



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Fahrmeir, Wagenpfeil:

Smoothing Hazard Functions and Time-Varying Effects in Discrete Duration and Competing Risks Models

Sonderforschungsbereich 386, Paper 7 (1995)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Smoothing Hazard Functions and Time-Varying Effects in Discrete Duration and Competing Risks Models

Ludwig FAHRMEIR and Stefan WAGENPFEIL

Address for correspondence: Prof. Fahrmeir
Institute of Statistics,
University of Munich
Ludwigstr. 33 / II
80539 Munich, Germany

Tel.: +89 / 2180 - 2220

Fax: +89 / 2180 - 3804

Email: ua311aa@sunmail.lrz-muenchen.de

Footnote:

Ludwig Fahrmeir is Professor of Statistics, and Stefan Wagenpfeil is Research Assistant in the Sonderforschungsbereich 386, Statistische Analyse diskreter Strukturen, Institute of Statistics, University of Munich, 80539 Munich, Germany. This research was supported by the German Science Foundation (DFG). The authors thank Leonhard Knorr-Held and the two referees for valuable comments.

State space or dynamic approaches to discrete or grouped duration data with competing risks or multiple terminating events allow simultaneous modelling and smooth estimation of hazard functions and time-varying effects in a flexible way. Full Bayesian or posterior mean estimation, using numerical integration techniques or Monte Carlo methods, can become computationally rather demanding or even infeasible for higher dimensions and larger data sets. Therefore, based on previous work on filtering and smoothing for multicategorical time series and longitudinal data, our approach uses posterior mode estimation. Thus we have to maximize posterior densities or, equivalently, a penalized likelihood, which enforces smoothness of hazard functions and time-varying effects by a roughness penalty. Dropping the Bayesian smoothness prior and adopting a nonparametric viewpoint, one might also start directly from maximizing this penalized likelihood. We show how Fisher scoring smoothing iterations can be carried out efficiently by iteratively applying linear Kalman filtering and smoothing to a working model. This algorithm can be combined with an EM-type procedure to estimate unknown smoothing- or hyperparameters. The methods are applied to a larger set of unemployment duration data with one and, in a further analysis, multiple terminating events from the German socio-economic panel GSOEP.

KEY WORDS: Fisher scoring; Iteratively weighted Kalman smoothing; Multiple modes of the terminating event; Penalized likelihood; Posterior mode smoothing; Survival analysis.

1. INTRODUCTION

In many applications duration or survival times are not observed continuously, but are only known to lie between a pair of consecutive follow ups. A typical example are data on duration of unemployment in the German socio-economic panel GSOEP, where time is measured in months. Often there are, say m , different types of terminating causes or competing risks as in our application to unemployment data from the GSOEP in Section 4, where we

distinguish between full-time jobs, part-time jobs and other causes for end of unemployment. Of course, in this example competing "chances" would be more appropriate than competing "risks".

In the following let time be divided into k intervals $[b_0, b_1), [b_1, b_2), \dots, [b_{q-1}, b_q), [b_q, \infty)$ of equal length where $q=k-1$. For the first interval $b_0 = 0$ may be assumed and b_q denotes the final follow up. Instead of continuous time, discrete time $T \in \{1, \dots, k\}$ is observed where $T = t$ denotes end of duration within the interval $[b_{t-1}, b_t)$. Distinct terminating causes, competing risks or types of failure are denoted by $R \in \{1, \dots, m\}$. Classical formulations of competing risks models introduce latent durations T_1, \dots, T_m , one for each terminating cause. The observed duration T and terminating cause R are then viewed as $T = \min(T_1, \dots, T_m)$ and $R = r$ if $T = T_r$. This classical concept is quite natural for example in an industrial study, where T_1, \dots, T_m are failure times of different components of a system or device, and it may or may not be meaningful in a medical study. However, in social sciences it generally does not appear to be meaningful. For example, in our application to duration of unemployment, the notion of a time it would take an unemployed person to get a part-time job given he is precluded to get a full-time job is not meaningful or of prime interest. Therefore, as e.g. in Prentice et al. (1978) and Lancaster (1990, pp. 99), we formulate duration and competing risks models in terms of cause-specific hazard functions as the basic characteristics for the observables T and R . The cause-specific discrete hazard function resulting from cause or risk r is given by the conditional probability

$$\lambda_r(t|\mathbf{x}_t) = \text{pr}(T = t, R = r | T \geq t, \mathbf{x}_t), \quad (1.1)$$

$r=1, \dots, m, t=1, \dots, q$, where \mathbf{x}_t is a vector of, possibly time-dependent, covariates. The overall hazard function, regardless of cause, is given by

$$\lambda(t|\mathbf{x}_t) = \sum_{r=1}^m \lambda_r(t|\mathbf{x}_t) = P(T = t | T \geq t, \mathbf{x}_t).$$

Modelling of cause-specific hazard functions can be based on multicategorical response models. A first candidate for unordered events is the multinomial logit model

$$\lambda_r(t|\mathbf{x}_t) = \frac{\exp(\gamma_{tr} + \mathbf{x}_t' \beta_r)}{1 + \sum_{i=1}^m \exp(\gamma_{ti} + \mathbf{x}_t' \beta_i)}. \quad (1.2)$$

The parameters $\gamma_{1r}, \dots, \gamma_{qr}$ represent the cause-specific baseline hazard function and β_r is the cause-specific effect. If the events are ordered, ordinal models like cumulative or sequential models are appropriate.

If there is only one type of terminating event, we have the situation of discrete or grouped survival data. Then (1.1) reduces to the discrete hazard function

$$\lambda(t|\mathbf{x}_t) = \text{pr}(T = t | T \geq t, \mathbf{x}_t),$$

$t = 1, \dots, q$, and (1.2) to the binary logit model

$$\lambda(t|x_t) = \frac{\exp(\gamma_t + x_t' \beta)}{1 + \exp(\gamma_t + x_t' \beta)}, \quad (1.3)$$

see e.g. Thompson (1977), Arjas and Haara (1987). Alternatively, one may consider the grouped proportional hazards or Cox model

$$\lambda(t|x_t) = 1 - \exp\{-\exp(\gamma_t + x_t' \beta)\}, \quad (1.4)$$

see e.g. Kalbfleisch and Prentice (1980). If the intervals are short, the models become very similar as has been shown by Thompson (1977). A detailed survey on discrete time duration data can also be found in Hamerle and Tutz (1989), a shorter introduction is in Fahrmeir and Tutz (1994, Ch.9).

Common static modelling and estimation, which treats baseline hazard coefficients and covariate parameters as fixed effects, are appropriate if the number of intervals is

comparatively small. In situations with many intervals - but not enough to apply models for continuous time - such unrestricted modelling and fitting of hazard functions will often lead to nonexistence and divergence of ML estimates due to the large number of parameters. This difficulty in real data problems becomes even more apparent if in addition covariate effects are assumed to be time-varying, as for example the effect of nationality, sex and age in our application to duration of unemployment. One may try to avoid such problems by a more parsimonious parameterization, e.g. using piecewise polynomials for hazard functions or varying effects. However, by imposing such parametric forms one may overlook unexpected patterns like peaks or seasonal effects.

In this paper we propose state space or dynamic models as a flexible technique, which makes simultaneous modelling and smooth estimation of hazard functions and covariate effects possible. The development is related to a dynamic version of the piecewise exponential model and extensions to point processes studied by Gamerman (1991, 1992) and, more closely, to dynamic grouped survival models with only one terminating event (Fahrmeir 1994), where a generalized Kalman filter and smoother (GKFS) is proposed for estimating hazard functions and time-varying effects. Here we extend this approach to models with multiple terminating events (Section 2) and develop a numerically efficient Fisher scoring smoothing algorithm (Section 3). It is obtained by extending iterative Kalman-type techniques for multicategorical time series (Fahrmeir and Tutz, 1994 Ch.8; Fahrmeir and Wagenpfeil, 1995) to the present situation. The smoothing algorithms can be derived as posterior mode estimators or, from a nonparametric point of view, as penalized likelihood estimators. For only one terminating event ($m=1$), they generally improve (GKFS) with regard to numerical accuracy and approximation quality. Data-driven choice of smoothing- or hyperparameters can be achieved by an EM-type algorithm or by cross-validation.

2. STATE SPACE MODELS FOR DISCRETE DURATION TIMES AND COMPETING RISKS

For individual units $i=1,\dots,n$, the data are given by (t_i, δ_i, r_i) , where t_i is the observed discrete duration time, δ_i is the censoring indicator with

$$\delta_i = \begin{cases} 1, & \text{failure in } [b_{t_i-1}, b_{t_i}) \\ 0, & \text{censoring in } [b_{t_i-1}, b_{t_i}) \end{cases},$$

and $r_i \in \{1, \dots, m\}$ indicates the terminating event. In addition, possibly time-varying covariates $\{x_{it}, t=1, \dots, t_i\}$ are observed. For the formulation of dynamic models and of smoothing algorithms we rewrite the data in different notation, similarly as in Arjas and Haara (1987) and Fahrmeir (1994). We define risk indicators u_{it} , $i=1, \dots, n$, $t=1, \dots, q$ by

$$u_{it} = \begin{cases} 1, & \text{if unit } i \text{ is at risk in } [b_{t-1}, b_t) \\ 0, & \text{otherwise} \end{cases},$$

the risk vector $u_t = (u_{it}, i=1, \dots, n) \in \mathbb{R}^n$, and the risk set $R_t = \{i: u_{it} = 1\}$ at t . We assume that censoring occurs at the end of the interval, so that the risk set R_t includes all individuals who are censored in $[b_{t-1}, b_t)$. Furthermore we introduce the column vectors

$$y_{it} = (y_{it1}, \dots, y_{itm})',$$

$i \in R_t$, $t=1, \dots, t_i$, of failure indicators

$$y_{itr} = \begin{cases} 1, & \text{failure of type } r \text{ occurs in } [b_{t-1}, b_t) \\ 0, & \text{no failure of type } r \text{ occurs in } [b_{t-1}, b_t) \end{cases},$$

$r=1, \dots, m$. If there is only one type of risk or failure ($m=1$) we simply write y_{it} for y_{it1} . We gather covariates and failure indicators of time interval t , that is $[b_{t-1}, b_t)$, in the column vectors

$$\mathbf{x}_t = (\mathbf{x}_{it}, \mathbf{i} \in \mathbf{R}_t), \mathbf{y}_t = (\mathbf{y}_{it}, \mathbf{i} \in \mathbf{R}_t)$$

and denote histories of covariates, failure and risk indicators up to t by

$$\mathbf{x}_t^* = (\mathbf{x}_1, \dots, \mathbf{x}_t), \mathbf{y}_t^* = (\mathbf{y}_1, \dots, \mathbf{y}_t), \mathbf{u}_t^* = (\mathbf{u}_1, \dots, \mathbf{u}_t).$$

Dynamic or state space duration models are now introduced in two stages: The observation model defines the conditional likelihood of the observations, given unknown baseline hazard parameters γ_t and a possibly time-varying vector β_t of covariate effects. It is supplemented by a stochastic transition model for these parameters.

