Fahrmeir, Knorr-Held:

Dynamic discrete-time duration models. (REVISED)

Sonderforschungsbereich 386, Paper 14 (1996)

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

Dynamic discrete–time duration models

Ludwig Fahrmeir and Leonhard Knorr–Held
Universität München
Institut für Statistik
Ludwigstr. 33, 80539 München

Abstract

Discrete–time grouped duration data, with one or multiple types of terminating events, are often observed in social sciences or economics. In this paper we suggest and discuss dynamic models for flexible Bayesian nonparametric analysis of such data. These models allow simultaneous incorporation and estimation of baseline hazards and time–varying covariate effects, without imposing particular parametric forms. Methods for exploring the possibility of time–varying effects, as for example the impact of nationality or unemployment insurance benefits on the probability of re–employment, have recently gained increasing interest. Our modeling and estimation approach is fully Bayesian and makes use of Markov Chain Monte Carlo (MCMC) simulation techniques. A detailed analysis of unemployment duration data, with full–time job, part–time job and other causes as terminating events, illustrates our methods and shows how they can be used to obtain refined results and interpretations.

Key words: Bayesian inference; Discrete–time duration data; Markov chain Monte Carlo; Multiple types of terminating events; Time–varying regression parameters.

1 Introduction

Regression models for duration data are an important and widely used tool for statistical analysis of life or event histories. Many well–known models are based on the assumption that duration, the time until some event occurs, is a continuous variable, see, for example, Kalbfleisch and Prentice (1980), Blossfeld, Hamerle and Mayer (1989), Lancaster (1990), and Andersen et. al. (1993). However, in many applications, in particular in social sciences, time is often measured as a discrete variable. The applicability of continuous–time duration models to discrete–time data is limited to special cases, where the number of ties is relatively small.

Terminating event	Duration of unemployment (months)	Sex	Nationality	Age	Unemployment insurance benefits received
full–time job	16	male	non–German	21	yes
full–time job	3	female	non–German	23	yes
full–time job	4	male	non–German	25	yes
housewife	3	female	non–German	29	yes
(censored)	36	male	German	30	until month 6
househusband	6	male	German	32	yes
full time job	2	male	German	36	yes
housewife	11	female	German	30	yes
(censored)	4	female	non–German	22	yes
part–time job	4	female	German	27	yes

Table 1: Some typical observations from the unemployment duration data.

Table 1 shows a small sample from data on duration of unemployment, taken from the German socio–economic panel GSOEP. Duration of unemployment is discrete and measured in months. Also there are several alternative types of terminating events or destination states, and one may distinguish between full–time jobs, part–time jobs and other causes which end unemployment. Typical questions, that arise here, are: What is the influence of the covariates (e.g. sex) on the probability of leaving the state of unemployment? Does the effect of covariates change over duration time? How does the shape of hazard and survival functions look like in the presence of such time–varying effects? Is it necessary to distinguish between different types of terminating events?

Conventional duration models with time–constant parameters are not flexible enough to answer questions of this type. Instead, both baseline hazards and, at least some, covariate effects have to be considered as functions of time, γ_t and β_t, $t = 1, 2, \ldots, q$, say. Even for a moderate number q of intervals, unrestricted modeling and fitting of $\{\gamma_t\}$ and $\{\beta_t\}$ will cause severe problems: Due to the large number of parameters involved, this will often lead to non–existence and divergence of ML estimates. These difficulties increase in situations with many intervals - but not enough to apply models for continuous time - and with multiple terminating events. One may try to avoid such problems by a more parsimonious parameterization, using piecewise polynomials or other parametric forms for hazard functions or varying effects (Yamaguchi 1993). Multiphase models may also be considered (Portugal and Addison 1995). However, by imposing such parametric forms one may overlook unexpected patterns like peaks, bumps, or seasonal effects. In this situation, non– or semiparametric approaches are useful for detecting and exploring such unknown patterns. Appropriate parametric models may then be developed in a second step.

In this paper we propose dynamic discrete time duration models as a flexible nonparametric Bayesian approach, which makes simultaneous modeling and smoothing of hazard functions and covariate effects possible. Dynamic models are regarded as nonparametric, since no particular functional form is specified for the dependence of the parameters on time. Instead only some smoothness is imposed in form of a stochastic process prior. No approximations based on asymptotic normality assumptions have to be made for statistical inference, and estimation of unknown smoothness or variance parameters is automatically incorporated. Thus, the proposed nonparametric Bayesian framework is a promising alternative to more traditional nonparametric methods like spline smoothing (Hastie and Tibshirani 1993), local likelihood estimation (Tutz 1995), discrete kernel smoothing (Fahrmeir and Tutz 1994, Ch. 9) or approaches based on counting processes (Aalen 1989, 1993, Huffer and Mc Keague 1991).

The models are obtained by adopting dynamic or state space models for categorical data to discrete time duration data, similarly to Gamerman (1991) for a dynamic version of the piecewise exponential model, and Fahrmeir (1994) and Fahrmeir and Wagenpfeil (1996), see Section 2. In contrast to the latter papers, inference is fully Bayesian using Markov chain Monte Carlo (MCMC) methods, based on ideas and suggestions of Knorr–Held (1995, 1996) (Section 3). Other Bayesian nonparametric approaches based on MCMC simulation have recently been suggested by Arjas and Liu (1995) for continuous–time duration data and by Berzuini and Larizza (1996) for joint modeling of time series and failure time data. MCMC techniques allow flexible and sophisticated inference: pointwise and simultaneous credible regions for covariate effects, predictive survival functions and other characteristics can be calculated based on posterior samples. We illustrate our approach in Section 4 with a detailed study of unemployment duration data, taken from the German socio–economic panel GSOEP. In a first analysis, only the terminating event "end of unemployment" regardless of a specific cause, is considered. Based on this nonparametric analysis, we also fit parametric models and compare results. In a second refined analysis we distinguish between three terminating events: employment in a full–time job, employment in a part–time job, and other causes. The analysis shows that it is important to differentiate between alternative terminating events in order to obtain correct interpretations and conclusions. The results suggest that some effects of covariates, characterizing individuals, change through time, whereas the impact of unemployment benefits is more or less constant. This is in contrast to findings of Narendranathan and Stewart (1993) for data from the British labour market. Section 5 concludes with a discussion of other estimation approaches, extensions to multiple time scales, the role of unobserved heterogeneity and some other comments.

Formal definitions of dynamic models in Section 2 rely on basic concepts for discrete duration data. For easier reference, we give a short review. Let time be divided into intervals $[a_0 = 0, a_1), [a_1, a_2), \ldots, [a_{q-1}, a_q)$ and $[a_q, \infty)$. Without loss of generality we assume that a_q denotes the end of the observation period. Often the intervals $[a_0, a_1), \ldots, [a_{q-1}, a_q)$ are of equal length but this is not an essential requirement. Instead of a continuous duration time, the discrete duration time $T \in \{1, \ldots, q+1\}$ is observed, where $T = t$ denotes end of duration within the interval $[a_{t-1}, a_t)$. In addition to duration T, a sequence of possibly time–varying covariate vectors $x_t = (x_{t1}, \dots, x_{tp}), t = 1, 2, \dots$, is observed. Let $x(t) = (x_1, \dots, x_t)$ denote the history of covariates up to interval $[a_{t-1}, a_t)$. If there is only one type of terminating event the discrete hazard function is given by

$$\lambda(t|x(t)) = \mathrm{pr}(T = t|T \geq t, x(t)), \quad t = 1, \dots q,$$

which is the conditional probability of the end of duration in interval $[a_{t-1}, a_t)$, given that the interval is reached and the history of the covariates. The discrete survival function

$$S(t|x(t)) = \mathrm{pr}(T > t|x(t)) = \prod_{s=1}^{t} (1 - \lambda(s|x(s))) \tag{1}$$

is the probability of surviving the interval $[a_{t-1}, a_t)$. A common specification for the hazard function is a binary logit model of the form

$$\lambda(t|x(t)) = \frac{\exp(\gamma_t + z_t'\beta)}{1 + \exp(\gamma_t + z_t'\beta)}, \tag{2}$$

see e.g. Thompson (1977) or Arjas and Haara (1987). The parameter γ_t represents a time–varying baseline effect and the design vector z_t is some function of $x(t)$, often simply $z_t = x_t$. Finally β is the corresponding vector of fixed covariate effects. A slightly different specification is the grouped proportional hazards or Cox model $\lambda(t|x(t)) = 1 - \exp\left(-\exp(\gamma_t + z_t'\beta)\right)$,

see e.g. Kalbfleisch and Prentice (1980). This model can be derived by assuming a latent proportional hazards model for durations on a continuous time scale, but durations are only observed in terms of whole time–intervals, like weeks or months. If intervals are short compared to the observation period, the models are very similar, as has been shown by Thompson (1977). A detailed survey on discrete–time duration data can be found in Hamerle and Tutz (1989), a shorter introduction in Fahrmeir and Tutz (1994, Ch. 9).

