SEMI-MARKOV AND COMPETING RISKS MODELS WITH APPLICATIONS TO OCCUPATIONAL MOBILITY

Gilg U.H. Seeber

Institut für Statistik
Universität Innsbruck
Innsbruck, Austria

I. INTRODUCTION AND SUMMARY

This paper attempts to give an introduction to two approaches of stochastic modelling of social processes: Semi Markov processes and competing risks models.

The first one has been widely used in mathematical sociology, at least implicitly. Our presentation is, however, more general, in that we do not postulate any assumptions about the process besides the semi Markov property. Section II.A briefly reports some basic definitions and features concerning semi Markov processes. Nonparametric maximum likelihood estimators for the initial distribution and the transition matrix are presented in section II.B.

Models of competing risks have been developed by biostatisticians and can be interpreted as mechanisms governing special semi Markov processes. Most of the competing risks literature, however, uses the concept of latent sojour times, each associated with

STOCHASTIC MODELLING
OF SOCIAL PROCESSES

ISBN 0-12-215490-8

a specific cause of termination of the actual, observable sojourn time, which is supposed to be the minimum of the latent times. Serious problems of identifiability arise. Our discussion in chapter III confines to a concise presentation of basic properties and problems of competing risks models. Cause specific hazard rates are proposed as constructs, which allow statistical inference without possibly restrictive assumptions.

Chapter IV collects references to the literature. We do not attempt to give a complete bibliography, but it is hoped, the reader will find the books or articles for his particular needs.

Finally, we apply some of the models to a large data set of job histories. Using the semi Markov model a transition matrix is estimated and found to be a useful summary statistic at least, reflecting some of the main paths of occupational mobility. In a second practical example a competing risks model is used to assess the effect of an employee's education to his chance of advancement to a higher position.

The methods proposed in this paper are no new things, but, as far as we know, have not been applied by sociologists. Their usefulness outside biostatistics and reliability theory has to be proven and we hope, that one or the other reader is encouraged to proceed in this way.

II. SEMI MARKOV PROCESSES

A. Basic Theory

Suppose we have a non empty but finite set X of possible states and a sequence $(X_n)_{n\in N}$ of random elements taking values in X. Furthermore, let $(T_n)_{n\in N}$ be a sequence of nonnegative random variables,[1] called sojourn times. For all $n\in N$ we define

$$\tau_n = \sum_{k=1}^{n} T_k \tag{1}$$

and

$$\tau(t) = 1 + \max\{n;\ n\in N \text{ and } \tau_n \le t\}. \tag{2}$$

The stochastic process $(X(t))_{t\geq 0}$ with

$$X(t) = X_{\tau(t)} \tag{3}$$

is called a semi Markov process if:

$$P\{X_{n+1} = y,\ T_n \le t \mid X_n, X_{n-1}, \dots, X_o, T_{n-1}, \dots, T_o\} = P\{X_{n+1} = y,\ T_n \le t \mid X_n\} \tag{4}$$

and

these probabilities are independent of n.

For illustration consider an individual's job history. A typology of working positions within a firm or even a national economy may yield a state space X. The sequence of positions held by one

[1] $(X_n)_{n\in N}$ and $(T_n)_{n\in N}$ originate from the same probability space, (Ω,Σ,P) say.

person is commonly regarded as the realization of a (stationary) Markov process, i.e. the probability of changing to a certain position only depends on the current state and the destination. A generalization of this concept involves the times spent in each position by means of conditional probabilities

$$F_{xy}(t) = P\{T_n \leq t \mid X_{n+1} = y, X_n = x\} \tag{5}$$

of the sojourn times T_n. Note that $F_{xy}(t)$ does not depend on n!

Figure 1 shows a typical trajectory of such a process.

Now let $Q(t) = (Q_{xy}(t))_{x,y\varepsilon X}$, where

$$Q_{xy}(t) = P\{X_{n+1} = y, T_n \leq t \mid X_n = x\} \tag{6}$$

be the matrix of transition probabilities (the semi Markov matrix). If we define

$$p_{xy} = \lim_{t\to\infty} Q_{xy}(t) \ , \tag{7}$$

then we have

$$F_{xy}(t) = \frac{Q_{xy}(t)}{p_{xy}} \ , \text{ if } p_{xy} \neq o. \tag{8}$$

It is sometimes convenient to define

$$F_{xy}(t) = \begin{cases} 0 & \text{if } 0 \leq t < 1 \\ 1 & \text{if } t \geq 1 \end{cases} \ , \text{ if } p_{xy} = 0. \tag{9}$$