Observation model

The duration process is modelled as a sequence of categorical responses \mathbf{y}_{it} , with response probabilities in form of a categorical regression model. For the case of only one type of risk ($m=1$), we assume for $\mathbf{i} \in \mathbf{R}_t$ that, conditional on parameters γ_t, β_t and the history $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*$,

$$\text{pr}(\mathbf{y}_{it} = 1 | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*, \gamma_t, \beta_t) = \mathbf{h}(\eta_{it}), \quad (2.1)$$

where $\mathbf{h}: \mathbb{R} \rightarrow (0,1)$ is a response function and

$$\eta_{it} = \gamma_t + \mathbf{w}_{it}' \beta_t \quad (2.2)$$

a linear predictor. In (2.2) the design vector \mathbf{w}_{it} is an appropriate function of $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*$. Often, \mathbf{w}_{it} will only be a simple function of covariates \mathbf{x}_{it} of object i . However, the model also gives the possibility to introduce the history \mathbf{y}_{t-1}^* of time-dependent failure indicators in form of covariates, a suggestion already made in Prentice et al. (1978). Common response functions $\mathbf{h}(\cdot)$ are the logistic function $\mathbf{h}(\cdot) = \exp(\cdot) / \{1 + \exp(\cdot)\}$ for the logit model (1.3), and $\mathbf{h}(\cdot) = 1 - \exp\{-\exp(\cdot)\}$ for the grouped Cox model (1.4). For competing risks ($m>1$) we assume

$$\text{pr}(\mathbf{y}_{itr} = 1 | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*, \gamma_t, \beta_t) = \mathbf{h}_r(\eta_{it}), \quad (2.3)$$

with response function $\mathbf{h}_r: \mathbb{R} \rightarrow (0,1)$ and linear prediction vector $\eta_{it} = (\eta_{it1}, \dots, \eta_{itm})$. For the multinomial logit model (1.2),

$$\mathbf{h}_r(\eta_{it}) = \frac{\exp(\eta_{itr})}{1 + \sum_{j=1}^m \exp(\eta_{itj})}, \quad (2.4)$$

$\eta_{itr} = \gamma_{tr} + \mathbf{w}'_{it}\beta_{tr}$, and $\gamma_t = (\gamma_{t1}, \dots, \gamma_{tm})'$, $\beta_t = (\beta_{t1}, \dots, \beta_{tm})'$. Again, the design vector \mathbf{w}_{it} is an appropriate function of $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*$, where \mathbf{y}_{t-1}^* is now the history of multicategorical failure indicators. If latent failure times are considered, this allows, at least in principle, to incorporate effects of censoring by the cause of failure for the other causes. Other multicategorical response models can also be written in the general form (2.3), see e.g. Fahrmeir and Tutz (1994, Ch. 3 and Section 9.3). Introducing a corresponding design vector \mathbf{z}_{it} , hazard functions and response probabilities for \mathbf{y}_{it} can be written as

$$\lambda_r(t|\mathbf{z}_{it}; \alpha_t) = h_r(\eta_{it}), \quad r = 1, \dots, m, \quad (2.5)$$

with $\eta_{it} = (\mathbf{z}'_{it1}\alpha_t, \dots, \mathbf{z}'_{itm}\alpha_t)' = \mathbf{Z}_{it}\alpha_t$ and design matrix

$$\mathbf{Z}_{it} = \begin{bmatrix} \mathbf{z}'_{it1} \\ \vdots \\ \mathbf{z}'_{itm} \end{bmatrix}.$$

For example, we have $\mathbf{z}'_{it} = (\mathbf{1}, \mathbf{w}'_{it})$ and $\alpha_t = (\gamma_t, \beta_t)'$ in the binary model (2.1), (2.2), and

$$\mathbf{Z}_{it} = \begin{bmatrix} \mathbf{1} & & \mathbf{0} & \mathbf{w}'_{it} & & \mathbf{0} \\ & \ddots & & & \ddots & \\ \mathbf{0} & & \mathbf{1} & \mathbf{0} & & \mathbf{w}'_{it} \end{bmatrix}$$

and $\alpha_t = (\gamma_{t1}, \dots, \gamma_{tm}, \beta'_{t1}, \dots, \beta'_{tm})'$ in the multinomial logit model (2.4).

Transition model

Let us first assume that time intervals are of the same length. Simple models for stochastic variation of baseline hazard parameters $\{\gamma_t\}$ are first or second order random walks

$$\gamma_t = \gamma_{t-1} + \tau_t \quad \text{or}$$

$$\gamma_t = 2\gamma_{t-1} - \gamma_{t-2} + \tau_t, \quad \tau_t \sim N(\mathbf{0}, \sigma_\tau^2)$$

and the local linear trend model

$$\gamma_t = \gamma_{t-1} + \lambda_{t-1} + \tau_t, \quad \lambda_t = \lambda_{t-1} + v_t, \quad v_t \sim N(\mathbf{0}, \sigma_v^2)$$

with mutually independent white noise processes $\{\tau_t\}, \{v_t\}$. For covariate effects the simplest choice is a first order random walk

$$\beta_t = \beta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, E), \quad E \text{ diagonal.}$$

The case of time-constant effects is obtained as a special case by setting corresponding variances in E to zero. Time intervals of unequal length can be handled by introducing variances $\sigma_\tau^2 \mathbf{h}_t, \sigma_v^2 \mathbf{h}_t$ etc., where \mathbf{h}_t is the length of $[b_{t-1}, b_t)$, $t = 1, \dots, q$. All these models can be written in the general form

$$\alpha_t = \mathbf{F}\alpha_{t-1} + \xi_t, \tag{2.6}$$

where the state vector α_t includes γ_t and β_t . F is a nonrandom transition matrix, $\{\xi_t \sim N(\mathbf{0}, \mathbf{Q}_t)\}$ is a white noise sequence which is independent of the initial state $\alpha_0 \sim N(\mathbf{a}_0, \mathbf{Q}_0)$. A joint first order random walk model for $\alpha_t = (\gamma_t, \beta_t)'$ is obtained by setting $\mathbf{F}=\mathbf{I}$, a second order random walk model for γ_t can be defined by

$$\alpha_t = \begin{pmatrix} \gamma_t \\ \gamma_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \gamma_{t-1} \\ \gamma_{t-2} \end{pmatrix} + \begin{pmatrix} \tau_t \\ \mathbf{0} \end{pmatrix}.$$

More complex forms, e.g. seasonal components, can also be written in state space notation as for linear Gaussian models (Harvey 1989).

For a complete model specification in terms of joint densities, we add the following basic assumptions:

(A1) Given $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*$ and α_t , individual failure indicators $y_{it}, i \in \mathbf{R}_t$ are conditionally independent within $[\mathbf{b}_{t-1}, \mathbf{b}_t)$, i.e.

$$\mathbf{p}(\mathbf{y}_t | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*; \alpha_t) = \prod_{i \in \mathbf{R}_t} \mathbf{p}(y_{it} | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*; \alpha_t).$$

(A2) Given $\mathbf{y}_{t-1}^*, \mathbf{x}_{t-1}^*, \mathbf{u}_{t-1}^*$, covariate \mathbf{x}_t and risk vector \mathbf{u}_t are independent of

$$\alpha_{t-1}^* = (\alpha'_0, \alpha'_1, \dots, \alpha'_{t-1})'.$$

(A3) Conditional on $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*$ and α_t , failure indicators \mathbf{y}_t are independent of α_{t-1}^* , i.e.

$$\mathbf{p}(\mathbf{y}_t | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*; \alpha_t^*) = \mathbf{p}(\mathbf{y}_t | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*; \alpha_t).$$

For fixed parameters, (A1) corresponds to Assumption 2 in Arjas and Haara (1987). It is weaker than the usual unconditional independence assumption for units, since it allows for interaction via the common history, and it is likely to hold if a common cause for failures is incorporated in the covariate process. Assumption (A2) corresponds to Assumption 1 of Arjas and Haara (1987). It will generally hold for non-informative censoring and external or time-independent covariates. Assumption (A3) is basic in state-space modelling. It says that conditional information of α_t^* on \mathbf{y}_t is already contained in the current parameter α_t alone, and is mostly not stated explicitly in the case of fixed parameters. The Markovian assumption in (2.6) for the transition models is also quite common in state space modelling. It implicitly assumes that

$$\mathbf{p}(\alpha_t | \alpha_{t-1}^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{u}_t^*) = \mathbf{p}(\alpha_t | \alpha_{t-1}) \quad (2.7)$$

with $\mathbf{p}(\alpha_t | \alpha_{t-1}) \sim \mathbf{N}(\mathbf{F}\alpha_{t-1}, \mathbf{Q}_t)$, since \mathbf{F} and \mathbf{Q}_t do not depend on $\mathbf{y}_{t-1}^*, \mathbf{x}_t^*$.

3. ESTIMATION OF TIME-VARYING PARAMETERS WITH FISHER SCORING

Let the hyperparameters Q_t , Q_0 and a_0 be known and fixed for the moment. The following smoothing criterion and algorithm are derived for competing risks. Survival data with only one terminating event ($m = 1$) are a special case.