For several, say m, alternative types of terminating events, causes or destinations, let $R \in \{1, \dots, m\}$ denote the distinct event. The basic quantities characterizing the duration process are now event–specific hazard functions

$$\lambda_r(t|x(t)) = \mathrm{pr}(T = t, R = r|T \geq t, x(t)), \tag{3}$$

$r = 1, \dots, m, t = 1, \dots, q$. Models for multicategorical responses can be used to model event–specific hazard functions. A common candidate for unordered events is the multinomial logit model (e.g. Allison 1982)

$$\lambda_r(t|x(t)) = \frac{\exp(\gamma_{tr} + z_t'\beta_r)}{1 + \sum_{j=1}^{m} \exp(\gamma_{tj} + z_t'\beta_j)}, \tag{4}$$

where γ_{tr} and β_r are event–specific baseline and covariate effects, respectively. A cause–specific generalization of the grouped Cox model is given e.g. in Fahrmeir and Tutz (1994, Ch. 9). Other discrete choice models like a probit or a nested multinomial logit model (Hill, Axinn and Thornton 1993) may also be considered.

Event–specific hazard functions need not necessarily correspond to latent duration times $T_1, \ldots, T_m$, one for each terminating event. The observed duration time can then be defined as $T = \min(T_1, \ldots, T_m)$ and the terminating event as $R = r$ if $T = T_r$, but this approach in general requires untestable assumptions on the independence of latent duration times. Therefore, we use event–specific hazard functions (3) as the basic characteristics for duration models, following Prentice et al. (1978), Kalbfleisch and Prentice (1980), and Lancaster (1990, p. 99).

2 Dynamic models for discrete–time duration data

For individual units $i = 1 \ldots, n$, let T_i denote duration times and U_i right–censoring times. Duration data with multiple terminating events are usually given by $(t_i, \delta_i, r_i, x_i(t_i))$, where $t_i = \min(T_i, U_i)$ is the observed discrete duration time, δ_i is the censoring indicator,

$$\delta_i = \begin{cases} 1, & T_i < U_i \\ 0, & T_i \geq U_i, \end{cases}$$

$r_i \in \{1, \ldots, m\}$ indicates the terminating event and $x_i(t_i) = \{x_{it}, t = 1, \ldots, t_i\}$ is the sequence of observed covariates. We rewrite the data in terms of stochastic processes: Let R_t denote the risk set, i.e. the set of units at risk in $[a_{t-1}, a_t)$. Censoring is assumed to occur at the end of the interval, so that the risk set R_t includes all individuals who are censored in $[a_{t-1}, a_t)$. We define event indicators $y_{it} \in \{0, 1, \ldots, m\}$, $i \in R_t$, $t = 1, \ldots, t_i$, by

$$y_{it} = \begin{cases} r, & \text{event of type } r \text{ occurs in } [a_{t-1}, a_t), \quad r = 1 \ldots, m \\ 0, & \text{no event occurs in } [a_{t-1}, a_t). \end{cases}$$

Then, from a dynamic point of view, duration can be interpreted as a stochastic process of multicategorical decisions between $y_{it} = 0$ and $y_{it} = r$, i.e., end of duration due to event $r \in \{1, \ldots, m\}$. Similarly, it is convenient to introduce censoring processes by

$$c_{it} = \begin{cases} 1, & U_i \geq a_t, \text{ i.e. } i \text{ not censored up to } [a_{t-1}, a_t) \\ 0, & U_i < a_t, \text{ i.e. } i \text{ censored in } [a_{t-1}, a_t) \text{ or earlier.} \end{cases}$$

We collect covariates, event and censoring indicators of time interval t, that is $[a_{t-1}, a_t)$, in the column vectors

$$x_t = (x_{it}, i \in R_t), \quad y_t = (y_{it}, i \in R_t), \quad c_t = (c_{it}, i \in R_t)$$

and denote histories up to t by

$$x_t^* = (x_1, \ldots, x_t), \quad y_t^* = (y_1, \ldots, y_t), \quad c_t^* = (c_1, \ldots, c_t).$$

Dynamic discrete duration models are defined hierarchically by an observation model, given the unknown baseline and covariate effects, a latent stochastic transition model for these possibly time–varying effects and priors for unknown hyperparameters of the transition model. The model specification is completed by several conditional independence assumptions.

Observation model

The duration process of each unit is viewed as a sequence of multicategorical decisions between remaining in the transient state $y_{it} = 0$, i.e. no event occurs, or leaving for one of the absorbing states $y_{it} = r$, $r = 1, \ldots, m$, i.e. end of duration at t due to terminating event of type r. Individual response probabilities for $y_{it} = r$ are modelled using categorical response models. For the special case of only one type of terminating event ($m = 1$), we assume for $i \in R_t$ that, conditional on parameters γ_t, β_t and the covariate x_{it}, response probabilities for $y_{it} = 1$ are in the form

$$\mathrm{pr}(y_{it} = 1 | x_{it}, \gamma_t, \beta_t) = h(\eta_{it}) \tag{5}$$

with linear predictor

$$\eta_{it} = \gamma_t + z_{it}' \beta_t \tag{6}$$

and link function $h : \mathbf{R} \mapsto (0,1)$, e.g. one of the common link functions for the logit or grouped Cox model. In (6), the design vector z_{it} is some appropriate function of the covariates x_{it}. The observation model can be extended to incorporate the history y_{t-1}^* of past event indicators into z_{it}, a suggestion made by Prentice et al. (1978). However, we do not make use of this possibility here. We assume that the censoring process is conditionally independent of y_{it}, given x_{it}, γ_t and β_t, so that z_{it} does not depend on c_t^*.

For multiple terminating events ($m > 1$) we assume for $r = 1, \ldots, m$

$$\mathrm{pr}(y_{it} = r | x_{it}, \gamma_t, \beta_t) = h_r(\eta_{it}), \tag{7}$$

with link function $h_r : \mathbf{R}^m \mapsto (0,1)$, and linear predictor vector $\eta_{it} = (\eta_{it1}, \ldots, \eta_{itm})$. For the multinomial logit model (4), we have

$$h_r(\eta_{it}) = \frac{\exp(\eta_{itr})}{1 + \sum_{j=1}^{m} \exp(\eta_{itj})} \tag{8}$$

with $\eta_{itr} = \gamma_{tr} + z_{it}' \beta_{tr}$. Other multicategorical response models can also be written in the general form (7). Again the design vector may be an appropriate function of covariates x_{it}, but not of c_t^*.

