent covariates are assumed to be a function of the duration then a Weibull, a Gompertz-(Makeham), a log-logistic, a log-normal, or a gamma model may be used to do the estimation.

If job shifts are studied, the duration v in a job may serve as a proxy variable for job-specific knowledge and skills which have to be newly acquired in each occupation. If the hypothesis that the inclination towards changing jobs decreases with increasing accumulation of job-specific knowledge is true, then we expect that the Weibull distribution will possess a significant $\hat{\alpha}$, lying between 0 and 1.

The Weibull model is very flexible and appropriate for a wide variety of situations. Like the Gompertz distribution, the Weibull distribution can also be used to model a monotonically falling ($0 < \alpha < 1$) or monotonically increasing ($\alpha > 1$) risk. For the special case $\alpha = 1$ one obtains an exponential distribution, and it is therefore possible to test the null hypothesis of a constant event risk against the alternative $\alpha \neq 1$.

In order to include covariates in the Weibull model, the parameter λ is related to the covariate vector \mathbf{x} log-linearly: $\lambda(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta}^*)$. The Weibull model is then expressed as:

$$\begin{aligned}\lambda(v \mid \mathbf{x}) &= \exp(\mathbf{x}'\boldsymbol{\beta})^\alpha \alpha v^{\alpha-1} \\ &= \exp(\mathbf{x}'\boldsymbol{\beta}) \alpha v^{\alpha-1}\end{aligned}$$

with

$$\boldsymbol{\beta} = \alpha \boldsymbol{\beta}^*.$$

The parameter estimations are carried out with the program system GLIM (Roger & Peacock, 1983) and presented as model 3 in Table 10.1.

The estimated coefficients for EDU, PRES, NOJ, LFX, COHO2, and COHO3 of model 3 are relatively consistent with those of models 1 and 2. A test of the null hypothesis $H_0: \alpha \geq 1$ against the alternative hypothesis $H_1: \alpha < 1$ gives a monotonically decreasing inclination to change jobs:

$$z = \frac{\hat{\alpha} - 1}{s(\hat{\alpha})} = \frac{0.8266 - 1}{0.1293} = -13.41.$$

This means that the inclination to change jobs decreases monotonically with increasing job-specific knowledge.

CONCLUDING REMARKS

The major advantage of event-history data is the fact that they permit an adequate representation of changes in qualitative variables which may occur at any point in time. The question nevertheless remains: Why has the event-oriented observation plan thus far only seldom been used in economic and social science research?

One explanation can certainly be found in the extensive and costly observation procedures necessary to record event histories. One way of doing this is to observe the process and follow the development of the characteristics of individuals with the survey instrument over a lengthy period of time. However, it then often takes a long time before the data are finally available and theory has sometimes developed in a different direction. Event-history data are therefore often collected retrospectively. As was also the case in the GLHS, the history of events is thus reconstructed over a long period of time. This type of data collection is sometimes the only way of obtaining event-oriented information (e.g., today, the educational and occupational careers of the 1929–31, 1939–41, and 1949–51 birth cohorts can only be recorded on a retrospective basis). However, such data are often criticized as being unreliable, in particular when the events to be recalled took place in the distant past. Retrospective recording of event-history data therefore requires a greater degree of care and control and this can generally only be achieved by extensive data checking and time-consuming data editing. Moreover, if the data are retrospectively recorded on only one occasion or for only one birth cohort, there is a considerable risk of the data base becoming obsolete relatively quickly.

This is why in the case of the socioeconomic panel (Hanefeld, 1987), the advantages of the traditional panel are combined with the retrospective recording of event-history data. Thus each new panel wave provides not only up-to-date information for the time of survey, but by retrospective questions one also records the most important changes and their precise point of time between these waves (for comparison of panel and retrospective studies see Featherman, 1979–80).

Regardless which of the described procedures to record event-history data is selected, it is always an *extensive and costly exercise*. But there exists a potentially strong demand for dynamic analysis of processes and courses in the fields of economics and the social sciences. This growing demand should lead to an increased supply of event-history oriented data structure in these fields in the future.

APPENDIX

List of variable names used in the example

Variable

name	Definition
NOJ	Number of previously held jobs
LFX	Labor force experience in months
EDU	Education, measured in years:
9 years	Lower secondary school qualification (completion of compulsory education)

10 years	Middle School qualification (certificate from <i>Realschule</i>)
11 years	Lower secondary school qualification with additional vocational training (apprenticeship or certificate from specialized vocational school)
12 years	Middle school qualification with additional vocational training degree (apprenticeship or certificate from specialized vocational school)
13 years	<i>Abitur</i> (included in this category are certificates from a <i>Gymnasium</i> , <i>Kolleg</i> or <i>Wirtschaftsgymnasium</i> ; also certificates from a secondary technical school, the <i>Fachoberschule</i> or the <i>Höhere Berufsfachschule</i>)
17 years	Professional college qualification (certificate from a higher technical college or a professional college, the <i>Fachhochschule</i> , <i>Ingenieurschule</i> or <i>Höhere Fachschule</i>)
19 years	University degree (from all institutions of higher education)
COHO2	Dummy variable for cohort 1939–41: 1 cohort 1939–41 0 other
COHO3	Dummy variable for cohort 1949–51: 1 cohort 1949–51 0 other
PRES	Wegener's (1985) prestige score for job <i>i</i>
MAR	Marital status: 0 not married 1 married
TDEP	Time dependence in parametric models
CONST	Intercept

REFERENCES

- Aalen, O. O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7, 1121–1137.
- Andersen, P. K. & Borgan, Ø. (1985). Counting process models for life history data: A review (with discussion). *Scandinavian Journal of Statistics*, 12, 97–158.
- Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10, 1100–1120.
- Blau, P. M. & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.
- Blossfeld, H.-P. (1986). Career opportunities in the Federal Republic of Germany: A dynamic approach to the study of life course, cohort, and period effects. *European Sociological Review*, 2, 208–225.
- Blossfeld, H.-P. (1987). Zur Repräsentativität der SFB-3-Lebensverlaufsstudie: Ein Vergleich mit Daten aus der amtlichen Statistik. *Allgemeines Statistisches Archiv*, 71, 126–144.