To estimate the time-varying parameters α_t , a full Bayesian solution would be to compute the posterior densities $p(\alpha_t | y_k^*, x_k^*, u_k^*)$. Because our observation model is binary or multinomial, and thus non-normal, direct numerical or Monte Carlo integration techniques can become infeasible for higher dimensions. Sampling-based approaches like MCMC simulation, which have recently been applied to dynamic generalized linear models for the pure time series situation, see e.g. Gamerman (1995), would be more promising, but have still to be developed for the models considered in this paper. We concentrate on posterior mode smoothing as suggested in Fahrmeir (1992). The iteratively weighted Kalman filter and smoother (IWKFS), an efficient Fisher scoring algorithm for penalized loglikelihood estimation presented in Fahrmeir and Wagenpfeil (1995) and shortly described in Fahrmeir and Tutz (1994, Ch. 8), is extended to the present context.

With $\alpha = \alpha_k^* = (\alpha'_0, \alpha'_1, \dots, \alpha'_k)'$, the posterior mode estimate is defined as

$$a = (a'_{0|k}, a'_{1|k}, \dots, a'_{k|k})' := \arg \max_{\alpha} \left\{ p(\alpha | y_k^*, x_k^*, u_k^*) \right\}.$$

Repeated application of Bayes' theorem and the model assumptions gives

$$p(\alpha | y_k^*, x_k^*, u_k^*) \propto \prod_{t=1}^k \prod_{i \in R_t} p(y_{it} | \alpha_t, y_{t-1}^*, x_t^*, u_t^*) \prod_{t=1}^k p(\alpha_t | \alpha_{t-1}) \cdot p(\alpha_0).$$

Incorporating the observation and transition models defined by (2.3), (2.6), assuming for the moment that Q_0 and Q_t are regular and taking logarithms, we obtain the penalized loglikelihood function

$$\begin{aligned}
\text{PL}(\alpha) = l(\alpha) - \frac{1}{2} \sum_{t=1}^k (\alpha_t - F\alpha_{t-1})' Q_t^{-1} (\alpha_t - F\alpha_{t-1}) \\
- \frac{1}{2} (\alpha_0 - a_0)' Q_0^{-1} (\alpha_0 - a_0)
\end{aligned} \tag{3.1}$$

with $l(\alpha) := \sum_{t=1}^k \sum_{i \in R_t} I_{it}(\alpha_t)$ and the individual log-likelihood contribution

$$I_{it}(\alpha_t) := \left(1 - \sum_{r=1}^m y_{itr} \right) \log \left\{ 1 - \sum_{r=1}^m h_r(\eta_{it}) \right\} + \sum_{r=1}^m y_{itr} \log h_r(\eta_{it}).$$

Thus posterior mode smoothers $\mathbf{a} = (a'_{0k}, a'_{1k}, \dots, a'_{kk})'$ are maximum penalized likelihood estimators, i.e.

$$\mathbf{a} = \underset{\alpha}{\text{argmax}} \{ \text{PL}(\alpha) \}, \tag{3.2}$$

in the framework of our model (2.3), (2.6).

One may, however, drop the Bayesian viewpoint and directly start from the penalized loglikelihood criterion (3.2). Then α could be regarded as a sequence of fixed but unknown parameters and $\text{PL}(\alpha)$ as a function of the variable α . The first term in (3.1) measures the deviance between the data and the fit. The second and third terms penalize deviations from the underlying transition mechanism for the parameters, thus acting as roughness penalties and enforcing smoothness of the parameter sequence.

A numerical solution of the nonlinear programming problem involved in (3.2) could in principle be obtained by various algorithms. In the following, we will derive a Fisher scoring algorithm for smoothing, where iteration steps can be performed efficiently by applying Kalman filtering and smoothing to a "working model", similarly as maximum likelihood estimation in static generalized linear models can be carried out by applying iteratively weighted least squares to "working observations". For $t = 1, \dots, k$ let

$$l_t(\boldsymbol{\alpha}_t) = \sum_{i \in R_t} l_{it}(\boldsymbol{\alpha}_t)$$

denote the log-likelihood contribution of individuals which are still under risk in time interval t .

Corresponding contributions for the score function and expected information matrix are

$$\begin{aligned} s_t(\boldsymbol{\alpha}_t) &= \partial l_t(\boldsymbol{\alpha}_t) / \partial \boldsymbol{\alpha}_t = \\ &= \sum_{i \in R_t} \mathbf{Z}'_{it} \mathbf{D}_{it}(\boldsymbol{\alpha}_t) \boldsymbol{\Sigma}_{it}^{-1}(\boldsymbol{\alpha}_t) \{y_{it} - \lambda_{it}(\boldsymbol{\alpha}_t)\} \end{aligned} \quad (3.3)$$

and

$$\begin{aligned} \mathbf{S}_t(\boldsymbol{\alpha}_t) &= \mathbf{E}\{-\partial^2 l_t(\boldsymbol{\alpha}_t) / \partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}_t'\} = \\ &= \sum_{i \in R_t} \mathbf{Z}'_{it} \mathbf{D}_{it}(\boldsymbol{\alpha}_t) \boldsymbol{\Sigma}_{it}^{-1}(\boldsymbol{\alpha}_t) \mathbf{D}'_{it}(\boldsymbol{\alpha}_t) \mathbf{Z}_{it}, \end{aligned} \quad (3.4)$$

where $\lambda_{it}(\boldsymbol{\alpha}_t) = \mathbf{h}(\boldsymbol{\eta}_{it}) = \{h_1(\boldsymbol{\eta}_{it}), \dots, h_m(\boldsymbol{\eta}_{it})\}'$ are individual hazard functions, $\mathbf{D}_{it}(\boldsymbol{\alpha}_t) = \partial \mathbf{h}(\boldsymbol{\eta}_{it}) / \partial \boldsymbol{\eta}$ is the Jacobian of $\mathbf{h}(\boldsymbol{\eta}) = \{h_1(\boldsymbol{\eta}), \dots, h_m(\boldsymbol{\eta})\}'$ evaluated at $\boldsymbol{\eta}_{it}$, and $\boldsymbol{\Sigma}_{it}(\boldsymbol{\alpha}_t)$ is the multinomial covariance matrix. The inverse $\boldsymbol{\Sigma}_{it}^{-1}(\boldsymbol{\alpha}_t)$ is available explicitly:

$$\boldsymbol{\Sigma}_{it}^{-1}(\boldsymbol{\alpha}_t) = \frac{1}{\sum_{r=1}^m h_r(\boldsymbol{\eta}_{it})} \mathbf{1} + \text{diag}\left\{\frac{1}{h_1(\boldsymbol{\eta}_{it})}, \dots, \frac{1}{h_m(\boldsymbol{\eta}_{it})}\right\},$$

where $\mathbf{1}$ is a matrix with all elements equal to 1, cf. Kredler (1986).

To describe Fisher scoring iterations in compact matrix notation, it is convenient to define

$$l_0(\boldsymbol{\alpha}_0) = -\frac{1}{2}(\boldsymbol{\alpha}_0 - \mathbf{a}_0)' \mathbf{Q}_0^{-1}(\boldsymbol{\alpha}_0 - \mathbf{a}_0)$$

and

$$s_0(\boldsymbol{\alpha}_0) = \partial l_0(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha}_0 = \mathbf{Q}_0^{-1}(\mathbf{a}_0 - \boldsymbol{\alpha}_0),$$

$$\mathbf{S}_0 = \mathbf{E}\{-\partial^2 l_0(\boldsymbol{\alpha}_0) / \partial \boldsymbol{\alpha}_0 \partial \boldsymbol{\alpha}_0'\} = \mathbf{Q}_0^{-1}.$$

In matrix notation, (3.1) can then be rewritten as

$$PL(\alpha) = \tilde{l}(\alpha) - \frac{1}{2}\alpha'K\alpha \quad (3.5)$$

with $\tilde{l}(\alpha) = l(\alpha) + l_0(\alpha_0)$ and the symmetric and block-tridiagonal penalty matrix

$$K = \begin{bmatrix} K_{00} & K_{01} & & & 0 \\ K_{10} & K_{11} & K_{12} & & \\ & K_{21} & \ddots & \ddots & \\ & & \ddots & \ddots & K_{k-1,k} \\ 0 & & & K_{k,k-1} & K_{kk} \end{bmatrix}$$

with

$$\begin{aligned} K_{t-1,t} &= K'_{t,t-1}, 1 \leq t \leq k, \\ K_{00} &= F'_1 Q_1^{-1} F_1, \\ K_{tt} &= Q_t^{-1} + F'_{t+1} Q_{t+1}^{-1} F_{t+1}, 1 \leq t \leq k, \\ F_{k+1} &= 0, \\ K_{t-1,t} &= -F'_t Q_t^{-1}, 1 \leq t \leq k. \end{aligned}$$

To express the score function $s(\alpha) = \partial \tilde{l}(\alpha) / \partial \alpha = \{s_0(\alpha_0)', s_1(\alpha_1)', \dots, s_k(\alpha_k)'\}'$ and the expected information matrix $S(\alpha) = E\{-\partial^2 \tilde{l}(\alpha) / \partial \alpha \partial \alpha'\} = \text{diag}\{S_0, S_1(\alpha_1), \dots, S_k(\alpha_k)\}$ in matrix form, we define the observation column vector y , augmented by a_0 , as

$$y' = (a'_0, y'_1, \dots, y'_k),$$

and the vector of hazard functions, augmented by α_0 , as

$$\lambda(\alpha)' = \{\alpha'_0, \lambda_1(\alpha_1)', \dots, \lambda_k(\alpha_k)'\}$$

with the column vector $\lambda_t(\alpha_t) = \{\lambda_{it}(\alpha_t), i \in R_t\}$ according to (2.5). The block-diagonal inverse covariance matrix is

$$\Sigma^{-1}(\alpha) = \text{diag}\{Q_0^{-1}, \Sigma_1^{-1}(\alpha_1), \dots, \Sigma_k^{-1}(\alpha_k)\}$$

with $\Sigma_t^{-1}(\alpha_t) = \text{diag}\{\Sigma_{it}^{-1}(\alpha_t), i \in R_t\}$. Jacobians are gathered in the block-diagonal matrix

$$D(\alpha) = \text{diag}\{I, D_1(\alpha_1), \dots, D_k(\alpha_k)\},$$

where $D_t(\alpha_t) = \text{diag}\{D_{it}(\alpha_t), i \in R_t\}$ contains individual Jacobians as diagonal blocks.