Transition model

Let α_t denote the state vector of unknown time–dependent parameters. Prior specifications for stochastic variation of $\{\alpha_t\}$ are in common linear Gaussian and Markovian form as for linear dynamic or state space models. The simplest model is a random walk of first–order $\alpha_t = \alpha_{t-1} + u_t$, $u_t \sim N(0, Q)$, here $\alpha_t = (\gamma_{t1}, \ldots, \gamma_{tm}, \beta'_{t1}, \ldots, \beta'_{tm})'$. An alternative approach is to take the process α_t to be the superposition of a first–order random walk and a local linear trend component with unknown time–changing slope τ_t, the local linear trend model (e.g. Fahrmeir and Tutz 1994, Ch. 8). An intermediate strategy is proposed in Berzuini and Larizza (1996), where the slope τ is assumed to be time–constant. Informative priors on τ can be used to incorporate prior beliefs that, say, a specific covariate effect is linear declining with time. Other interesting transition models are second–order random walks and seasonal models.

In general, we admit a multivariate Gaussian autoregressive model of order k for α_t, $t \geq k$:

$$\alpha_t = \sum_{l=1}^{k} F_l \alpha_{t-l} + u_t, \quad u_t \sim N(0, Q_t). \tag{9}$$

The error variables u_t are assumed to be mutually independent and independent of initial values α_t, for which diffuse priors $\alpha_t \propto$ const, $t = 1, \ldots, k$, are assumed. The matrices $F_1, \ldots, F_k$ are known. If time intervals $[a_{t-1}, a_t)$ are of the same length, we set $Q_t = Q$, otherwise $Q_t = h_t Q$, where h_t is the length of $[a_{t-1}, a_t)$. Usually Q is unknown and is considered as a hyperparameter. In a full Bayesian setting, a prior specification for Q completes the transition model. Products of inverse gamma or inverted Wishart distributions are the usual choice.

For full Bayesian inference, the joint distribution of $y = (y_1, \ldots, y_q)$, $x = (x_1, \ldots, x_q)$, $c = (c_1, \ldots, c_q)$, $\alpha = (\alpha_1, \ldots, \alpha_q)$, where q is the number of intervals, and Q has to be completely defined. This is achieved by adding a number of conditional independence assumptions. To see what assumptions are useful and how they can be interpreted, we recursively factorize the joint distribution. Let

$$L_t = \mathrm{p}(y_t^*, x_t^*, c_t^*, \alpha_t^*, Q), \qquad t = 1, \ldots, q,$$

denote the joint distribution up to the interval $[a_{t-1}, a_t)$. By repeated conditioning, we get the factorization

$$L_t = L_{t-1} \mathrm{p}(y_t|\cdot) \mathrm{p}(\alpha_t|\cdot) \mathrm{p}(x_t, c_t|\cdot)$$

with

$$
\begin{aligned}
\mathrm{p}(y_t|\cdot) &= \mathrm{p}(y_t|y_{t-1}^*, x_t^*, c_t^*, \alpha_t^*, Q), \\
\mathrm{p}(\alpha_t|\cdot) &= \mathrm{p}(\alpha_t|\alpha_{t-1}^*, y_{t-1}^*, x_t^*, c_t^*, Q), \\
\text{and } \mathrm{p}(x_t, c_t|\cdot) &= \mathrm{p}(x_t, c_t|y_{t-1}^*, x_{t-1}^*, c_{t-1}^*, \alpha_{t-1}^*, Q).
\end{aligned}
$$

We now make the following conditional independence assumptions:

A1 Conditional on x_{it} and α_t, individual event indicators y_{it} are independent of α_{t-1}^* and Q, i.e.

$$\mathrm{p}(y_{it}|y_{t-1}^*, x_t^*, c_t^*, \alpha_t^*, Q) = \mathrm{p}(y_{it}|x_{it}, \alpha_t).$$

A2 Given y_{t-1}^*, x_t^*, c_t^*, α_t^* and Q, individual event indicators y_{it}, $i \in R_t$ are conditionally independent, i.e.

$$\mathrm{p}(y_t|y_{t-1}^*, x_t^*, c_t^*, \alpha_t^*, Q) = \prod_{i \in R_t} \mathrm{p}(y_{it}|y_{t-1}^*, x_t^*, c_t^*, \alpha_t^*, Q).$$

A3 The sequence α_t is Markovian of order k, i.e.

$$\mathrm{p}(\alpha_t|\alpha_{t-1}^*, y_{t-1}^*, x_t^*, c_t^*, Q) = \begin{cases} \mathrm{p}(\alpha_t|\alpha_{t-1}, \ldots, \alpha_{t-k}, Q) & t > k \\ \mathrm{p}(\alpha_t) & t = 1, \ldots, k. \end{cases}$$

A4 Given y_{t-1}^*, x_{t-1}^*, c_{t-1}^*, covariates x_t and censoring indicator c_t are independent of α_{t-1}^* and Q.

A5 Initial values $\alpha_1, \ldots, \alpha_k$, x_1, c_1 and Q are mutually independent.

Assumption (**A1**), which is implicitly assumed in the observation model, is common for dynamic or state space modeling. It says that conditional information of α_t^* on y_t is already contained in α_t, and is usually not stated for fixed parameters. Since only individuals i in the risk set R_t contribute likelihood information in time period t, i.e. $c_{it} = 1$ if $i \in R_t$, c_t^* can be omitted on the right hand side of (**A1**). Note that the covariates x_{it} may contain information on covariate values of other individuals or from the past. As stated above, we do not include the history y_{t-1}^* of failure indicators in form of covariates. The conditional independence assumption (**A2**) is weaker than usual unconditional independence assumptions among units, since it allows for interaction via common history, and it is likely to hold if a common cause for failures is incorporated in the covariate process. For fixed parameters (**A2**) corresponds to Assumption 2 of Arjas and Haara (1987). Assumption (**A3**) is already implied by the transition model (13). Assumption (**A4**) corresponds to Assumption 1 of Arjas and Haara (1987). It will generally hold for noninformative censoring and external or time independent covariates. It may not hold for internal covariates. Independence of initial values $\alpha_1, \ldots, \alpha_k$ in (**A5**) has already been stated in the transition model and is supplemented by the additional independence assumption on x_1, c_1 and Q.

Summarizing (**A1**) and (**A2**), we get

$$\mathrm{p}(y_t | y_{t-1}^*, x_t^*, c_t^*, \alpha_t^*, Q) = \prod_{i \in R_t} \mathrm{p}(y_{it} | x_{it}, \alpha_t).$$

Under assumptions (**A1**) – (**A5**) the joint distribution of y, x, c, α and Q is now proportional to a product of individual conditional likelihood contributions, defined by the observation model, a smoothness prior for α as a product of transition densities by (13), and the prior for Q:

$$\begin{aligned} \mathrm{p}(y, x, c, \alpha, Q) \quad \propto \quad & \left\{ \prod_t \prod_{i \in R_t} \mathrm{p}(y_{it} | x_{it}, \alpha_t) \right\} \times \\ & \left\{ \prod_{t > k} \mathrm{p}(\alpha_t | \alpha_{t-1}, \ldots, \alpha_{t-k}, Q) \right\} \times \mathrm{p}(Q). \end{aligned} \tag{10}$$

A graphical representation of our model is shown in Figure 1. Individual densities in the first factor are given by the observation model (5), implying they are independent of the right–censoring mechanism c_t^*, and transition densities in the second factor are given by

$$\mathrm{p}(\alpha_t | \alpha_{t-1}, \ldots, \alpha_{t-k}, Q) \sim N(\sum_{l=1}^{k} F_l \alpha_{t-l}, Q). \tag{11}$$

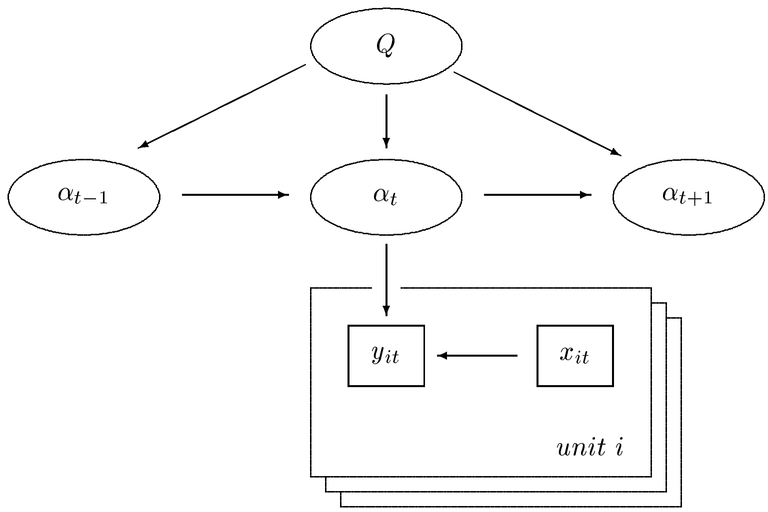

Figure 1: Directed graphical representation of a dynamic model with lag $k = 1$.