It follows, that F_{xy} is a proper distribution in any case. If, for a state x

$$p_{xy} = 0 \qquad \text{for all } y\varepsilon X \ , \tag{10}$$

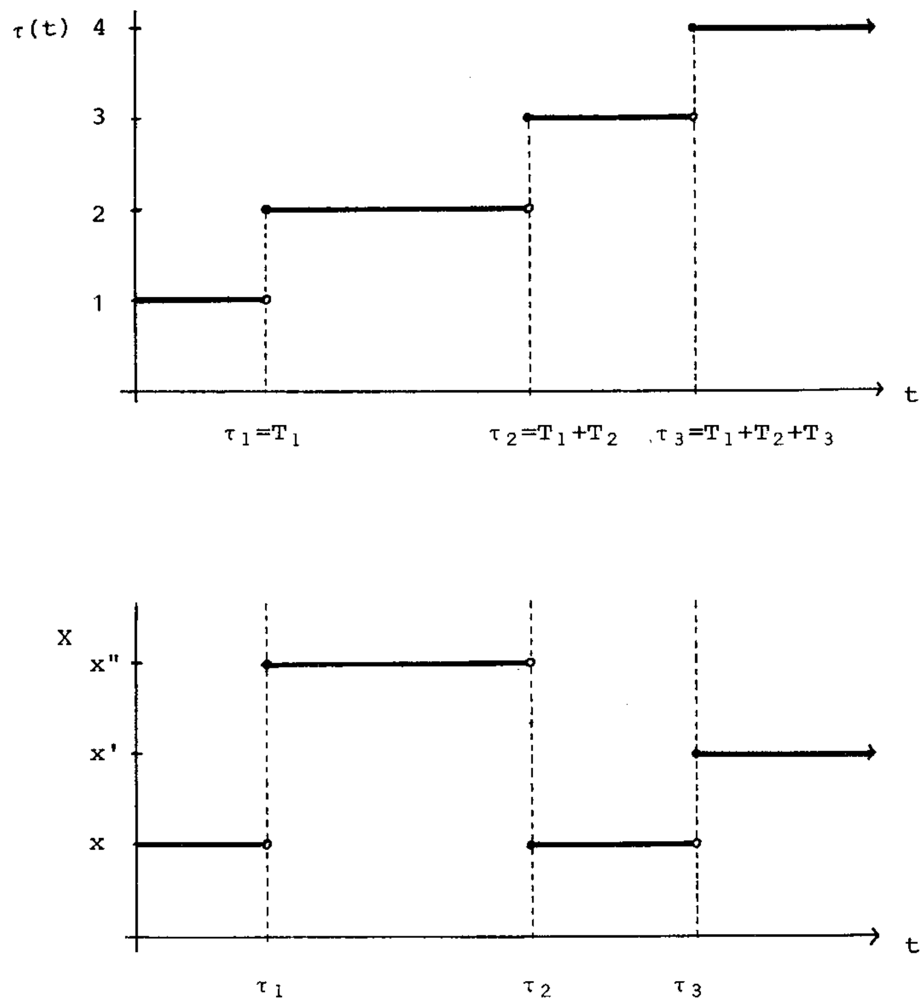

FIGURE 1. Trajectory of a semi Markov process.

then x is called absorbing. In this case, the probability of leaving x before or at time t

$$P_x(t) = P\{T_n \leq t \mid X_n = x\}$$

$$= \sum_{y\varepsilon X} Q_{xy}(t)$$

$$= \sum_{y\varepsilon X} p_{xy}F_{xy}(t) \qquad (11)$$

is identical O (for all t).

A semi Markov process is completely characterized either by the matrix of transition probabilities Q(t) and the vector $\Pi = (\Pi_x)_{x\varepsilon X}$ of its initial distribution or by the conditional distributions $F_{xy}(t)$ of the sojourn times, the probabilities p_{xy} of transitions from x to y, and Π.

The rest of this section introduces some generalizations of concepts used in the statistical analysis of single sojourn times. For mathematical convenience only, we assume all distributions $Q_{xy}(t)$ to be absolutely continuous, i.e. having a density with respect to Lebesgue measure.

First consider the hazard rates

$$\lambda_{xy}(t) = \lim_{\Delta t \downarrow 0} \frac{P\{t \leq T_n < t+\Delta t,\ X_{n+1} = y \mid T_n \geq t,\ X_n = x\}}{\Delta t} \qquad (12)$$

which may be interpreted as the instantaneous risk of leaving state x to state y at time t, provided that no transition has occured up to t. Clearly we have

$$\lambda_x(t) = \lim_{\Delta t \downarrow 0} \frac{P\{t \leq T_n < t+\Delta t \mid T_n \geq t, X_n = x\}}{\Delta t} = \sum_{y\varepsilon X} \lambda_{xy}(t) \qquad (13)$$

i.e. the "destination specific" hazard rates $\lambda_{xy}(t)$ add up to the "overall" hazard rate for leaving state x.

Let

$$S_x(t) = P\{T_n > t \mid X_n = x\}$$

$$= 1 - P_x(t) \qquad (14)$$

be the survival function for state x and $Q'_{xy}(t)$ the density of $Q_{xy}(t)$, i.e.

$$Q_{xy}(t) = \int_0^t Q'_{xy}(u)\ du\ , \qquad (15)$$

then the following identities hold

$$\lambda_{xy}(t) = \frac{Q'_{xy}(t)}{S_x(t)} \qquad (16)$$

$$S_x(t) = \exp\ (-\int_0^t \lambda_x(u)\ du) \qquad (17)$$

$$Q'_{xy}(t) = \lambda_{xy}(t)\ \cdot \prod_{z\varepsilon X} \exp\ (-\int_0^t \lambda_{xz}(u)\ du). \qquad (18)$$

An observation $(x_1,t_1,\ldots x_{s-1},t_{s-1},x_s)$ where x_s is absorbing is called complete. It's probability is

$$P\{X_1=x_1,\ T_1=t_1,\ldots,X_{s-1}=x_{s-1},T_{s-1}=t_{s-1},X_s=x_s\} = \qquad (19)$$

$$= \Pi_{x_1} \cdot Q'_{x_1x_2}(t_1)\ldots Q'_{x_{s-1}x_s}(t_{s-1})\ .$$

If an absorbing state has not been reached, we only may record a censored observation $(x_1,t_1,\ldots,x_s,t_s)$, for which

$$P\{X_1=x_1, T_1=t_1, \ldots, X_s=x_s, T_s>t_s\} =$$

$$\Pi_{x_1} \cdot Q'_{x_1x_2}(t_1) \ldots Q'_{x_{s-1}x_s}(t_{s-1}) \cdot S_{x_s}(t_s) . \qquad (20)$$

Hence, assuming an independent censoring mechanism, the likelihood function for independent observations is (proportional to) the product of terms of the form (19) and (20). Note that, disregarding initial probabilities, the likelihood function can be written entirely in terms of the hazard rates $\lambda_{xy}(t)$, c.f. (18).

Finally we remark, that under quite general conditions there exist mutually independent, nonnegative random variables $Y_{xy}(x,y\varepsilon X)$ with distribution functions

$$G_{xy}(t) = 1 - \exp(-\int_0^t \lambda_{xy}(u)\, du) \ , \qquad (21)$$

such that

$$P_x(t) = P\{\min\{Y_{xy};\ y\varepsilon X\} \leq t\} \qquad (22)$$

and

$$Q_{xy}(t) = \int_0^t G'_{xy}(u) \cdot \prod_{\substack{z\varepsilon X \\ z\neq y}} (1-G_{xz}(u))\, du, \qquad (23)$$

where $G'_{xy}(t)$ is the density of $G_{xy}(t)$. (22) may be interpreted as a mechanism governing the semi Markov process. A slightly generalized situation is treated in chapter III.