Furthermore we introduce the block-diagonal, rectangular design matrix

$$Z = \text{diag}(I, Z_1, \dots, Z_k),$$

where I is the unit matrix and

$$Z_t = \begin{pmatrix} Z_{ht} \\ \vdots \\ Z_{it} \\ \vdots \\ Z_{jt} \end{pmatrix}$$

contains all design matrices $Z_{ht}, \dots, Z_{it}, \dots, Z_{jt}$ of individuals h, \dots, i, \dots, j still in the risk set R_t .

The score function $s(\alpha) = \{s_0(\alpha_0)', s_1(\alpha_1)', \dots, s_k(\alpha_k)'\}'$ of the log-likelihood $\tilde{l}(\alpha)$ in (3.5) is then

$$s(\alpha) = Z'D(\alpha)\Sigma^{-1}(\alpha)\{y - \lambda(\alpha)\},$$

and the block-diagonal (expected) information matrix $S(\alpha) = \text{diag}\{S_0, S_1(\alpha_1), \dots, S_k(\alpha_k)\}$ is

$$S(\alpha) = Z'W(\alpha)Z$$

with block-diagonal weight matrix

$$W(\alpha) = \text{diag}\{W_0, W_1(\alpha_1), \dots, W_k(\alpha_k)\} = D(\alpha)\Sigma^{-1}(\alpha)D'(\alpha).$$

The first derivative of $PL(\alpha)$ in (3.5) can be written as

$$u(\alpha) = \partial PL(\alpha)/\partial \alpha = s(\alpha) - K\alpha$$

and the block-tridiagonal expected information matrix is

$$U(\alpha) = E\{-\partial^2 PL(\alpha)/\partial \alpha \partial \alpha'\} = S(\alpha) + K.$$

A single Fisher-scoring step from the current iterate $\alpha^0 = \{(\alpha_0^0)', (\alpha_1^0)', \dots, (\alpha_k^0)'\}'$, say, to the next iterate $\alpha^1 = \{(\alpha_0^1)', (\alpha_1^1)', \dots, (\alpha_k^1)'\}'$ is then

$$U(\alpha^0)\{\alpha^1 - \alpha^0\} = u(\alpha^0) .$$

This can be rewritten as

$$\alpha^1 = \{Z'W(\alpha^0)Z + K\}^{-1} Z'W(\alpha^0)\tilde{y}(\alpha^0) , \quad (3.6)$$

with "working observation"

$$\tilde{y}(\alpha^0) = D^{-1}(\alpha^0)\{y - \lambda(\alpha^0)\} + Z\alpha^0 .$$

The linear equation system (3.6) has the same structure as the one obtained for a single Fisher scoring iteration for exponential family state space models, cf. Fahrmeir and Tutz (1994, Ch. 8.3.1). To solve (3.6) in a numerically efficient way, that is without explicitly inverting the block-tridiagonal system matrix $U(\alpha^0)$, Fahrmeir and Wagenpfeil (1995) developed a "working Kalman filter and smoother" (WKFS). We extend this method for the competing risks data situation and arrive at the following "working Kalman filter and smoother" (WKFS):

(WKFS) for competing risks:

Initialization: $a_{0|0} = a_0 , V_{0|0} = Q_0 .$

For $t = 1, \dots, k$:

$$\begin{aligned} \text{prediction step: } a_{t|t-1} &= F_t a_{t-1|t-1}, \\ V_{t|t-1} &= F_t V_{t-1|t-1} F_t' + Q_t. \\ \text{correction step: } V_{t|t} &= \{V_{t|t-1}^{-1} + S_t(\alpha_t^0)\}^{-1}, \\ a_{t|t} &= a_{t|t-1} + V_{t|t} \tilde{s}_t(\alpha_t^0), \end{aligned}$$

with the "working score function" $\tilde{s}_t(\alpha_t^0) := s_t(\alpha_t^0) - S_t(\alpha_t^0)\{a_{t|t-1} - \alpha_t^0\}$ and $s_t(\alpha_t^0)$, $S_t(\alpha_t^0)$

from (3.3), (3.4). For smoothing we may use the classical fixed interval smoother

For $t = k, \dots, 1$:

$$\begin{aligned} \mathbf{a}_{t-1|T} &= \mathbf{a}_{t-1|t-1} + \mathbf{B}_t (\mathbf{a}_{t|T} - \mathbf{a}_{t|t-1}), \\ \mathbf{V}_{t-1|T} &= \mathbf{V}_{t-1|t-1} + \mathbf{B}_t (\mathbf{V}_{t|T} - \mathbf{V}_{t|t-1}) \mathbf{B}_t', \end{aligned}$$

where $\mathbf{B}_t = \mathbf{V}_{t-1|t-1} \mathbf{F}_t' \mathbf{V}_{t|t-1}^{-1}$, yielding $\boldsymbol{\alpha}^1 = (\mathbf{a}'_{0|k}, \mathbf{a}'_{1|k}, \dots, \mathbf{a}'_{k|k})'$.

Remark:

- (i) We supposed that \mathbf{Q}_0 and \mathbf{Q}_t are regular so that the penalized loglikelihood (3.1) is well-defined. However, in (WKFS) these covariance matrices from the transition model (2.6) are not inverted and thus \mathbf{Q}_0 and \mathbf{Q}_t are allowed to be singular within (WKFS).
- (ii) The correction step is given in "scoring" form, i.e. in terms of $s_t(\boldsymbol{\alpha}_t)$, $\mathbf{S}_t(\boldsymbol{\alpha}_t)$. Applying the matrix inversion lemma, e.g. from Anderson and Moore (1979), the Kalman gain form of (WKFS) can be obtained.

As we want to solve (3.2), we have to iterate (WKFS), where the solution $\boldsymbol{\alpha}^{(j)} = \{(\boldsymbol{\alpha}_0^{(j)})', (\boldsymbol{\alpha}_1^{(j)})', \dots, (\boldsymbol{\alpha}_k^{(j)})'\}'$ of the previous iteration is the starting vector for the next loop. This results in the final algorithm, an iteratively weighted Kalman filter and smoother (IWKFS):

(IWKFS) for competing risks:

Initialization: Compute $\boldsymbol{\alpha}^0 = (\mathbf{a}_{0|k}^0, \mathbf{a}_{1|k}^0, \dots, \mathbf{a}_{k|k}^0)'$ with the generalized extended Kalman filter and smoother (GKFS) presented in Fahrmeir (1994).

Set iteration index $j = 0$.

Step 1: Starting with $\boldsymbol{\alpha}^j$, compute $\boldsymbol{\alpha}^{j+1}$ by application of (WKFS) for competing risks.

Step 2: If a convergence criterion is fulfilled: STOP,
else set $j = j+1$ and go to Step 1.

(IWKFS) for competing risks allows for joint estimation of discrete hazard functions and time-varying effects in grouped duration and competing risks models by solving the nonlinear programming problem involved in (3.2) with all the advantages of a complete Fisher scoring algorithm. Furthermore we take into account the block-tridiagonal form of $U(\alpha)$ and therefore we save storage and computation time. The iterative process is conveniently initialized with (GKFS) since it does not require a starting vector. Only the hyperparameters, i.e. Q_0, a_0 and Q_t , from the transition model (2.6) have to be known.

In the following we outline the EM-type algorithm for hyperparameter estimation as already suggested in Fahrmeir (1992) in a related context. The procedure for joint estimation of α , Q_0, a_0 and $Q = Q_t$ for all $1 \leq t \leq k$ can be summarized as follows:

EM-type algorithm:

1. Choose starting values $Q^{(0)}, Q_0^{(0)}, a_0^{(0)}$ and set iteration index $p = 0$.
2. Smoothing: Compute $a_{t|k}^{(p)}, V_{t|k}^{(p)}$, $t = 1, \dots, k$ by (GKFS) or (IWKFS) for competing risks, with unknown parameters replaced by their current estimates

$$Q^{(p)}, Q_0^{(p)} \text{ and } a_0^{(p)}.$$

3. EM step: Compute $Q^{(p+1)}, Q_0^{(p+1)}$ and $a_0^{(p+1)}$ by

$$a_0^{(p+1)} = a_{0|k}^{(p)},$$

$$Q_0^{(p+1)} = V_{0|k}^{(p)},$$

$$Q^{(p+1)} = \frac{1}{k} \sum_{t=1}^k \left\{ \left(a_{t|k}^{(p)} - F_t a_{t-1|k}^{(p)} \right) \left(a_{t|k}^{(p)} - F_t a_{t-1|k}^{(p)} \right)' + V_{t|T}^{(p)} \right. \\ \left. - F_t B_t^{(p)} V_{t|k}^{(p)} - V_{t|k}^{(p)} B_t^{(p)'} F_t' + F_t V_{t-1|k}^{(p)} F_t' \right\}$$

with $B_t^{(p)}$ defined as in the fixed interval smoother.

4. If some termination criterion is reached: STOP, else set $p = p+1$ and go to 2.

Numerical experience shows that the estimate $\mathbf{a}_{0|k}$ can be quite sensitive to the choice of the starting value $\mathbf{a}_0^{(0)}$. If there is no or little prior information about α_0 , as in our application in Section 4, it is therefore reasonable to use a flat prior for α_0 , i.e. $\mathbf{Q}_0^{-1} = 0$. The initialization of the (WKFS)-algorithm has then to be modified to

$$\mathbf{V}_{1|1} = \{\mathbf{S}_1(\alpha_1^0)\}^{-1},$$

$$\mathbf{a}_{1|1} = \alpha_1^0 + \mathbf{V}_{1|1}\mathbf{s}_1(\alpha_1^0).$$

The prediction and the smoothing step for $t = 1$ can be dropped. Within the EM-type algorithm the formulas for $\mathbf{a}_0^{(p+1)}$ and $\mathbf{Q}_0^{(p+1)}$ are no more necessary. To compute $\mathbf{Q}^{(p+1)}$ we now have to sum from $t = 2$ to k only. This EM-type algorithm and (WKFS) for the diffuse case were applied to analyze the data in Section 4.