3 Estimating hazard functions and covariate effects by MCMC simulation

Smoothing time–varying parameters, i.e. estimation of the sequence $\alpha = \{\alpha_t\}$ given the data y, x, c is of prime interest. Full Bayesian inference will be based on the posteriors $p(\alpha|y, x, c)$ or $p(\alpha_t|y, x, c)$, which are proportional to the right hand side of (10). Since the normalizing factor has rather complex structure, direct approaches using numerical integration or ordinary static Monte Carlo methods are computationally infeasible.

We suggest a Markov chain Monte Carlo (MCMC) sampling scheme, that allows to draw samples from posteriors of time–varying parameters, hazard functions and similar quantities of interest, thus making full Bayesian inference possible. We start this section with a short review, since some of the readers might not be so familiar with MCMC. The reader is also referred to the tutorial expositions of Casella and George (1992), Chib and Greenberg (1995) and Gilks, Richardson and Spiegelhalter (1996). A more theoretical study of MCMC techniques can be found in Tierney (1994).

3.1 Basic ideas of MCMC

MCMC techniques have revolutionized general Bayesian inference in the last few years. Bayesian inference starts with a prior distribution $\mathrm{p}(x)$ for an unknown parameter vector x. In our context, the unknown parameters are $\alpha_1, \ldots, \alpha_q$ and Q and the corresponding prior distribution is

$$\left\{ \prod_{t>k} \mathrm{p}(\alpha_t | \alpha_{t-1}, \ldots, \alpha_{t-k}, Q) \right\} \times \mathrm{p}(Q).$$

Having observed data D, Bayes's theorem tells us, that the posterior distribution, conditioning on D, is given by

$$\mathrm{p}(x|D) = \frac{\mathrm{p}(D|x)\mathrm{p}(x)}{\int \mathrm{p}(D|x)\mathrm{p}(x)dx}. \tag{12}$$

Here $\mathrm{p}(D|x)$ is the likelihood, in our case equal to

$$\left\{ \prod_t \prod_{i \in R_t} \mathrm{p}(y_{it} | x_{it}, \alpha_t) \right\}.$$

The right hand side of equation (10) corresponds to the nominator $\mathrm{p}(D, x) = \mathrm{p}(D|x)\mathrm{p}(x)$ in (12).

The posterior distribution contains all the information Bayesian inference is based on. Typically, summary characteristics of the posterior, such as the posterior mean

$$\mathrm{E}(x|D) = \frac{\int x \; \mathrm{p}(D|x)\mathrm{p}(x)dx}{\int \mathrm{p}(D|x)\mathrm{p}(x)dx}$$

are of primary interest. However, computation of such expectations involves integrations, which can be very hard to solve, especially if x is of high dimension. Therefore, classical Bayesian inference was restricted to rather simple models, where analytic computation of characteristics of the posterior distribution was possible. Accurate approximations by numerical techniques are available only for problems, where the dimension of the parametervector is not greater than, say, 10 or 20. However, for a lot of applied problems, the posterior is analytically and numerically intractable. Monte Carlo methods circumvent the integration problem by generating samples from the posterior distribution. However, ordinary Monte Carlo methods, such as importance sampling, are often computationally infeasible for complex, highly structured models. Here MCMC methods are more appropriate.

In this subsection let $\mathrm{p}(x)$ be the posterior distribution of a random vector x, suppressing the conditioning on the data D. The basic idea of MCMC is to generate a sample $x^{(k)}, k = 1, 2, \ldots$ by a Markov transition function $Q(x^{(k)} \rightarrow x^{(k+1)})$ such that $\mathrm{p}(x)$ is the stationary distribution of the Markov chain X. Thus, after a sufficiently long 'burn–in phase' of length m, the generated states $x^{(k)}, k = m + 1, \ldots, n$ are dependent samples from the posterior. For example, the posterior mean can now be estimated by the arithmetic average

$$\frac{1}{n-m} \sum_{k=m+1}^{n} x^{(k)}.$$

Other quantities of interest can be also be estimated by the appropriate empirical versions.

For construction of such a Markov chain it is necessary to find a suitable transition function $Q(x^{(k)} \rightarrow x^{(k+1)})$, such that the posterior distribution $\mathrm{p}(x)$ is the stationary distribution of X. There are surprisingly many different choices of Q for a given distribution $\mathrm{p}(x)$, but most of them, including the Gibbs sampler, are special cases of the Hastings (1970) algorithm. Most methods split up x into components $x_1, \ldots, x_T, \ldots, x_H$ of possibly differing dimension. In our context, these components could be chosen as $\alpha_1, \ldots, \alpha_q$ and Q, leading to a so–called single move updating scheme. These components are updated one by one using the Hastings algorithm. The posterior distribution $\mathrm{p}(x)$, typically high dimensional and rather complicated, is not needed; only so called *full conditional distributions* enter in the Hastings algorithm. A full conditional distribution, short *full conditional*, is the distribution of one component, conditioning on all the remaining components, such as $\mathrm{p}(x_T|x_1, \ldots, x_{T-1}, x_{T+1}, \ldots, x_H)$. Besag (1974) showed, that $\mathrm{p}(x)$ is uniquely determined by the set of its full conditional distribution. This gives an intuitive justification for the fact, that only full conditional distributions and not the posterior itself are needed for MCMC simulation. In hierarchical models, defined by conditional independence assumptions, these full conditionals often have a much simpler structure than the posterior itself. This provides an important computational advantage.

The Gibbs sampling algorithm, probably the most prominent member of MCMC algorithms, iteratively updates all components by samples from their full conditionals. Markov chain theory shows that under very general conditions the so generated sequence of random numbers converges to the posterior. However, often these full conditionals are themselves still quite complex, so generation of the required random numbers might be a difficult task. Relief lies in the fact that it is not necessary to sample from the full conditionals; A member of the much more general class of Hastings algorithms can be used to update the full conditionals. Such a Hastings step is typically easier to implement and often makes a MCMC algorithm more efficient in terms of CPU time. A Hastings step proposes a new value for a given component and accepts it with a certain probability. A Gibbs step (i.e. a sample from a full conditional) turns out to be a special case where the proposal is always accepted.

Let $\mathrm{p}(x_T|x_{-T})$ be the full conditional of a component x_T of x, given the rest of the components, denoted by x_{-T}. To update $x_T = x_T^{(k)}$ in iteration step k, it is sufficient to generate a proposal x'_T from an arbitrarily chosen transition kernel $\mathrm{P}(x_T \rightarrow x'_T; x_{-T})$ and accept the generated proposal with probability

$$\delta = \min\left\{1, \frac{\mathrm{p}(x'_T|x_{-T})\mathrm{P}(x'_T \rightarrow x_T; x_{-T})}{\mathrm{p}(x_T|x_{-T})\mathrm{P}(x_T \rightarrow x'_T; x_{-T})}\right\},$$

otherwise leave x_T unchanged. This is the Hastings algorithm used for updating full conditionals. Only a ratio of the full conditional of x_T enters in δ, so $\mathrm{p}(x_T|x_{-T})$ need to be known just up to scale and need not to be normalized, a very convenient fact for implementation. Note that both the current state x_T and the proposed new state x'_T as well as the current states of the other components x_{-T} affect δ.