- Blossfeld, H.-P., Hamerle, A. & Mayer, K. U. (1989). *Event history analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Carlsson, G. (1958). *Social mobility and class structure*. Lund: C. W. K. Gleerup.
- Carroll, G. & Meyer, K. U. (1986). Job shift patterns in the Federal Republic of Germany: The effects of class, industrial sector, and organizational size. *American Sociological Review*, 51, 323–347.
- Clayton, D. (1988). The analysis of event-history data: A review of progress and outstanding problems. *Statistics in Medicine*, 7, 819–841.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, 34, 187–220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62, 269–276.
- Cox, D. R. & Hinkley, D. V. (1974). *Theoretical statistics*. London: Chapman & Hall.
- Cox, D. R. & Oakes, D. (1974). *Analysis of survival data*. London: Chapman & Hall.
- Featherman, D. I. & Hauser, R. M. (1978). *Opportunity and change*. New York: Academic Press.
- Glass, D. V. (ed.) (1954). *Social mobility in Britain*. London: Routledge & Kegan Paul.
- Goldthorpe, J. H. (1980). *Social mobility and class structure in modern Britain*. London: Routledge & Kegan Paul.
- Hanefeld, U. (1987). *Das Sozio-ökonomische Panel. Grundlagen und Konzeption*. Frankfurt am Main and New York: Campus.
- Heckman, J. J. & Singer, B. (1984). Econometric duration analysis. *Journal of Econometrics*, 24, 63–132.
- Kalbfleisch, J. D. & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.
- Lawless, J. F. (1982). *Statistical models and methods for life-time data*. New York: Wiley.
- Manton, K. G. & Stallard, E. (1988). *Chronic disease modelling: Measurement and evaluation of the risks of chronic disease processes*. London: Griffin.
- Mayer, K. U. et al. (1987). Lebensverläufe und Wohlfahrtsentwicklung. Materialien zur Konzeption, Design und Methodik der Hauptuntersuchung 1981/82. Max Planck Institute for Human Development and Education, Berlin.
- Mayer, K. U. & Brückner, E. (1989). *Lebensverläufe und Wohlfahrtsentwicklung* (Materialien aus der Bildungsforschung Nr. 35). Berlin: Max-Planck-Institut für Bildungsforschung.
- Papastefanou, G. (1980). Zur Güte von retrospektiven Daten – Eine Anwendung gedächtnispsychologischer Theorie und Ergebnisse einer Nachbefragung. Working paper No. 29 of the Sfb 3 'Mikroanalytische Grundlagen der Gesellschaftspolitik'. Frankfurt am Main and Mannheim.
- Roger, J. H. & Peacock, S. D. (1983). Fitting the scale as a GLIM parameter for Weibull, extreme value, logistic and log-logistic regression models with censored data. *GLIM-Newsletter*, 6, 30–37.
- Rogoff, N. (1953). *Recent trends in occupational mobility*. New York: Free Press.

- Sørensen, A. B. (1977). The structure of inequality and the process of attainment. *American Sociological Review*, 42, 965–978.
- Sørensen, A. B. (1984). Interpreting time dependency in career processes. In A. Diekmann & P. Mitter (eds.), *Progress in stochastic modeling of social processes*. New York: Academic Press.
- Sørensen, A. B. (1986). Theory and methodology in stratification research. In U. Himmelstrand (ed.), *The sociology of structure and action*. New York: Sage Publications.
- Sørensen, A. B. & Tuma, N. B. (1981). Labor market structures and job mobility. *Research in Social Stratification and Mobility*, 1, 67–94.
- Spilerman, S. (1977). Careers, labor market structure and socioeconomic achievement. *American Journal of Sociology*, 83, 551–593.
- Svalastoga, K. (1959). *Prestige, class and mobility*. Copenhagen: Scandinavian University Books.
- Töike, A. (1980). Zuverlässigkeit retrospektiver Verlaufsdaten – Qualitative Ergebnisse einer Nachbefragung. Working paper No. 30 of the Sfb 3 'Mikroanalytische Grundlagen der Gesellschaftspolitik'. Frankfurt am Main and Mannheim.
- Tuma, N. B. (1985). Effects of labor market structure on job-shift patterns. In J. Heckman & B. Singer (eds.), *Longitudinal analysis of labor market data*. Cambridge, MA: Cambridge University Press.
- Tuma, N. B. & Hannan, M. T. (1984). *Social dynamics: models and methods*. New York: Academic Press.
- Vaupel, J. W., Mauton, U. G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439–454.
- Wegner, B. (1985). Gibt es Sozialprestige? *Zeitschrift für Soziologie*, 14, 209–235.