Other approaches for data-driven hyperparameter estimation such as cross-validation proposed by Kohn and Ansley (1989) for linear state space models and extended in Wagenpfeil (1995) to exponential family state space models can also be used.

As in the closely related situation of choosing smoothing parameters or bandwidths in non- or semiparametric methods, one should not use such estimates blindly. In particular for higher-dimensional problems, the surface corresponding to the criterion to be optimized may be rather flat around the maxima or minima, so that different values of hyperparameters can produce similar values of the criterion function. Therefore careful subjective modifications of data-driven estimates remain within the responsibility of the statistician.

4. APPLICATION: DURATION OF UNEMPLOYMENT

We analyze data on duration of unemployment of 1419 persons, older than 16 years, which are observed from January 1983 until December 1988 in the German socio-economic

panel GSOEP. Duration of unemployment is measured in months. The following covariates are included:

sex S: S = 1 for males, S = 0 for females;

nationality N: N = 1 for Germans, N = 0 for others;

age at the beginning of unemployment, grouped in four categories and coded by 0-1 dummies:

A1 = 1 for "age ≤ 30 years", 0 else;

A2 = 1 for "41 ≤ age ≤ 50 years", 0 else;

A3 = 1 for "age ≥ 51 years", 0 else;

with reference category "31 ≤ age ≤ 40 years" coded by (A1, A2, A3) = (0,0,0).

In a first analysis, only the terminating event "end of unemployment", regardless of a specific cause, is considered. We apply a binary dynamic logit model

$$\lambda(t|x_i) = \frac{\exp(\gamma_t + x_i' \beta_t)}{1 + \exp(\gamma_t + x_i' \beta_t)},$$

where the dummy vector $x_i' = (N_i, S_i, A1_i, A2_i, A3_i)$ describes nationality, sex and age of person i. We do not incorporate the history y_{t-1}^* of failure indicators in form of covariates. This means that end of unemployment duration up to t-1 for unemployed with covariate vector x_i does not affect the probabilities for end of unemployment for persons with the same covariate vector x_i . The baseline-effect γ_t and time-varying effects $\beta_t' = (\beta_{N,t}, \beta_{S,t}, \beta_{A1,t}, \beta_{A2,t}, \beta_{A3,t})$ are modelled by first order random walks. Smoothed estimates of $\{\gamma_t, \beta_t\}$ were obtained by the (IWKFS) - algorithm, combined with EM iterations. Final estimates for the random walk variances were $\hat{\sigma}_\gamma^2 = 0.196$, $\hat{\sigma}_N^2 = 0.026$, $\hat{\sigma}_S^2 = 0.017$, $\hat{\sigma}_{A1}^2 = 0.002$, $\hat{\sigma}_{A2}^2 = 0.005$, $\hat{\sigma}_{A3}^2 = 0.021$. The dotted lines in each graph represent pointwise confidence bands about the smoothed estimates. They were obtained, as in Fahrmeir and Wagenpfeil (1995), by taking $V_{t|T}$ ($t = 1, \dots, T$) from the last (IWKFS)-iteration, i.e. the diagonal blocks of the inverse Fisher information matrix $\{U(\mathbf{a})\}^{-1}$.

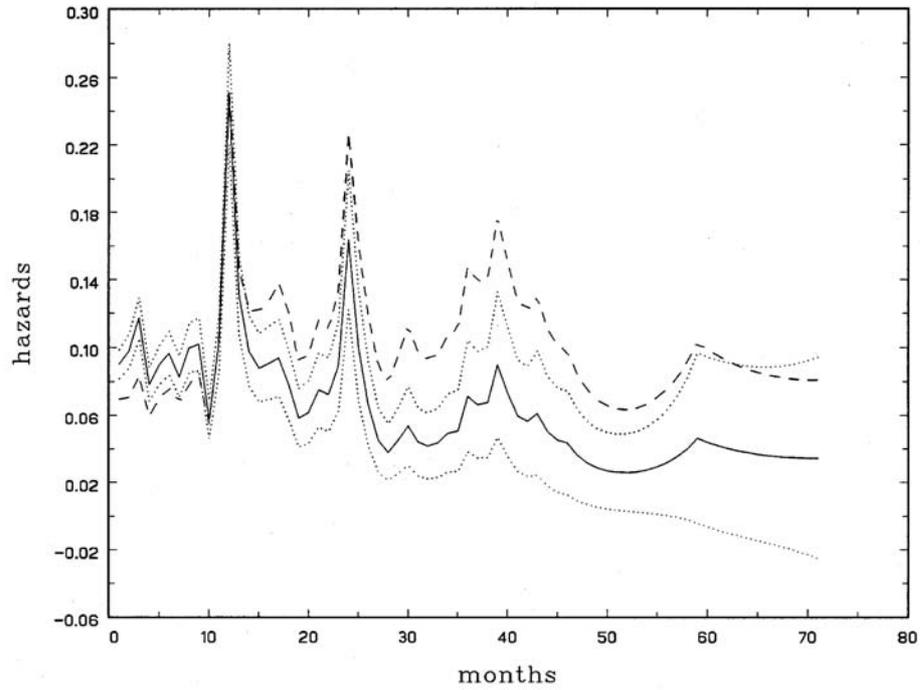


Figure 1 . Duration of unemployment, estimated rates of terminating unemployment for women in the reference group of age: Germans (——) with confidence band (·····) and foreigners (- - - - -).

Figure 1 shows the estimated hazard functions for German (——) and foreign (- - - -) females, with age between 31 and 40 years. Since foreign females with this age are coded by $x'_i=(0,0,0,0,0)$, their hazard function corresponds to the baseline effect, i.e. $\lambda(t|x_i) = \exp(\gamma_t)/\{1 + \exp(\gamma_t)\}$, while the hazard function for Germans is parametrized by $\exp(\gamma_t + \beta_{N,t})/\{1 + \exp(\gamma_t + \beta_{N,t})\}$. The smoothed estimate of the time-varying nationality effect $\beta_{N,t}$ is given in Figure 2. Both figures indicate that Germans have clearly better chances

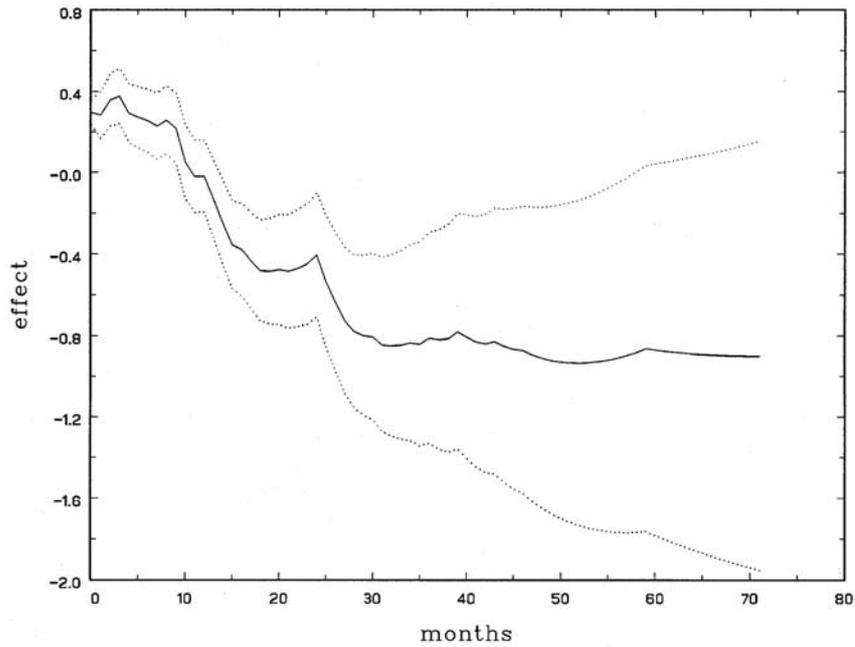


Figure 2 . Duration of unemployment, estimated nationality effects $\hat{\beta}_{N,t}$ (—) with confidence band (·····). Reference group: foreigner.

of leaving the state of unemployment during the first 10 months, but that foreigners have better chances after about 15 months. Of course, this may be partly explained by the fact that Germans with good chances on the labour market have already obtained a job earlier, while many of the remaining Germans are long-term unemployment persons. At a first look, the distinct and perhaps unexpected peaks at the end of the first and second year are rather striking. It will be seen that they can be better explained by a refined analysis with a competing risks model. Figure 3 gives clear evidence that the effect $\{\beta_{s,t}\}$ of sex is time-varying: In the