Gibbs sampling corresponds to the specific choice

$$\mathrm{P}(x_T \rightarrow x'_T; x_{-T}) = \mathrm{p}(x'_T|x_{-T}),$$

so that δ becomes 1 and therefore all proposals are accepted. Here the current state of x_T does not affect the new one x'_T.

There is a great flexibility in the choice of the transition kernel P. Common choices are random walk Metropolis proposals and (conditional) independence proposals (Tierney 1994).

Random walk Metropolis proposals are generated from a distribution, that is symmetric about the current value x_T. Often used are Gaussian or Rectangular distributions. In contrast, conditional independence proposals do not depend on the current state of x_T; they may however depend on the current values of x_{-T}. As we have seen above, the Gibbs sampling kernel is a specific conditional independence proposal. However, it is crucial that for a chosen P, the acceptance probability δ not be too small (in average) and that both convergence and mixing behavior of the whole simulated Markov chain be satisfactory.

Somewhat surprising is the fact that one is allowed to use hybrid procedures, that is, use different versions of Hastings proposals for updating different components of x. One strategy is to sample from the full conditionals, that is a "Gibbs step", as long as this is easy and fast. If not, a specific Hastings step with a simple proposal distribution mostly works faster in CPU time. As long as all components are updated in a deterministic or even random order (which may ensure better mixing of the chain), the chain converges to the posterior.

3.2 MCMC simulation in dynamic discrete time duration models

In this subsection we propose a hybrid MCMC procedure for simulating the (unnormalized) posterior (10). Time–varying parameters α_t, $t = 1, \ldots, q$, are updated using specific conditional independence proposals, while a Gibbs step is used for updating Q. Consider the full conditional

$$\mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q, y, x, c) \propto \prod_{i\in R_t} \mathrm{p}(y_{it}|x_{it}, \alpha_t) \times \mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q). \tag{13}$$

While the first factor corresponds to the observation model at time t, the second reflects the dependence of underlying parameters through the transition model and does not depend on the data y, x and c.

This second factor, the conditional distribution $\mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q)$, can be derived from (9). It is Gaussian, with density function $\varphi(\alpha_t; \mu_t, \Sigma_t)$, where the mean μ_t and covariance matrix Σ_t depend on the current values of Q and of neighboring parameters $\alpha_{s\neq t}$. Different transition models result in different formulae for μ_t and Σ_t. For example, a random walk of first–order $\alpha_t = \alpha_{t-1} + u_t$, $u_t \sim N(0, Q)$, has conditional distribution

$$N(\mu_t, \Sigma_t) = \begin{cases} N(\alpha_{t+1}, Q) & (t = 1) \\ N(\frac{1}{2}\alpha_{t-1} + \frac{1}{2}\alpha_{t+1}, \frac{1}{2}Q) & (t = 2, \ldots, q-1) \\ N(\alpha_{t-1}, Q) & (t = q) \end{cases} . \tag{14}$$

We add a short derivation of this result for $t = 2, \ldots, q-1$. The first–order random walk prior on α can be written as

$$\mathrm{p}(\alpha|Q) \propto \exp\left(-\frac{1}{2}\sum_{t=2}^{q}(\alpha_t - \alpha_{t-1})' Q^{-1} (\alpha_t - \alpha_{t-1})\right).$$

Since $\mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q) \propto \mathrm{p}(\alpha|Q)$, it follows that

$$\mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q) \propto \exp\left(-\frac{1}{2}\left((\alpha_t - \alpha_{t-1})' Q^{-1}(\alpha_t - \alpha_{t-1}) + (\alpha_{t+1} - \alpha_t)' Q^{-1}(\alpha_{t+1} - \alpha_t)\right)\right),$$

which gives the desired result. Note that (14) has an independent appealing interpretation as a stochastic interpolation rule (Besag and Kooperberg 1995).

We use a specific conditional independence proposal, namely a sample from the conditional distribution $\mathrm{p}(\alpha_t|\alpha_{s\neq t}, Q)$, to update α_t via a Hastings step. The acceptance probability simplifies in this case to

$$\delta = \min\left\{1, \frac{\mathrm{p}(y_t|\alpha_t')}{\mathrm{p}(y_t|\alpha_t)}\right\},$$

with

$$\mathrm{p}(y_t|\alpha_t) := \prod_{i\in R_t} \mathrm{p}(y_{it}|x_{it}, \alpha_t)$$

as the conditional likelihood of objects under risk in interval t, defined by the observation model. Such a proposal has a natural interpretation due to the hierarchical structure of the model: α_t' is drawn independently of the observation model and just reflects the specific autoregressive prior specification. It is therefore called a conditional prior proposal (Knorr–Held 1996). If it produces improvement in the likelihood at time t, it will always be accepted, if not the acceptance probability is equal to the likelihood ratio. This algorithm shows good performance for duration data with an acceptance rate ranging from 0.3 to 0.9. We also experienced with a slightly different MCMC sampling scheme, where blocks $\alpha_a, \ldots, \alpha_b$ are updated simulataneously rather then updating each α_t one at a time. Such a blocking strategy often improves mixing and convergence considerably. Conditional prior proposals can be generalized to this case conveniently.

Sampling from the full conditional

$$\mathrm{p}(Q|\alpha, y, x, c) \sim \mathrm{p}(Q|\alpha)$$

is straightforward for conjugate priors like inverse gamma or inverted Wishart distributions. If Q is assumed to be diagonal, an inverse gamma prior $Q_{jj} \sim IG(a, b)$ for the j–th diagonal entry in Q is computational convenient, since the resulting full conditional is still inverse gamma with parameters $a + (q - k)/2$ and $b + \sum u_{tj}^2/2$. Note the transformation from α to u_t, $t = k+1, \ldots, q$ via the transition model (9). The inverse gamma distribution has density

$$\mathrm{p}(Q_{jj}) \propto Q_{jj}^{-a-1} \exp(-b/Q_{jj})$$

and has a unique mode at $b/(a+1)$. In all our examples we start with the values $a = 1$ and $b = 0.005$, so that $\mathrm{p}(Q_{jj})$ is highly dispersed but still proper. This choice reflects sufficient prior ignorance about Q but avoids problems arising with improper priors, see Raftery and Banfield (1991) for a more detailed discussion. We then add a sensitivity analysis and rerun the algorithm with different choices for b, such as 0.05 or 0.0005. The parameter b determines, how close to zero the variances are allowed to be a priori. Note that the inverse gamma distribution has no expectation for $a = 1$, so our prior guess is rather diffuse for every value of b.

It is very important to carefully check convergence and mixing behavior of any MCMC algorithm. Theoretical considerations are typically limited to rather simple models; therefore empirical output analysis is more practical. This is still an active research area, the reader is referred to Raftery and Lewis (1996), Gelman (1996), Cowles and Carlin (1996) and the relevant parts of Gilks, Richardson and Spiegelhalter (1996). We always look at several plots such as time series plots of the sampled values and calculate routinely autocorrelation functions for every parameter. Figure 2 shows the time series plot of a specific parameter of our first analysis ($m = 1$, Section 4). Shown are the stored values of the 100th parameter in our model, the effect of nationality at time $t = 28$. Low values of the autocorrelation function indicate good mixing. Plots for other parameters look quite similar.