B. Nonparametric Estimation.

In this section we describe, without going into too great detail, non parametric maximum likelihood estimators for the initial distribution and the transition matrix $Q_{xy}(t)$ of a semi Markov process. We do not require any assumptions about the underlying distributions of the sojourn times, in particular they need not be continuous. However, in rearranging the likelihood function to a form suitable for maximization, our presentation assumes continuity.

Without loss of generality, let $X = \{1,\ldots,s\}$. From (19-20) it follows, that the likelihood of N independent observations factors into three products

$$L = \prod_{x=1}^{s} a_x \cdot \prod_{x,y=1}^{s} b_{xy} \cdot \prod_{x=1}^{s} c_x, \qquad (24)$$

where each a_x is of the form $(\Pi_x)^{l_x}$, l_x being the number of initial entries in state x. The first product may be maximized separately of the remaining parts and leads to

$$\hat{\Pi}_x = \frac{l_x}{N} \ , \qquad x=1,\ldots,s \qquad (25)$$

as ML - estimator for the initial distribution Π.
Each b_{xy} is a product of terms $Q'_{xy}(t)$, and c_x is a product of terms $S_x(t)$, where the t's are observed or censored sojourn times, respectively. The terms c_x make an explicit solution of the likelihood difficult to determine, so we use the distribution functions $G_{xy}(t)$ from (21) for reparametrization. Defining

$$H_{xy}(t) = 1 - G_{xy}(t) \qquad (26)$$

to be the survivorship function corresponding to $G_{xy}(t)$ it follows that

$$Q'_{xy}(t) = G'_{xy}(t) \cdot \prod_{\substack{z=1\\z\neq y}}^{s} (1 - G_{xz}(t))$$

$$= - H'_{xy}(t) \cdot \prod_{\substack{z=1\\z\neq y}}^{s} H_{xz}(t) \tag{27}$$

and

$$S_x(t) = P\{T_n > t \mid X_n = x\}$$

$$= \prod_{y=1}^{s} P\{Y_{xy} > t\}$$

$$= \prod_{y=1}^{s} (1-G_{xy}(t))$$

$$= \prod_{y=1}^{s} H_{xy}(t) \tag{28}$$

This allows to rearrange the likelihood function to be

$$L = \prod_{x=1}^{s} a_x \cdot \prod_{x,y=1}^{s} L_{xy} \quad , \tag{29}$$

where

$$L_{xy} = \prod_{i=1}^{k} [- H'_{xy}(t_i)^{m_{xyi}} \cdot \prod_{\substack{z=0\\z\neq y}}^{s} H_{xz}(t_i)^{m_{xzi}}] \tag{30}$$

$t_1 < \ldots < t_k$ are the distinct sojourn times in x

m_{xyi} is the number of transitions from x into y with sojourn time t_i, and

m_{xOi} is the number of censored observations in state x with time t_i .

To find the ML-estimator for $H_{xy}(t)$ we only have to consider (30). Its solution must be a discrete distribution with mass points only at $t_1,\ldots,t_k$. Now, let

$$N_{xyi} = m_{xOi} + \sum_{z=y}^{s} m_{zxi} + \sum_{z=o}^{s} \sum_{j=i+1}^{k} m_{xzj} \tag{31}$$

be the number of individuals in state x which have not moven into a new state before time t_i, i.e. are exposed to the risk of leaving state x. (The sum in the middle of formula (31) allows distributions underlying the sojourn times, which are not continuous as well.)

Define

$$u_{xyi} = \begin{cases} 1 - \dfrac{m_{xyi}}{N_{xyi}} & \text{if } N_{xyi} \neq O \\ 1 & \text{otherwise} \end{cases} \tag{32}$$

(the estimated conditional probability of not leaving state x into state y at time t_i), then the ML-estimator for $H_{xy}(t_i)$ is obtained by

$$\hat{H}_{xy}(t_i) = \prod_{j=1}^{i} u_{xyj} \quad . \tag{33}$$

Using

$$\hat{H}_{xy}(t_i) = \hat{H}_{xy}(t_{i-1}) - \hat{H}_{xy}(t_i) \tag{34}$$

and

$$v_{xyi} = (1 - u_{xyi}) \cdot \prod_{z=1}^{y-1} u_{zxi} \cdot \prod_{z=1}^{s} \prod_{j=1}^{i-1} u_{xzj} \tag{35}$$

yields to

$$\hat{p}_{xy} = \sum_{j=1}^{k} v_{xyj} \tag{36}$$

and

$$\hat{Q}_{xy}(t) = \hat{p}_{xy} - \sum_{j=i+1}^{k} v_{xyj} \quad \text{for all } t \text{ such that} \tag{37}$$

$t_i \le t < t_{i+1}$ as non parametric ML-estimators for p_{xy} and $Q_{xy}(t)$.

If the largest sojourn time in state x is censored, then

$$\sum_{z=1}^{s} \hat{p}_{xz} < 1 \ . \tag{38}$$

In this case, $\hat{p}_{xy}$ may be normalized by setting

$$\hat{p}_{xy} = \frac{\hat{p}_{xy}}{\sum_{z=1}^{s} \hat{p}_{xz}} \tag{39}$$

If we have as a special case only two states, i.e. $X = \{1,2\}$, and

$$\Pi_1 = 1,$$
$$p_{12} = 1, \text{ and}$$
$$\text{state 2 is absorbing,} \tag{40}$$

then $\hat{Q}_{12}(t)$ reduces to the product-limit estimator.

Approximate variances/covariances of the above estimators are available, see the bibliographical remarks.

III. COMPETING RISKS

A. Potential Sojourn Times

Consider the case with state space $X = \{0,1,\dots,m\}$, where 0 is transient and $1,\dots,m$ are absorbing states. Every individual is supposed to enter the process in state 0. From (21-22) we know, that there exist under certain regularity conditions mutually independent non-negative random variables $Y_1,\dots,Y_m$ such that

$$P(t) = P\{\min\{Y_1,\dots,Y_m\} \le t\}. \tag{41}$$

(Note, that we have dropped the subscript 0!) $Y_1,\dots,Y_m$ may be interpreted as potential or latent sojourn times corresponding to specific causes terminating state 0, namely the transitions to one of the states $1,\dots,m$.