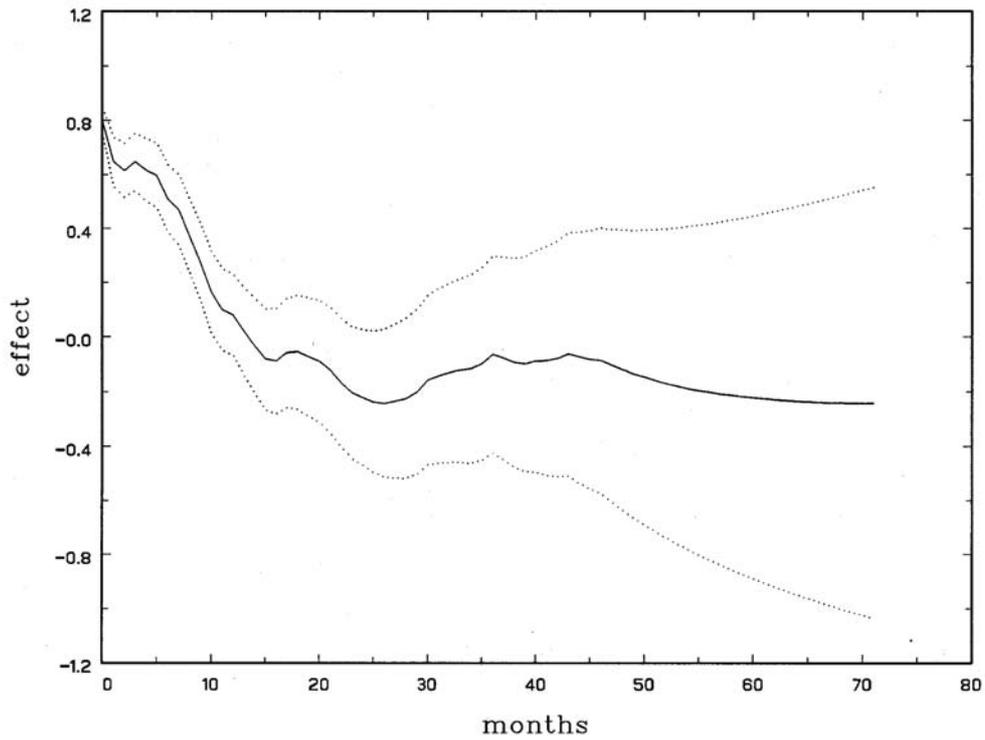


Figure 3 . Duration of unemployment, estimated effects of sex $\hat{\beta}_{s,t}$ (——) with confidence band (·····). Reference group: female.

first year, men have better chances. Later on chances are more or less the same, since the effect does not differ significantly from zero. This is also reflected in the smoothed hazard functions for men with German (——) or other (- - -) nationality in Figure 4. Again, the distinct

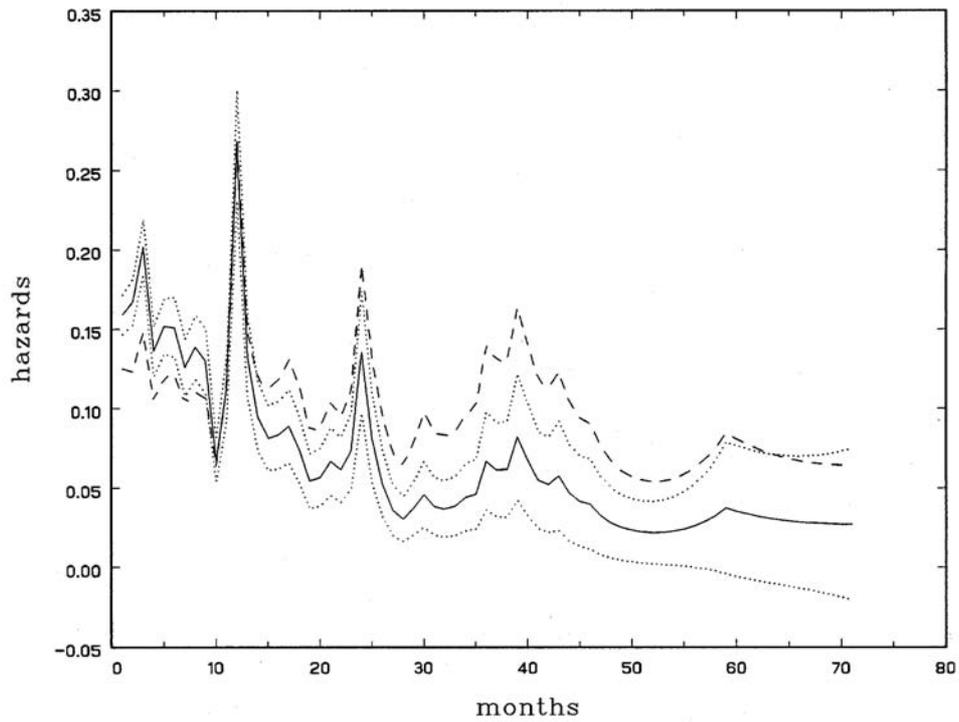


Figure 4 . Duration of unemployment, estimated rates of terminating unemployment for men in the reference group of age: Germans (——) with confidence band (·····) and foreigners (-----).

peaks are eye-catching. Finally the effect of age is displayed in Figure 5: As one may expect,

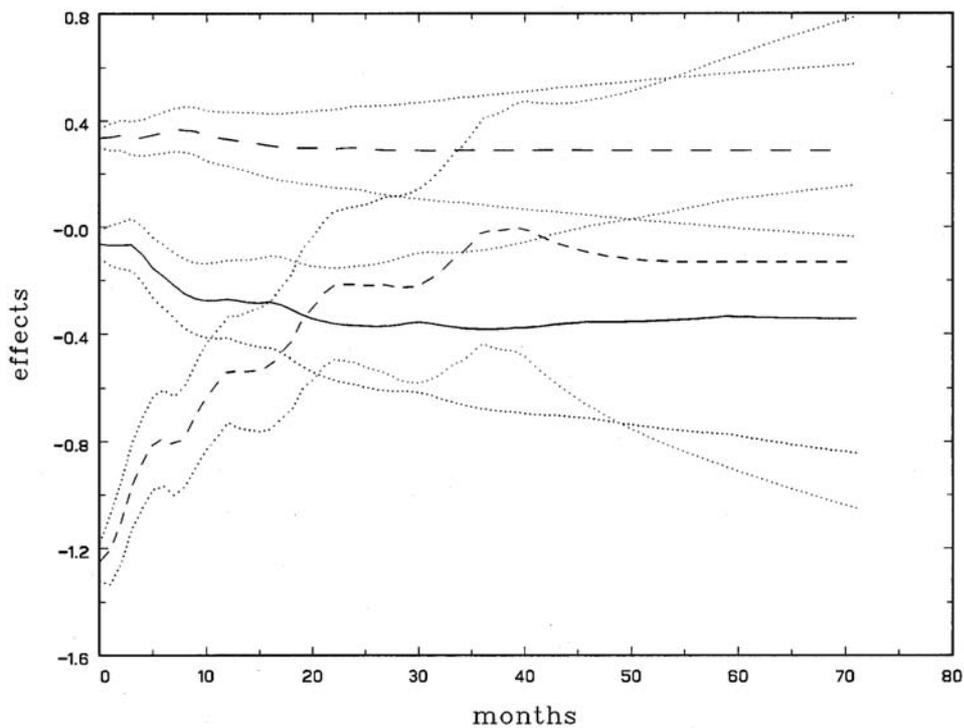


Figure 5 . Duration of unemployment, estimated effects of age group 1 $\hat{\beta}_{A1,t}$ (---), of age group 2 $\hat{\beta}_{A2,t}$ (—) and age group 3 $\hat{\beta}_{A3,t}$ (- - - - -) with confidence bands (·····). Reference group: "31 ≤ age ≤ 40 years".

younger persons (age ≤ 30 years) have constantly better chances of getting a job compared to the reference group (age from 31 to 40), while chances for persons between 41 and 50 are slightly decreasing and lower. Somewhat surprising is the time-varying effect for persons older than 50: At the beginning chances are bad, but later on they seem to be the same as for the reference group. Again, the following refined analysis will allow to interpret this effect.

We now distinguish between three causes for the ending of unemployment:

- (1) employment in a full-time job;
- (2) employment in a part-time job;
- (3) further causes like retraining or going to university, completing military or civil service, retiring, working as a housewife/houseman, and others.

To study cause-specific differences in hazard rates and covariate effects, we apply a multinomial dynamic logit model (2.4) with $m = 3$ categories defined by cause 1 (full-time job), 2 (part-time job) and 3 (others). Thus, the observation model is

$$h_r(t|x_i) = \frac{\exp(\eta_{itr})}{1 + \sum_{j=1}^3 \exp(\eta_{itj})}, \quad r = 1,2,3,$$

with cause-specific predictors $\eta_{itr} = \gamma_{tr} + x_i \beta'_{tr}$. Cause-specific baseline effects γ_{tr} and covariate effects $\beta'_{tr} = (\beta_{N,t}^{(r)}, \beta_{S,t}^{(r)}, \beta_{A1,t}^{(r)}, \beta_{A2,t}^{(r)}, \beta_{A3,t}^{(r)})$, $r = 1,2,3$, were again modelled by first order random walks. As before, we do not incorporate the history y_{t-1}^* of end of unemployment indicators, which are now cause-specific, into the model. Thus it is assumed that, for each of the causes, end of unemployment up to $t-1$ to a specific cause for persons with covariate vector x_i does not affect cause-specific future probabilities for end of unemployment due to other causes for persons with the same covariate constellation. If one thinks in terms of latent duration times for

each of the causes, this means that, for each of the causes, the other two are treated as noninformative censoring.

Smoothed hazard functions and time-varying effects are displayed in Figures 6 - 11.

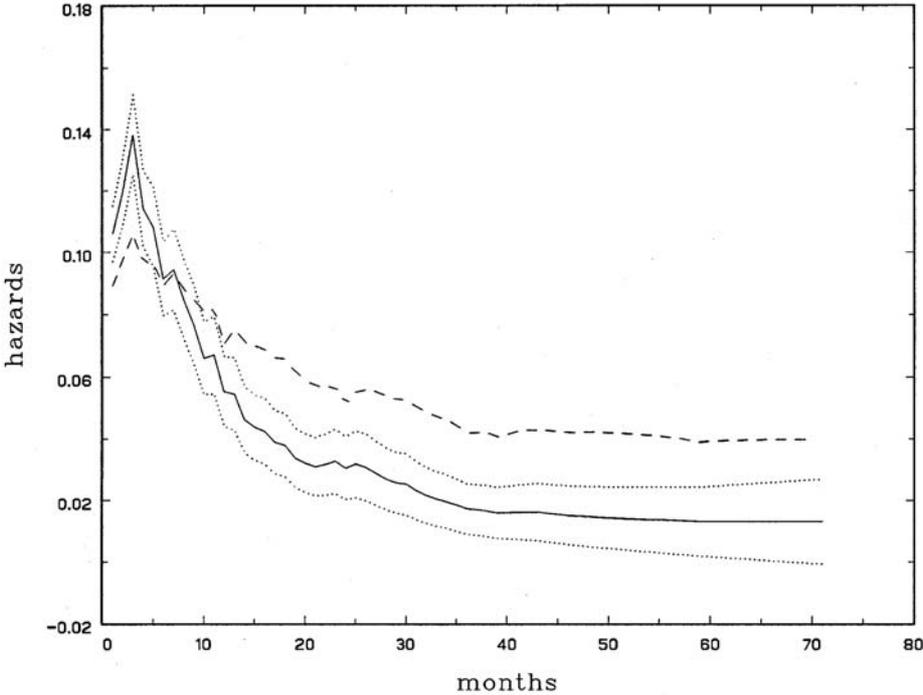


Figure 6 . Competing risks model, estimated rates for men in the reference group of age for the transition from unemployment to a full-time job: Germans (——) with confidence band (·····) and foreigners (- - - - -).