After convergence, the simulated random numbers are samples from the marginal distributions $\mathrm{p}(\alpha|y, x, c)$ and $\mathrm{p}(Q|y, x, c)$ and are used to estimate characteristics of the posterior distribution. Note that for a given covariate sequence $x_i(t)$ of a specific unit i, samples from its hazard function are calculated by plugging in the samples from $\mathrm{p}(\alpha|y, x, c)$ in (5) or (7). Even samples from the survivor function can be obtained by using the samples from

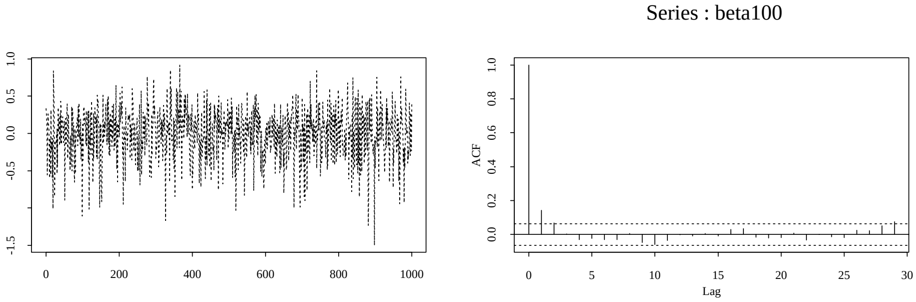

Figure 2: Time series plot and estimated autocorrelation function of a selected parameter

the hazard function in the dynamic version of (1). Furthermore it is possible to construct simultaneous credible regions for covariate effects, hazard or survival functions by using the method described in Besag et. al. (1995).

4 Applications to duration of unemployment

We analyze data on duration of unemployment of 1416 persons, older than 16 years and living in West Germany, which are observed from January 1983 until December 1993 in the German socio–economic panel GSOEP. Only persons with single spells of unemployment are considered. Duration of unemployment is measured in months. 296 observations are censored. Only a small fraction of persons are unemployed for more than three years, so that there is very little information on such long–term unemployment. Therefore only durations up to 36 months are considered and longer durations are considered as censored. In total, 49 from the 54 persons, still unemployed in month 36, are censored. Based on previous analysis with time–constant effects (Fahrmeir and Tutz 1994, Ch. 9) we include the covariates sex, age and nationality, coded as follows:

Sex S : $S = 1$ for males , $S = 0$ for females;

Nationality N: $N = 1$ for German , $N = 0$ for foreigner;

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 "$41 \leq$ age ≤ 50 years", 0 else;

$A3 = 1$ for "age ≥ 51 years", 0 else;

with reference category "$31 \leq$ age $\leq$ 40 years" coded by $(A1, A2, A3) = (0, 0, 0)$. The observed frequency counts for these covariates are 56 %, 64 %, 50 %, 15 % and 18 % for S, N, $A1$, $A2$ and $A3$, respectively.

Most often covariates are expected to have much the same impact over the course of unemployment. We let the data decide whether this is really so and admit that the effects of these covariates may vary over time. In particular, we are able to check if unemployment benefits have effects that vary or erode over time. Results of Narendranathan and Stewart (1993) and Portugal and Addison (1995) based on British Labor market data, provide empirical evidence for declining effects of unemployment benefits. In Germany there are two major types of unemployment benefits: unemployment insurance ("Arbeitslosengeld") and unemployment assistance ("Arbeitslosenhilfe"). Unemployment insurance regularly pays a certain proportion of last income for a first period of unemployment, with receipt of benefits depending on how much has been contributed to the system beforehand. After this period, unemployment assistance is paid, but the amount of support is considerably less. Under certain circumstances, there may be no financial support at all. For more details on unemployment compensation in Germany see Zimmermann (1993). In our sample, there are only few persons with no financial support during some time. Therefore, we collapse the categories "unemployment assistance" and "no financial support" and include the time–varying binary variable

B_t: Unemployment insurance benefit in month t received $(B_t = 1)$ or not $(B_t = 0)$

as a further regressor.

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

$$\lambda(t|x(t)) = \frac{\exp(\gamma_t + z'_{it}\beta_t)}{1 + \exp(\gamma_t + z'_{it}\beta_t)},$$

where $z'_{it} = (S_i, N_i, S_i N_i, A1_i, A2_i, A3_i, B_{it})$ contains the fixed or time–varying covariates above, and S_iN_i is an interaction effect between sex and nationality, with $S_iN_i = 1$ for German males (34 % observed frequency), $S_iN_i = 0$ else. The baseline–effect γ_t and time–varying covariate effects β_t are modelled by first–order random walks. We prefer first order random walks for the following reasons: Although estimates tend to be less smooth than with second–order random walks, they react more flexibly in the presence of unexpected peaks or other dynamic patterns. Furthermore, first–order random walk models reduce to traditional models with constant parameters, if corresponding error variances tend to zero. Thus, smoothness priors defined by first order random walks are in favor of horizontal lines. Our analysis is based on a final run of 41000 iterations with a burn–in period of 1000. We stored every 40th sample. We also calculated posterior mode estimators, which were in close agreement with the MCMC results.

Figure 3a) shows the estimated baseline effect γ_t. It corresponds to the hazard function for foreign females, with age between 31 and 40 years and receiving no unemployment insurance benefit. Apart from a peak at about one year of unemployment, the baseline effect is declining until month 30. The subsequent increase should not be overinterpreted: data becomes sparse at that observation period, and also censoring due to unemployment spells of more than 36 months may introduce some bias. The effects of sex and nationality in Figures 3b) and 3c) have to be interpreted together with the interaction effect of sex and nationality in Figure 3d). Figure 3c) shows that Germans have generally better chances of leaving the state of unemployment than foreigners, but this effect is vanishing over time. Employment chances are further enhanced for German men during the first year of unemployment (Figure 3d). However, this effect also vanishes later on. This may partly be explained by the fact that Germans with good chances in the labour market have already obtained a job earlier, while many of the remaining Germans are long–term unemployment persons.

Figure 3e) - g) displays the effect of age. As one might expect, younger individuals (age ≤ 30) have better chances of getting a job compared to the reference group (age from 31 to 40), especially for the first 15 months, but this effect vanishes later on. The effect of age between 41 and 50 is negative and almost constant. More surprising is the effect of age $>$ 50: It is negative at the beginning but increases distinctly towords zero with duration of