In this chapter we are concerned with a slightly more general situation: Given m not necessarily independent non-negative random variables $Y_1,\dots,Y_m$ we can observe their identified minimum, i.e.

$$T = \min\{Y_1,\dots,Y_m\} \tag{42}$$

and

$$J = \begin{cases} j & \text{if } T = Y_j \\ 0 & \text{otherwise,} \end{cases} \tag{43}$$

where we assume that

$$P\{Y_i = Y_j\} = 0 \qquad \text{if } i \ne j \ . \tag{44}$$

In survival analysis terminology this situation is called "competing risks": With each variable Y_j there is associated a specific risk or cause of death. The reader should be warned, however, that the (marginal) distribution of each Y_j depends in general on the risks acting simultaneously; e.g. if one cause is removed, the distribution of Y_j need not remain the same as before.

B. Identifiability

One of the basic problems in the theory of competing risks is the following: Does the joint distribution of the identified minimum (T,J) uniquely determine the joint distribution of the Y_i's?

If $Y_1,\ldots,Y_m$ are mutually independent the answer is "yes": for the marignal distribution function $G_i(t)$ of Y_i we have

$$G_i(t) = 1 - \exp(-\int_o^t \frac{Q_i'(u)}{S(u)}\, du)$$

$$= 1 - \exp(-\int_o^t \lambda_i(u)\, du), \tag{45}$$

where $Q_i(t)$ is the joint distribution function of (T,J), sometimes called incidence function, and

$$S(t) = 1 - \sum_{j=1}^{m} Q_j(t) \quad , \tag{46}$$

as we had before. (As a remark we note, that, in this case, the cause specific hazard function λ_i is identical to the (marginal) hazard function of the sojourn time Y_i.) The joint distribution function of $Y_1,\ldots,Y_m$ is simply the product of the marginal distribution functions.

Now, if the Y_i's are dependent, then there exist mutually independent random variables having an identified minimum with distribution function identical to $Q_i(t)$; a result which is not surprising in view of formulae (21-23). It follows, that in general

the dependent and the independent model are indistinguishable by means of the observable quantities (T,J).

In order to avoid any complications a large portion of the competing risks literature assumes independence of the potential sojourn times. Other possibilities to circumvent the problem of identifiability include restriction to certain parametric families or postulation of proportional hazards, i.e. to assume that there exist constants c_i between 0 and 1 such that

$$\lambda_i(t) = c_i \cdot \lambda(t) = c_i \cdot (\sum_{j=1}^{m} \lambda_j(t)) \; . \tag{47}$$

In this case we have

$$F_i(t) = P\{Y_i \le t | J = i\}$$

$$= P\{T \le t\} = P(t) \tag{48}$$

and

$$1 - G_i(t) = 1 - \exp(-c_i \cdot \int_o^t \lambda(u)\, du)$$

$$= (1 - P(t))^{c_i}$$

$$= S(t)^{c_i} \; . \tag{49}$$

It can be shown, that c_i is equal to p_i, the expected proportion of transition to state i / deaths from cause i.

The assumptions one is likely to underly his investigations depends on a careful inspection of the real world phenomena to be analyzed.

C. Estimation

As we have seen in the discussion on semi Markov processes, the incidence function $Q_i(t)$, hence the destination or cause specific hazard rates $\lambda_i(t)$ and the survival function $S(t)$ are estimable whether or not the potential sojourn times are independent. For instance, in this situation the nonparametric model discussed in section II. B is applicable as well.

Recall that the likelihood function can be written entirely in terms of the hazard rates $\lambda_i(t)$. Suppose we have independent observations (t_n,δ_n,j_n), $n=1,\dots,N$, where t_n is the observed time, δ_n a censoring indicator with

$$\delta_n = \begin{cases} 1 & \text{if } t_n \text{ is a complete observation} \\ 0 & \text{if } t_n \text{ is a censored observation} \end{cases} \tag{50}$$

and j_n is the destination or cause (if $\delta_n = 1$), then

$$\begin{aligned} L &= \prod_{n=1}^{N} \lambda_{j_n}(t_n)^{\delta_n} \cdot S(t_n) \\ &= \prod_{n=1}^{N} \left[\lambda_{j_n}(t_n)^{\delta_n} \cdot \prod_{i=1}^{m} \exp(-\textstyle\int_0^t \lambda_i(u)\,du)\right] . \end{aligned} \tag{51}$$

Using

$$\zeta_{ni} = \begin{cases} 1 & \text{if } j_n = i \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

we rearrange (51) to

$$\begin{aligned} L &= \prod_{n=1}^{N} \left[\prod_{i=1}^{m} \lambda_i(t_n)^{\delta_n\zeta_{ni}} \cdot \prod_{i=1}^{m} \exp(-\textstyle\int_0^t \lambda_i(u)\,du)\right] \\ &= \prod_{i=1}^{m} \left[\prod_{n=1}^{N} \lambda_i(t_n)^{\delta_n\zeta_{ni}} \cdot \exp(-\textstyle\int_0^t \lambda_i(u)\,du)\right] . \end{aligned} \tag{53}$$

Hence the likelihood factors into separate components for each cause and, furthermore, the component for cause i may be obtained by regarding all sojourn times terminated by other causes than i as censored. As one important result we find, that statistical inference on the hazard function $\lambda_i(t)$ can be carried out by means of the methods used in the analysis of univariate sojourn times.

Covariates are easily incorporated in the model by considering all probabilities to be conditional on a vector z of concomitant variables, e.g.

$$\lambda_i(t;z) = \lim_{\Delta t\downarrow 0} \frac{P\{t\le T<t+\Delta t, J=i \mid T\ge t, z\}}{\Delta t} \tag{54}$$

Again, a variety of standard models and methods are available to assess the relationship between the cause specific hazard rates and covariates, including proportional hazards and accelerated failure time models.