Figure 6 gives hazard functions for men with German (——) or other (- - - -) nationality, with age between 31 and 40, for the transition to a full-time job ($r = 1$). It shows the typical pattern of unemployment data with a short increase and slow decrease, without peaks as in Figures 2 and 4. However, the change of chances after about a year is still valid. In Figure 7 we

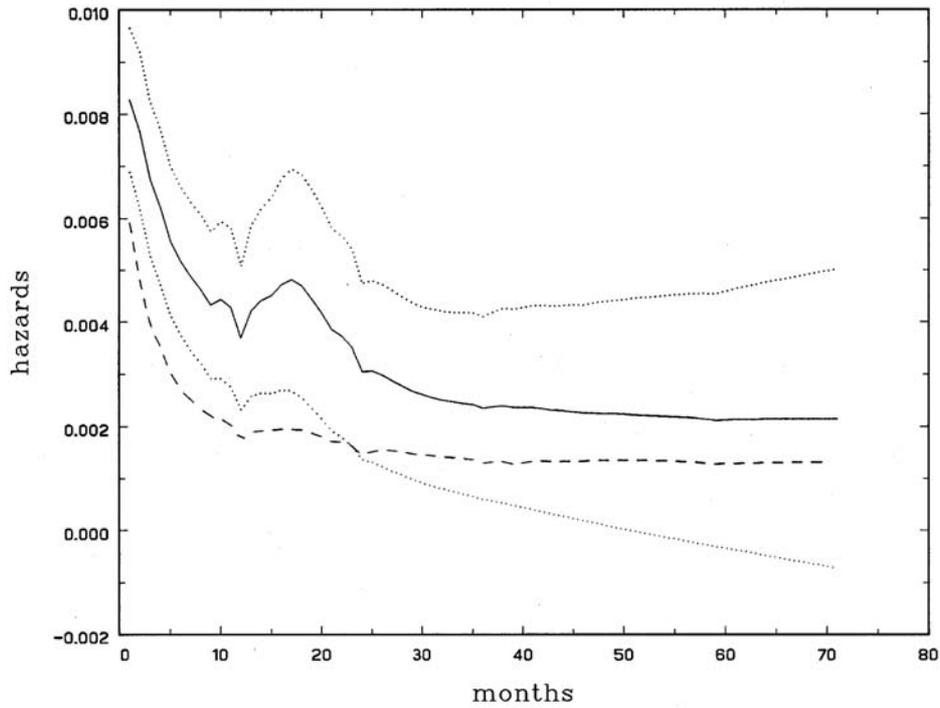


Figure 7 . Competing risks model, estimated rates for men in the reference group of age for the transition from unemployment to a part-time job: Germans (——) with confidence band (·····) and foreigners (- - - - -).

observe that slightly more German men in this age group get part-time jobs than men from other countries, but absolute values of hazards are quite low. The distinct peaks at the end of the first and second year in Figures 2 and 4 now reappear only in the hazard functions of both groups for transition to further causes (3) in Figure 8. There is no significant difference

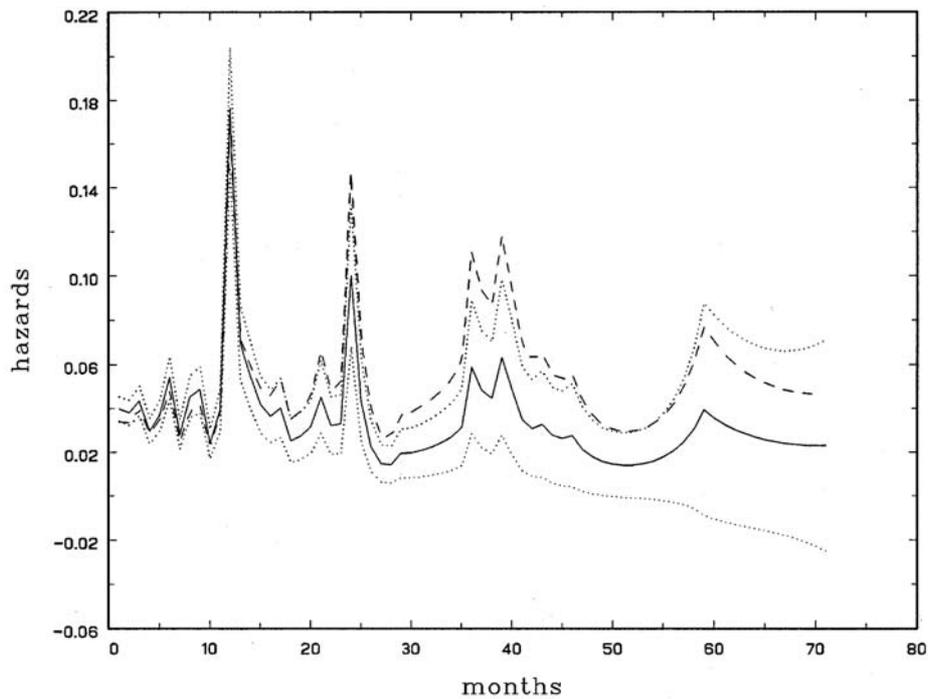


Figure 8 . Competing risks model, estimated rates for men in the reference group of age for the transition from unemployment to other causes: Germans (—) with confidence band (·····) and foreigners (- - - - -).

between Germans and foreigners in the reference group of age for the first year, whereas the hazard function for foreigners is slightly higher later on. A closer look at the data shows that transition to "retirement", "housewife/houseman" and "other reasons" are mainly responsible for the first peak, and transition to "other reasons" also for the second one. A possible explanation may be financial support whilst unemployment, which becomes significantly lower after one year of unemployment. A second reason may be due to the specific kind of questions on employment status in the GSOEP: Participants of the panel fill out questionnaires for each year and have to give answers on employment status retrospectively for each month of this year. If they do not like to name a certain month, they will perhaps simply name the beginning or end of a year as the time of leaving the status of unemployment. The smoothed estimates of the

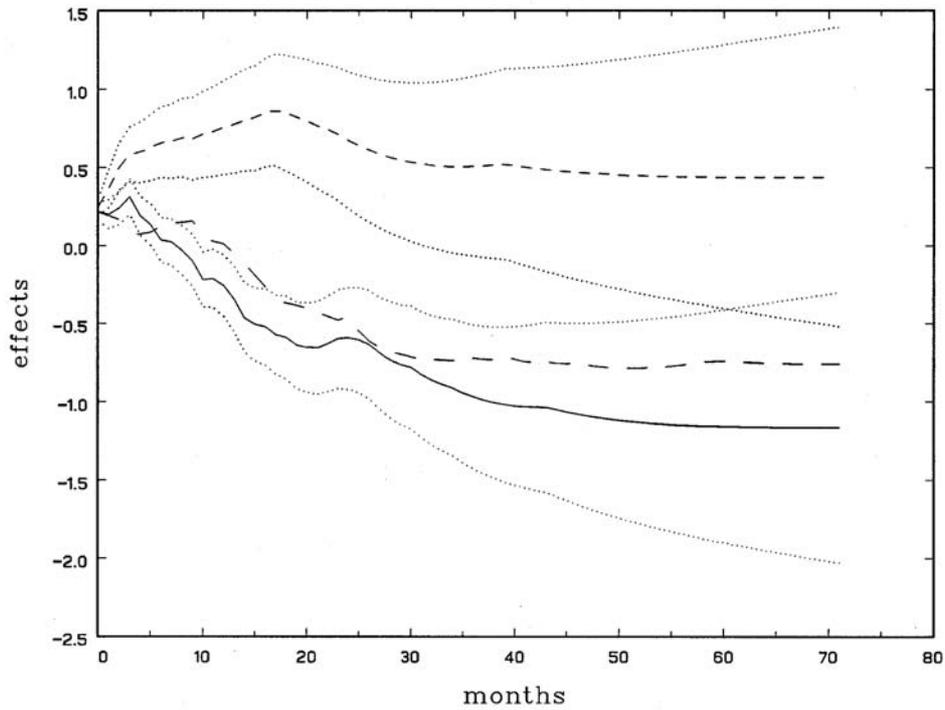


Figure 9 . Competing risks model, estimated nationality effects for the transition from unemployment to a full-time job $\hat{\beta}_{N,t}^{(1)}$ (——) with confidence band (·····), to a part-time job $\hat{\beta}_{N,t}^{(2)}$ (- - - - -) with confidence band (·····) and to other causes $\hat{\beta}_{N,t}^{(3)}$ (- · - · -).

effects of nationality $\hat{\beta}_{N,t}^{(r)}$, $r = 1,2,3$, in Figure 9 correspond to differences or similarities of hazard functions in Figures 6 - 8. For example, the effect $\hat{\beta}_{N,t}^{(2)}$ of German nationality on the chance of getting a part-time job is positive and increasing during the first 20 months and decreasing but still positive thereafter, compared to foreigners. Figure 10 shows the effect of