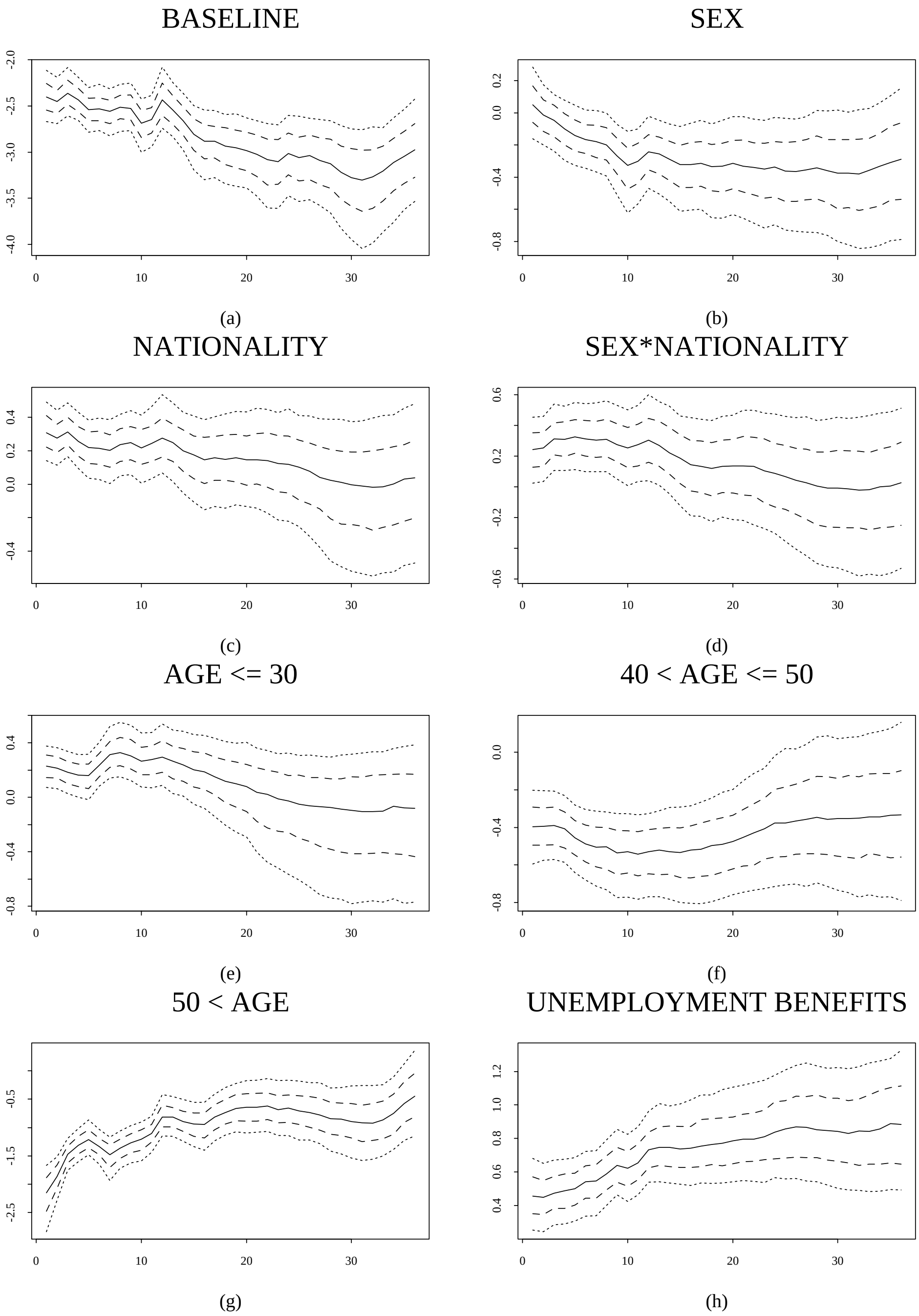

Figure 3: Time–varying effects of several covariates. Shown is the posterior median (—) within 50 % and 80 % pointwise credible regions.

unemployment. How can this be interpreted? Perhaps even more surprising is the effect of unemployment benefits: It is positive throughout and even increasing with duration of unemployment. This is in contrast to speculations that unemployment benefits foster apathy in leaving the state of unemployment. As we will see, this effect and other questions, for example the peak at about month 12 in the baseline effect, can be better interpreted and answered by a refined analysis, that distinguishes between different types of terminating events.

Figure 4 shows Bayesian pointwise credible regions for the hazard function and simultaneous credible regions for the survival function. Considered are persons aged between 31 and 40 who receive unemployment insurance benefits. All calculations are based on posterior samples from the corresponding quantities. We see that German men are likely to have a higher hazard in the first month of unemployment than German women. However, we observe the inverse trend for the second and third year, where the hazard for women seems to be even slightly higher. Consequently, the survival function is steeper for men than for women. Note that for foreign men and women hazard and survival functions are much more similar.

We explored the dependence of our conclusions upon prior specifications by a sensitivity analysis as discussed in Section 3.2. We rerun our algorithm with the value of b changed to 0.0005 for all eight variances. The results can be summarized as follows: In general, all estimated effects show a very similar pattern as for $b = 0.005$ (Figure 3). Both point and interval estimates are visually indistinguishable for parameters with a relatively high temporal variation such as the baseline effect or the effect of age > 50. The new parameter b mainly changes a lower limit for the variances, which is much smaller than the estimated variances anyway. Covariates with less temporal variation show a slightly smoother pattern with smaller variance estimates. The corresponding credible regions are slightly narrower; mainly for $t > 24$. This can be explained by the fact, that the data tends to be sparse towards the end of the observation period, so prior assumptions are still inherent in the posterior. Smaller variances therefore cause smaller credible bands. For $b = 0.05$, the patterns are now rougher for covariates with low temporal variation such as the effect of nationality.

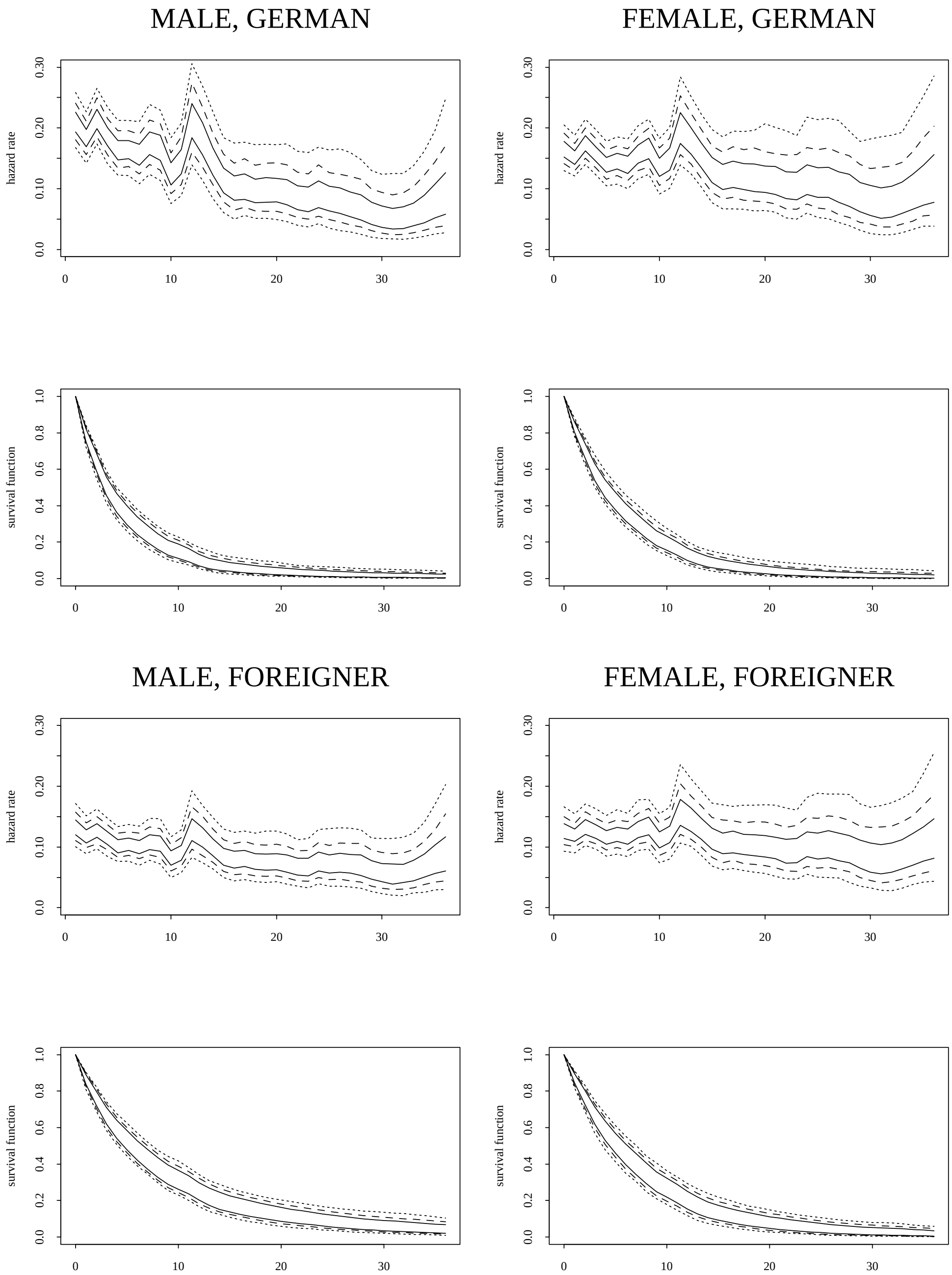

Figure 4: Hazard and survival functions for several covariate combinations. Considered are persons who received unemployment insurance benefits with age between 31 and 40. Shown are 50 %, 80 % and 95 % pointwise (hazard) and simultaneous (survival) credible regions.