In our analysis of the labour market problem we have used a fully parametric proportional hazards model with rates stemming from a Weibull distribution, i.e. we have assumed

$$\lambda_i(t;z_1,\dots,z_k) = \alpha_i \cdot t^{\alpha_i - 1} \cdot \exp(\beta_{oi} + \sum_{l=1}^{k} \beta_{li}\, z_l) . \tag{55}$$

Using GLIM we have obtained maximum likelihood estimates for the parameters. Standard asymptotic theory may be applied for significance testing.

IV. BIBLIOGRAPHICAL REMARKS

In our presentation of semi Markov processes we have followed Nollau (1981). The nonparametric model described in section II.B was developed by Lagakos, Sommer and Zelen (1978), where approximate variances/covariances of the estimates may be found. Gill (1980) proves uniform consistency and weak convergence under mild censoring conditions. A generalization allowing incomplete observations of destinations as well is described in Dinse (1982).

Chapters on competing risks models are contained in the books of Kalbfleisch and Prentice (1980) and Elandt-Johnson and Johnson (1980). The first reference concentrates on cause specific hazard rates, the second book treats identifiability questions and proportional hazard models in more detail. Review articles include David and Moeschberger (1976), Gail (1975), Elandt-Johnson (1976), and Prentice et al. (1978). Parametric competing risks models are extensively studied in the monograph of David and Moeschberger (1978). An early reference is Cox (1959).

The identifiability problem is treated in Elandt-Johnson (1981), Miller (1977), Nádas (1971), and Tsiatis (1975). Langberg, Proschan and Quinzi (1978) describe the transformation of a dependent model into an independent one, preserving essential features. Basu and Ghosh (1978) examine various multivariate distributions for identifiability knowing the distribution of the identified minimum, including bivariate and trivariate normal distributions, the bivariate exponential distribution of Marshall and Olkin (1967) etc.

Estimation in independent models is considered by Boardman and Kendell (1970) and Herman and Patell (1971) among others.

Elandt-Johnson (1976), Langberg, Proschan and Quinzi (1981) and Moeschberger (1974) treat dependent sojourn times. The subject in a semi Markov process context is mentioned by Aalen (1976).

For the study of models incorporating concomitant information we refer to Beck (1979), Holt (1978), Lagakos (1978), and Prentice, Williams and Peterson (1981). Elandt-Johnson (1978) and Nádas (1970) consider proportional hazard models.

Two further topics we have not mentioned may be of interest: Nelson (1970, 1972) developes hazard plotting techniques for multivariate sojourn times and Johnson and Koch (1978) apply linear model techniques to grouped, i.e. interval censored times.

Besides the references cited, there is, naturally, a lot of literature concerning the statistical analysis of single sojourn times, which may be helpful in working with competing risks problems. A book on that subject written for (mathematical) sociologists is Coleman (1981). We have followed Aitkin and Clayton (1981) in fitting the Weibull models by use of GLIM.

V. APPLICATIONS TO OCCUPATIONAL MOBILITY

A. Stochastic Models for Occupational Careers.

This part of the paper applies some of the statistical models discussed earlier to problems of intragenerational occupational mobility. Our concern is to assess certain features of typical careers or job histories using a stochastic process approach. As a job history we consider the sequence of occupational positions a person held during his life. One way to find out typical careers is to assume an individual job history to be the realization of a stationary Markov process. Inspection of the estimated transition matrix could give insight about the main paths of occupational mobility.

Our approach is somewhat more general in that we use a semi Markov representation for job histories for two reasons: First, the times spent in certain positions may follow specific patterns. Within the primary sector of labour market there may exist typical service times preceding an advancement to a position of higher qualification, whereas in the secondary sector changes in job status may occur rather arbitrarily. The second reason is of technical significance: available data on job histories typically include incomplete observations, up to date records of complete careers are not at all easy to obtain. Semi Markov models can be adapted to allow censoring. If one is mainly interested in transition probabilities the nonparametric model of section II. B is recommended. Moreover, if little is known about the distributional shape of the sojourn times, the nonparametric estimates provided by that model give first insights. This model is used for the analysis of the data described in section V. B below.

The second practical example illustrates the use of a competing risks model to assess the impact of an employee's education to his chance of promotion to a higher position.

For an extensive treatment on stochastic modelling of social processes we refer to the book by Bartholomew (1973). Stewman (1976) reviews Markov models for careers. The work of Michael J. Piore on mobility chains and labour market segmentation we feel gives a thorough sociological background for our analyses, c.f. Piore (1978) for instance.

B. The Data.

Our data originate from an interrogation conducted by "Österreichisches Statistisches Zentralamt" in September 1972, where about eighty thousand Austrian inhabitants were asked about some features of their occupational careers. Each period of an individual job history should be described by three characteristics with respect to an economic typology of enterprises, a typology of occupations and the qualification of the working position. Each change in one of the characteristics was taken as the beginning of a new period for which the year had to be recorded.

For our purposes we have extracted all those males from the sample, who entered their occupational career between 1955 and 1970. This lead to a (sub-) sample of 9345 people.

Our main interest lies in the qualification of the working position held by each person. The corresponding variable, briefly called "status" subsequently, takes the following values and meanings:

Table 1: Outcomes of the variable "status"
(For details we refer to ÖStZ (1974))

1	*working as an apprentice*
2	*working as an unskilled worker*
3	*working as a skilled worker*
4	*working as an employee with easy activities*
5	*working as an employee with medium activities*
6	*working as an employee with qualified activities*
7	*working as an employee with managing activities*
8	*working as an entrepreneur*
9	*assisting (without working contract)*
10	*unknown*

C. Analysis of Changes in Status

We now apply the nonparametric semi Markov model described in section II. B to analyze the job histories with respect to the variable "status". Note, that the beginning of a new period in one's occupational career is constituted by a change in one of the three characterizing variables, so it may happen that there is no change in the "status". It follows, that the probability of moving to a state identical to the current state need not be zero.

Our analysis is based on 9345 job histories with 27814 sojourn times, of which 9345 are right censored. (We do not have any complete job histories). We first compute the maximum likelihood estimates for the initial distribution, which are just the proportions entering each state.