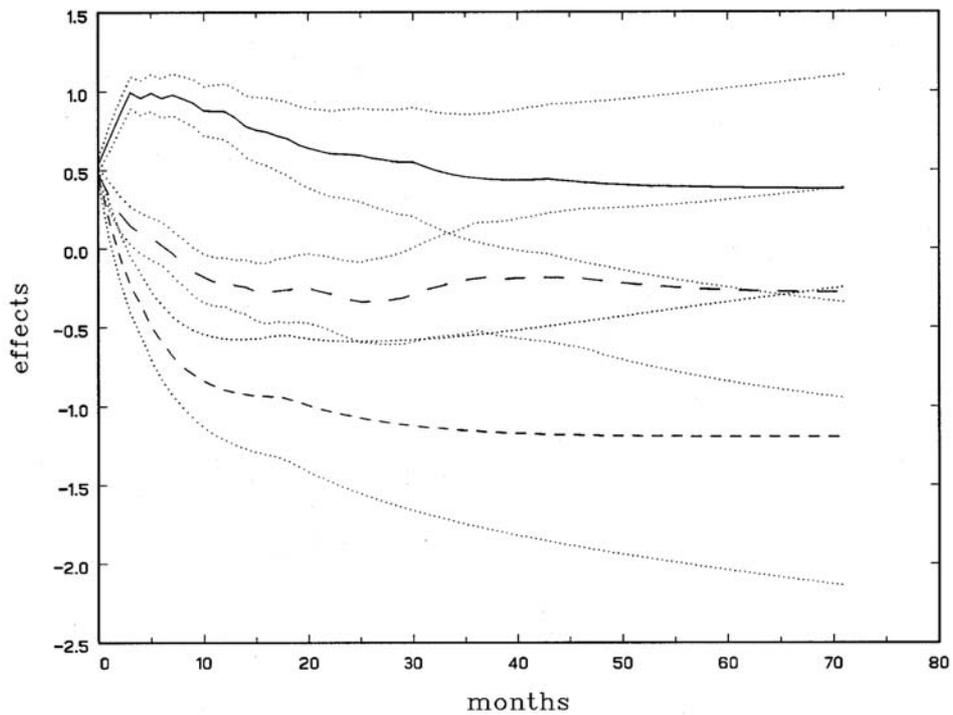


Figure 10 . Competing risks model, estimated effects of sex for the transition from unemployment to a full-time job $\hat{\beta}_{S,t}^{(1)}$ (——) with confidence band (·····), to a part-time job $\hat{\beta}_{S,t}^{(2)}$ (- - - - -) with confidence band (·····) and to other causes $\hat{\beta}_{S,t}^{(3)}$ (- · - · -).

being male compared to female. The effect $\hat{\beta}_{S,t}^{(1)}$, corresponding to transition to a full-time job, has a similar pattern as in Figure 3. However, the other effects show that, apart from the beginning, transition rates to part-time jobs or further causes are lower for men than for

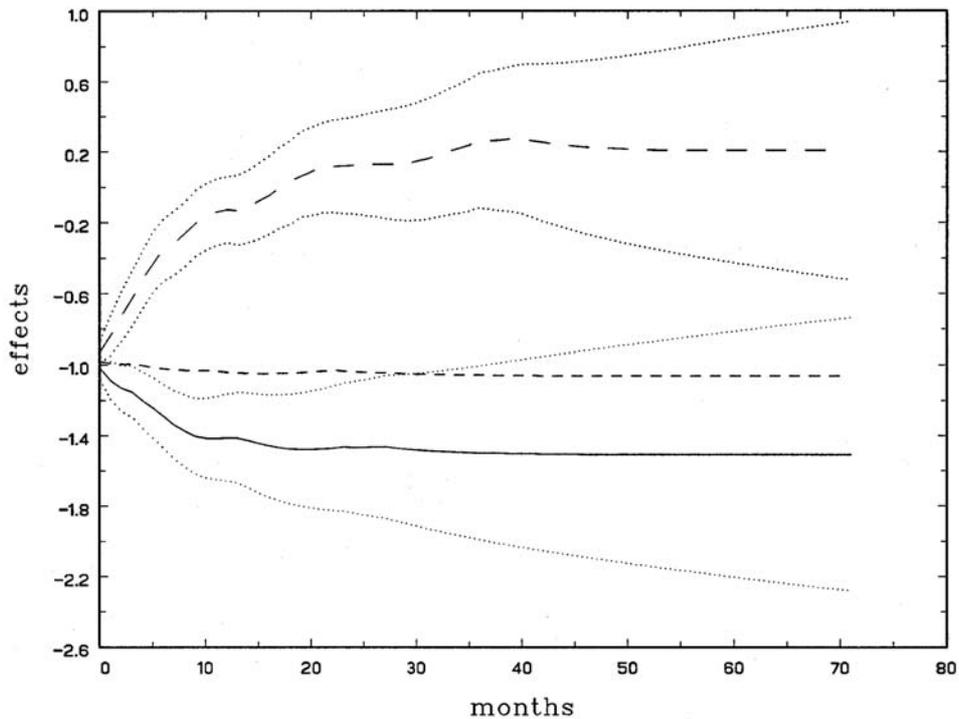


Figure 11 . Competing risks model, estimated effects of age group 3 for the transition from unemployment to a full-time job $\hat{\beta}_{A3,t}^{(1)}$ (—) with confidence band (·····), to a part-time job $\hat{\beta}_{A3,t}^{(2)}$ (- - - -) with confidence band (·····) and to other causes $\hat{\beta}_{A3,t}^{(3)}$ (- - - -).

women. Effects of age are quite similar as in Figure 5 for persons up to 50 years. Figure 11 shows that the time-varying effect of persons older than 50 years already seen in Figure 5 is mainly caused by transitions to "further causes". As one might expect, unemployed in this age group are willing to retire, i.e., to leave unemployment status, if they do not get work again within reasonable time. Compared to Figure 5, however, this time-varying effect is now less distinct. It appears that the stronger time-variation in Figure 5 is also caused by mixing the three effects in Figure 11 together.

Summing up, the results of this application demonstrate that dynamic duration and competing risks models are a useful tool for better interpretation and insight.

5. CONCLUSIONS

In this paper, we showed how smoothed estimates of time-varying hazards and effects can be obtained by a dynamic modelling approach. Another interesting question would be prediction, for example prediction of the survival function

$$S(t|\mathbf{x}) = \text{pr}(T \geq t|\mathbf{x}) = \prod_{s=1}^{t-1} \left\{ 1 - \sum_{r=1}^m \lambda_r(\mathbf{s}|\mathbf{x}) \right\}$$

of a new person, with covariate vector \mathbf{x} , becoming unemployed. A simple point estimate for $S(t,\mathbf{x})$ is obtained by plugging in the estimates $\hat{\gamma}_t$ and $\hat{\beta}_t$ for γ_t and β_t in the model for $\lambda_r(\mathbf{s}|\mathbf{x}) = \lambda_r(\mathbf{s}|\mathbf{x}; \gamma, \beta)$. The delta-method can be used to obtain approximate variances from the estimated covariance matrix of $\hat{\gamma}_t, \hat{\beta}_t$, which is provided by the (IWKFS)-algorithm. However, it is not possible to obtain the Bayesian predictive distribution, as e.g. in Gamerman (1991), since this requires the posterior distribution of γ_t, β_t , at least approximately. Two possibilities, both based on posterior mode smoother as an ingredient, are integration-based methods (Frühwirth-Schnatter, 1994) and simulation-based MCMC approaches (Gamerman, 1995), developed in a time series context. We will consider MCMC techniques in subsequent research.

Another extension concerns models for continuous time distributions. A flexible model, also allowing for simultaneous modelling and estimation of time-varying baseline and covariate effects, is the dynamic piecewise exponential model in Gamerman (1991), which is applied to unemployment duration in Gamerman and West (1987). It could be combined with our estimation approach by introducing the piecewise exponential distribution as observation model. Another possibility for further research is to consider continuous-time transition models, e.g. integrated Wiener process instead of random walks.

REFERENCES

- Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*, New Jersey: Prentice-Hall.
- Arjas, E., and Haara P. (1987), "A Logistic Regression Model for Hazard: Asymptotic Results," *Scandinavian Journal of Statistics*, 14, 1-18.
- Fahrmeir, L. (1992), "Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models," *Journal of the American Statistical Association*, 87, 501-509.
- (1994), "Dynamic Modelling and Penalized Likelihood Estimation for Discrete Time Survival Data," *Biometrika*, 81, 317-330.
- Fahrmeir, L., and Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer-Verlag.
- Fahrmeir, L., and Wagenpfeil, S. (1995), "Penalized Likelihood Estimation and Iterative Kalman Smoothing for Non-Gaussian Dynamic Regression Models," Discussion Paper, 5, Sonderforschungsbereich 386, Statistische Analyse diskreter Strukturen.
- Frühwirth-Schnatter, S. (1994), "Applied State Space Modelling of Non-Gaussian Time Series Using Integration-Based Kalman-Filtering," *Statistics and Computing*, 4, 259-269.
- Gamerman, D. (1991), "Dynamic Bayesian Models for Survival Data," *Applied Statistics*, 40, 63-79.
- (1992), "A Dynamic Approach to the Statistical Analysis of Point Processes," *Biometrika*, 79, 39-50.
- (1995), "Monte Carlo Markov Chains for Dynamic Generalized Linear Models," unpublished manuscript.
- Gamerman, D., and West, M. (1987), "An Application of Dynamic Survival Models in Unemployment Studies," *The Statistician*, 36, 269-274.
- Hamerle, A., and Tutz, G. (1989), *Diskrete Modelle zur Analyse von Verweildauern und Lebenszeiten*, New York: Campus.

- Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.
- Kohn, R., and Ansley, C. F. (1989), "A Fast Algorithm for Signal Extraction, Influence and Cross-validation in State Space Models," *Biometrika*, 76, 65-79.
- Kredler, Ch. (1986), "Behaviour of Third Order Terms in Quadratic Approximations of LR-statistics in Multivariate Generalized Linear Models," *The Annals of Statistics*, 14, 326-335.
- Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge, Cambridge university Press.
- Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, 541-554.
- Thompson, W. A., Jr. (1977), "On the Treatment of Grouped Observations in Life Studies," *Biometrics*, 33, 463-470.
- Wagenpfeil, S. (1995), "Hyperparameter Estimation in Exponential Family State Space Models," , Discussion Paper, 6, Sonderforschungsbereich 386, Statistische Analyse diskreter Strukturen.