Table 2: Distribution Π on initial states with respect to variable "status".

State		Frequency	Proportion
1	apprentice	5417	0.580
2	unskilled worker	1407	0.151
3	skilled worker	127	0.014
4	employee/easy activities	358	0.038
5	employee/medium activities	489	0.052
6	employee/qualified activities	378	0.040
7	employee/managing activities	86	0.010
8	entrepreneur	101	0.011
9	assisting	879	0.094
10	unknown	103	0.011

Next we record the frequencies of transitions for two purposes. First, to get an impression about how often each kind of transition occurs. Second to compare an estimation of the transition matrix by singly calculating the proportions of observed transitions in each row (i.e. disregarding censored observations) to the estimates obtained by the semi Markov model:

Table 3: Absolute frequencies of transitions.

	State									
State	1	2	3	4	5	6	7	8	9	10
1	100	451	3604	390	100	14	3	12	43	372
2	100	2306	337	257	71	16	6	203	49	580
3	18	653	714	316	244	116	25	218	32	1118
4	3	74	34	295	301	93	30	47	6	202
5	1	13	10	30	255	245	76	53	1	70
6	1	4	8	4	24	164	148	76	2	31
7	1	2	1	0	7	6	61	34	0	5
8	0	42	14	7	11	10	12	29	0	3
9	43	268	19	21	8	2	0	248	7	180
10	21	759	1006	298	170	76	22	42	141	124

One further descriptive statistics should be computed for diagnostic reasons: for each state the pattern of censored sojourn times. Striking irregularities call for careful inspection of the data and the underlying model. Since we did not find any conspiciousness we drop the reproduction here.

Now we may calculate the matrix of transition probabilities:

Table 4: Matrix (p_{xy}) of estimated transition probabilities; estimated standard errors are quoted within brackets.

State	1	2	3	4	5	6	7	8	9	10
1	0.020	0.104	0.656	0.062	0.023	0.003	0.001	0.008	0.007	0.118
	(0.003)	(0.009)	(0.015)	(0.005)	(0.005)	(0.001)	(0.000)	(0.004)	(0.002)	(0.010)
2	0.018	0.518	0.074	0.063	0.016	0.003	0.001	0.101	0.013	0.114
	(0.002)	(0.012)	(0.005)	(0.006)	(0.002)	(0.001)	(0.001)	(0.013)	(0.004)	(0.005)
3	0.004	0.147	0.162	0.082	0.067	0.033	0.008	0.083	0.006	0.195
	(0.001)	(0.006)	(0.007)	(0.006)	(0.005)	(0.003)	(0.002)	(0.008)	(0.001)	(0.005)
4	0.001	0.055	0.021	0.200	0.250	0.071	0.025	0.048	0.003	0.109
	(0.001)	(0.012)	(0.004)	(0.012)	(0.018)	(0.008)	(0.005)	(0.008)	(0.001)	(0.008)
5	0.001	0.009	0.007	0.023	0.209	0.246	0.154	0.051	0.001	0.045
	(0.001)	(0.003)	(0.002)	(0.005)	(0.015)	(0.018)	(0.030)	(0.008)	(0.001)	(0.005)
6	0.001	0.004	0.008	0.005	0.029	0.205	0.306	0.126	0.002	0.030
	(0.001)	(0.002)	(0.003)	(0.002)	(0.006)	(0.020)	(0.044)	(0.016)	(0.001)	(0.005)
7	0.002	0.008	0.003	0.	0.018	0.021	0.193	0.154	0.	0.020
	(0.002)	(0.007)	(0.003)		(0.007)	(0.010)	(0.025)	(0.032)		(0.010)
8	0.	0.109	0.018	0.009	0.017	0.024	0.042	0.045	0.	0.003
		(0.035)	(0.005)	(0.004)	(0.006)	(0.010)	(0.027)	(0.010)		(0.002)
9	0.038	0.293	0.020	0.023	0.010	0.002	0.	0.381	0.006	0.178
	(0.006)	(0.018)	(0.005)	(0.005)	(0.003)	(0.002)		(0.020)	(0.002)	(0.013)
10	0.006	0.246	0.282	0.106	0.071	0.047	0.020	0.063	0.036	0.083
	(0.002)	(0.012)	(0.010)	(0.008)	(0.008)	(0.008)	(0.005)	(0.015)	(0.004)	(0.013)

The attentive reader may have observed, that the entries in each row do not sum up to one, as it is necessarily the case for a stochastic matrix. This occured because in each state the largest observed sojourn times are censored.

For state 8 (working as an entrepreneur) the sum of the row elements equals only 0.266. Inspection of the tables 2 and 3 leads to an explanation: We have 1034 initial entries or transition to state 8 from distinct states but only 128 transition from state 8, which is not at all surprising, since in Austria entrepreneurs rarely loose their position. It follows that most of the sojourn times recorded in state 8 must be censored.

If one prefers to report a stochastic matrix as transition matrix, he may obtain this simply by normalization. (Table 5)

Since we have a very large number of observed sojourn times we may apply asymptotic theory for significance testing of transition probabilities. No one will wonder, that, in view of the enormous sample size, about three quarters of the entries are significantly different from 0 at the 1 % level.

As is intrinsic to Austria's educational system,a majority of the beginners seek for a position as apprentice. Nevertheless, a portion of 10 % start their occupational life as unskilled workers. Having finished apprenticeship, most of the young people will be able to find jobs as skilled workers. A minority of about 10 % has to accept positions as unskilled workers. Possibly, some of these have not been successful as apprentices.

Table 6 contains the estimated conditional survival function for apprentices becoming skilled workers ($1-F_{13}(t)$ in our notation). It is easily verified that apprenticeship usually lasts for three or four years.

Table 5: Matrix of transition probabilities adjusted for censoring. Values less than 0.01 have been dropped for readability. Entries not significantly different from O in Table 4 are quoted within brackets.

State	1	2	3	4	5	6	7	8	9	10
1	0.02	0.10	0.65	0.06	0.02			(0.01)	0.01	0.12
2	0.02	0.56	0.08	0.07	0.02			0.11	0.01	0.12
3		0.19	0.21	0.10	0.09	0.04	0.01	0.11	0.01	0.25
4		0.07	0.03	0.26	0.32	0.09	0.03	0.06		0.14
5		0.01	0.01	0.03	0.28	0.33	0.21	0.07		0.06
6		(0.01)	(0.01)	(0.01)	0.04	0.29	0.43	0.18		0.04
7	(0.01)	(0.02)	(0.01)		(0.04)	(0.05)	0.46	0.37		(0.05)
8		0.41	0.07	(0.04)	(0.06)	(0.09)	(0.16)	0.17		(0.01)
9	0.04	0.31	0.02	0.02	0.01			0.40	0.01	0.19
10	0.01	0.26	0.29	0.11	0.07	0.05	0.02	0.07	0.04	0.09

Table 6: Estimated conditional survival function for apprentices becoming skilled workers.

T	S(T)	SE(S(T))	0	S(T)	1
1	.998	.001	I	. . .	S
2	.993	.001	I	. . .	S
3	.974	.003	I	. . .	SI
4	.524	.012	I	. . S .	I
5	.233	.018	I	S. . .	I
6	.163	.019	I	S . . .	I
7	.124	.019	I	S . . .	I
8	.109	.019	I	S . . .	I
9	.084	.019	I S	. . .	I
10	.073	.018	I S	. . .	I
11	.068	.018	I S	. . .	I
12	.055	.018	I S	. . .	I
13	.044	.017	I S	. . .	I
14	.034	.016	IS	. . .	I
15	.014	.013	IS	. . .	I
16	.014	.013	IS	. . .	I
17	.014	.013	IS	. . .	I

At this point it should be noted, that these results, although predictive in nature, need not be valid in the eighties. Today, it can be difficult to find a place as apprentice, and, furthermore, successfully having finished apprenticeship is not at all a garantee for a position as a skilled worker.

If an unskilled worker changes his position he is likely to labour as an unskilled worker again, although a minority can advance to positions as skilled workers or employees. For several reasons it can be claimed, that this minority mainly consists of former skilled workers. Using the Markov approach an answer can only be given by suitably altering the state space. A second minority will become entrepreneurs.

Skilled workers, if changing, advance to entrepreneurs or employees, remain skilled workers or attain jobs as unskilled workers. The latter does not necessarily signify a worsening of their (financial) situation, but always means, that they move to positions outside their learned profession.

Careers of employees typically follow the pattern of permanent advancement. It is nearly impossible, for an employee, to descend to a minor position and the chance for advancement is fairly large. The probability of becoming an entrepreneur increases with the position held.

What has been said up to this point can be summarized as follows: There are two segments of labour market: one consists of (blue collar) workers, the second of (white collar) employees, and there is a narrow "one way" path from the first to the second sector to be passed primarily by skilled workers.

The majority of entrepreneurs, who loose their positions, reenter labour force as unskilled workers, often a hard lot. But, as we have seen above, the absolute frequency of such changes is small.

Assisting activities (value 9 of the status-variable) occur mainly within the agricultural sector of Austria's economy. There, young people usually help their parents before they take possession of the farm or they leave home to become unskilled workers. This movement out of the agricultural sector caused one of the major changes in the Austrian society during the period after World War II.

Finally we add a remark concerning state 10 (unknown). Times spent for military service are recorded as unknown. Therefore the probability of changing to state 10 is larger for states which are likely to be held by the majority of younger people. It is supposed that a larger portion returns to their former position after service.

D. Effects of Education.

For the purposes of this section we have extracted from our (sub-) sample 332 men, who have entered their occupational career as employees with medium activities between 1950 and 1960. The question to be asked is: does a person's education influence his chance of advancing to a higher position at the first period of his occupational career. We give an answer by using a competing risks model, in particular we assess the effect of education on cause specific hazard functions.

"Education" is taken as binary variable with outcomes

1 for a person with university or AHS-degree (college degree)

0 otherwise. (56)

Table 7: Crosstabulation of education and second position.

	Education: *0 Low*	*1 High*	*Total*
Advancement	*53*	*81*	*134*
No advancement	*47*	*28*	*75*
Censored	*73*	*50*	*123*
Total	*173*	*159*	*332*

Table 7 gives a crosstabulation of the two variables we are interested in. A chi square test rejects the null hypothesis of independence at the 1 % level.

Now, let t_i be the sojourn time in the i-th person's first position and z_i his education. We use the parametric proportional hazards model

$$\lambda_j(t_i;z_i) = \alpha_j \cdot t^{\alpha_j - 1} \cdot \exp(\beta_{0j} + \beta_{1j} \cdot z_i) \quad j=1,2 \qquad (57)$$

where j=1 for advancement and j=2 otherwise. Therefore, the cause specific hazard rates for both causes stem from a Weibull distribution with shape parameters α_1, α_2 respectively. The hazard function is strictly increasing, if the shape parameter is larger than 1, constant, if it is equal to 1 and strictly decreasing otherwise.

The parameters of each cause specific hazard rate are estimated by regarding all sojourn times terminated by the other cause as censored and maximizing the corresponding likelihood functions. We have used GLIM and the method described in Aitkin and Clayton (1980). The results are given in table 8.

Table 8: Estimates for the parameters of the cause specific hazard rates, estimated asymptotic standard errors are quoted within brackets.

Parameter	*Cause: Advancement*		*No advancement*	
$\hat{\alpha}_j$	*1.2936*	*(0.0651)*	*0.99013*	*(0.0525)*
$\hat{\beta}_{0j}$	*-4.197*	*(0.2145)*	*-3.561*	*(0.1946)*
$\hat{\beta}_{1j}$	*0.6165*	*(0.1767)*	*-0.3335*	*(0.2387)*

As results from an analysis of deviance, the shape parameter is significantly different from 1 for the advancement rate, which is, therefore, strictly increasing. The corresponding null hypothesis for the second rate cannot be rejected.

Using the asymptotic normal distribution of the parameter estimates, one can test, that all the β-parameters besides β_{12} are different from 0. It follows in particular, that education has a significant influence on the advancement hazard rate but not on the no advancement hazard function.

REFERENCES

Aalen, O. (1976). "Nonparametric Inference in Connection with Multiple Decrement Models". Scand.J.Statist. 3 (1976), 15-27.

Aitkin, M., and Clayton, D. (1980). "The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data using GLIM". Appl.Statist. 29 (1980), 156-163.

Bartholomew, D.J. (1973). Stochastic Models for Social Processes. John Wiley, New York

Basu, A.P., and Ghosh, J.K. (1978). "Identifiability of the Multinormal and Other Distributions under Competing Risks Model". J. Multivariate Analysis 8 (1978), 413-429.

Beck, G.J. (1979). "Stochastic Survival Models with Competing Risks and Covariates". Biometrics 35 (1979), 427-438.

Boardman, T.J., and Kendell, P.J. (1970): "Estimation in Compound Exponential Failure Models." Technometrics 12 (1970), 891-900.

Coleman, J.S. (1981). Longitudinal Data Analysis. Basic Books, New York

Cox, D.R. (1959). "The Analysis of Exponentially Distributed Life-Times with Two Types of Failure". J.R.Statist.Soc. B 21 (1959), 411-421.

David, H.A., and Moeschberger, M.L. (1976). "Some aspects of the theory of competing risks. Proc. 9th Int. Biometric Conference, 1976, 379-390.

David, H.A., and Moeschberger, M.L. (1978). The Theory of Competing Risks. Griffin, London.

Dinse, G.E. (1982). "Nonparametric Estimation for Partially-Complete Time and Type of Failure Data". Biometrics 38 (1982), 417-431.

Elandt-Johnson, R.C. (1976a). "Some Models in Competing Risks Theory: Multiple Causes of a Single Death." Proc. 9th Int. Biometric Conference, 1976, 391-407.

Elandt-Johnson, R.C. (1976b). "Conditional Failure Time Distribution under Competing Risks Theory with Dependent Failure Times and Proportional Hazard Rates". Scand. Actuar. J. 1976, 37-51.
Elandt-Johnson, R.C. (1978). "Some Properties of Bivariate Gumbel Type A Distributions with Proportional Hazard Rates". J.Multivariate Analysis 8 (1978), 244-254.
Elandt-Johnson, R.C. (1981), "Equivalence and Nonidentifiability in Competing Risks. A Review and Critique". Aligarh J.Statist. 1, (1981), 28-42.
Elandt-Johnson, R.C., and Johnson, N.L. (1980). Survival Models and Data Analysis. John Wiley, New York.
Gail, M. (1975). "A Review and Critique of Some Models Used in Competing Risks Analysis". Biometrics 31 (1975), 209-222.
Herman, R.J., and Patell, R.K.N. (1971). "Maximum Likelihood Estimation for Multi-Risk Model". Technometrics 13 (1971), 385-396.
Holt, J.D. (1978). "Competing Risk Analyses with Special Reference to Matched Pair Experiments". Biometrika 65 (1978), 159-165.
Johnson, W.D., and Koch, G.G. (1978). "Linear Models Analysis of Competing Risks for Grouped Survival Times." Int.Statist.Rev. 46 (1978), 21-51.
Kalbfleisch, J.D., and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. John Wiley, New York.
Lagakos, S.W. (1978). "A Covariate Model for Partially Censored Data Subject to Competing Causes of Failure". Appl. Statist. 27 (1978), 235-241.
Lagakos, S.W., Sommer, C.J., and Zelen, M. (1978). "Semi-Markov Models for Partially Censored Data". Biometrika 65 (1978), 311-317.
Langberg, N., Proschan, F., and Quinzi, A.J. (1978). "Converting Dependent Models into Independent Ones, Preserving Essential Features. Ann. Prob. 6 (1978), 174-181.
Langberg, N., Proschan, F., and Quinzi, A.J. (1981). "Estimating Dependent Life Lengths, With Applications to the Theory of Competing Risks". Ann. Statist. 9 (1981), 157-167.
Marshall, A.W., and Olkin, I. (1967). "A Multivariate Exponential Distribution". JASA 63 (1967), 30-44.
Miller, D.R. (1977). "A Note on Independence of Multivariate Lifetimes in Competing Risks Models. Ann.Statist. 5 (1977) 576-579.
Moeschberger, M.L. (1974). "Life Tests Under Competing Causes of Failure". Technometrics 16 (1974), 39-47.
Nadas, A. (1970). "On Proportional Hazard Functions." Technometrics 12 (1970), 413-416.
Nadas, A. (1971). "The Distribution of the Identified Minimum of a Normal Pair determines the Distribution of the Pair." Technometrics 13 (1971), 201-202.
Nelson, W. (1970). "Hazard Plotting Methods for Analysis of Life Data with Different Failure Modes." J.Quality Technology 2 (1970), 126-149.
Nelson, W. (1972). "Theory and Applications of Hazard Plotting for Censored Failure Data". Technometrics 14 (1972), 945-966.

Nollau, V. (1981). Semi-Markovsche Prozesse. Vlg. Harri Deutsch, Thun 1981.
ÖStZ, Österreichisches Statistisches Zentralamt (1974). Berufslaufbahn.Ergebnisse des Mikrozensus September 1972. Wien 1974.
Piore, M.J. (1978). Lernprozesse, Mobilitätsketten und Arbeitsmarktsegmente. In W. Sengenberger: Der gespaltene Arbeitsmarkt. Campus Verlag, Frankfurt 1978.
Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N.,
Farewell, V.T., and Breslow, N.E. (1978). "The Analysis of Failure Times in the Presence of Competing Risks. Biometrics 34 (1978), 541-554.
Prentice, R.L., Williams, B.J., and Peterson, A.V. (1981). "On the Regression Analysis of Multivariate Failure Time Data." Biometrika 68 (1981), 373-379.
Stewman, S. (1976). "Markov Models of Occupational Mobility: Theoretical Development and Empirical Support. Part 1: Careers." J.Math.Sociol. 4 (1976), 201-245.
Tsiatis, A. (1975). "A Nonidentifiability Aspect of the Problem of Competing Risks". Proc.Nat.Acad.Sci. (1975), 20-22.