To compare results obtained from our Bayesian nonparametric approach, we now reanalyze the data with a more conventional parametric model. Based on the estimated effects and associated credible regions displayed in Figure 3, we model the baseline effect and the effects of sex, age ≤ 30, age > 50 and unemployment benefits as simple functions of time, whereas the remaining effects are assumed to be constant over time. It should be noted that specification of appropriate functional forms for time–varying effects will generally be a rather difficult task without exploring patterns nonparametrically in advance.

For the baseline effect we assume a cubic polynomial

$$\gamma_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3.$$

A look at the credible regions in Figure 3h) suggests that a simple linear trend function

$$\beta_t^B = \beta_0^B + \beta_1^B t$$

is appropriate for the effect β_t^B of unemployment benefits. The effects β_t^S of sex and $\beta_t^{\text{age } >50}$ show more variation during the first 12 months of unemployment than later on. Therefore we model them by a simple regression spline, consisting of a cubic polynomial up to the cutpoint $t = 12$ and a linear trend for $t > 12$; i.e. we assume

$$\beta_t^S = \beta_0^S + \beta_1^S t + \beta_2^S \min(0, t-12)^2 + \beta_3^S \min(0, t-12)^3$$

for the effect of sex, and an analogous model for the effect of age > 50. Since there is less time–variation for the effect of age ≤ 30, we choose a piecewise constant function with a jump at $t = 12$:

$$\beta^{\text{ age} \leq 30} = \begin{cases} \beta_1 & \text{for } t \leq 12 \\ \beta_2 & \text{for } t > 12 \end{cases} .$$

Using the relation between discrete–time duration models and sequential binary models (see Fahrmeir and Tutz 1994, Ch. 9), maximum likelihood estimation can be carried out with standard software for generalized linear models. Figure 5 shows the estimated effects of baseline, sex, age > 50 and unemployment benefits. The overall shape of the baseline effect in Figure 5a) reflects the nonparametric estimate in Figure 3a), but obviously peculiarities like the peak around $t = 12$ cannot be detected by a cubic polynomial. Detailed modeling of this peak will require a more complex but less parsimonious parametric specification. The effect of unemployment benefits in Figure 5d) is quite close to the estimate obtained form the dynamic model for the first year. Later on the increase of this effect is less distinct for the nonparametric fit. This can be explained as follows: A large number of observations has duration less than about one year, while data become sparse towards the end of the observation period. The fit of the global parametric linear trend model is influenced to a large extent by the majority of observations with shorter durations. On the other hand, with a dynamic model the influence of these observations on the fit is declining as time increases.

Similar considerations have also be taken into account when comparing the effects of sex and age > 50. The effect of age > 50 in Figure 5c) is quite similar in shape to that in Figure 3g), whereas the effect of sex differs a little bit from that of Figure 3b), but is still in agreement with credible regions. Table (2) gives estimates and standard errors for the remaining effects. Comparison with Figure 3 shows again quite reasonable agreement.

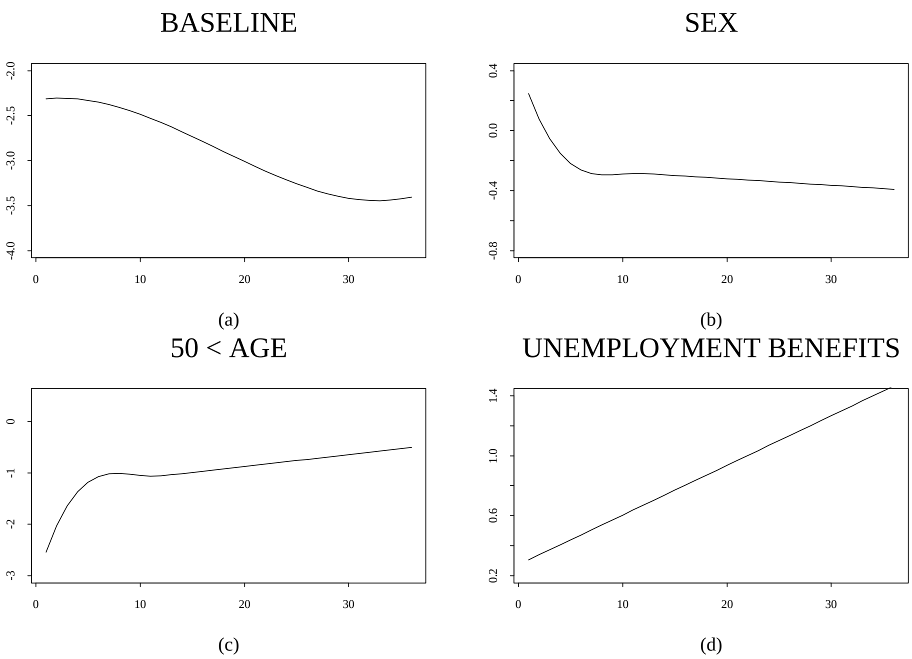

Figure 5: Estimated effects for the parametric model

As shown by this example, conventional parametric modeling of dynamic effects is possible

	Estimate	Std. Err.	95% CI	
age $\geq$ 30 years, $t \leq 12$	0.2173	0.0893	0.0422	0.3924
age $\geq$ 30 years, $t > 12$	-0.0075	0.2431	-0.4839	0.4688
$41 \leq$ age ≤ 50 years	-0.4623	0.1147	-0.6871	-0.2375
Nationality	0.2536	0.1078	0.0424	0.4649
Sex*Nationality	0.2421	0.1410	-0.0343	0.5186

Table 2: *Parameter Estimates*

and can be useful as a second step after having explored time–varying structures with non-parametric approaches in a first analysis. Without the first step however, it will generally be often quite difficult or even hopeless to specify appropriate but still parsimonious functional forms and to obtain adequate conclusions.

In our second analysis we now distinguish between three terminating events:

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/househusband, and others.

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

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

with event–specific predictors $\eta_{itr} = \gamma_{tr} + z_{it}'\beta_{tr}$. Covariate vectors z_{it} are the same as in the first analysis. Event specific baseline effects γ_{tr} and covariate effects β_{tr}, $r = 1, 2, 3$, are again modelled by first order random walks.

The baseline effect for transitions to a full–time or a part–time job show the typical smooth decreasing pattern often observed with unemployment data (Figure 6). The peak at month 12 appears only in the baseline effect for transitions to other causes. A closer look at the data shows that transitions "to retirement", "housewife/househusband" and "other reasons" are mainly responsible for this peak. A possible explanation may be that these individuals would lose unemployment insurance benefits after one year and prefer, for example, to retire. A second reason may be due to the specific kind of questions on employment status in GSOEP: Participants of the panel fill out questionnaires for every year and have to give answers on employment status retrospectively for each month. This group tends not to name a certain month but instead simply name the beginning or end of a year as the time of leaving the status of unemployment. The effects of sex and nationality are also now much better to interpret. For example, there is a distinct positive effect for transitions to a full–time job for men, but also a distinct negative effect for transitions to a part–time job. The nationality effect provides clear evidence that German females have highly increased chances of getting a part–time job; maybe they are much more interested in getting part–time jobs.

Also the effects of age can be better explained now (Figure 7). In particular, looking at the effects of age > 50, we see that chances for getting full–time and part–time jobs are significantly deteriorated and do not improve with increasing duration of unemployment. However, the effect of transitions to other causes is near zero and even slightly increasing. This supports presumptions that older individuals prefer to retire, to become housewife/househusband or to leave the unemployment register for other reasons. We also see that the time–varying effect of age > 50 in Figure 3g) is largely caused by confounding the effect of the three types of transitions into the effect of only one terminating cause. Also, the effects of unemployment insurance benefits can now be interpreted correctly: The effect is constantly positive for transitions to a full–time job, presumably since individuals with unemployment insurance benefits had regular jobs earlier and thus get easier offers for a new full–time job. On the other side, the effect is clearly negative for transitions to part–time jobs. A possible explanation is that some individuals with good financial support from unemployment insurance are less motivated to get a part–time job.

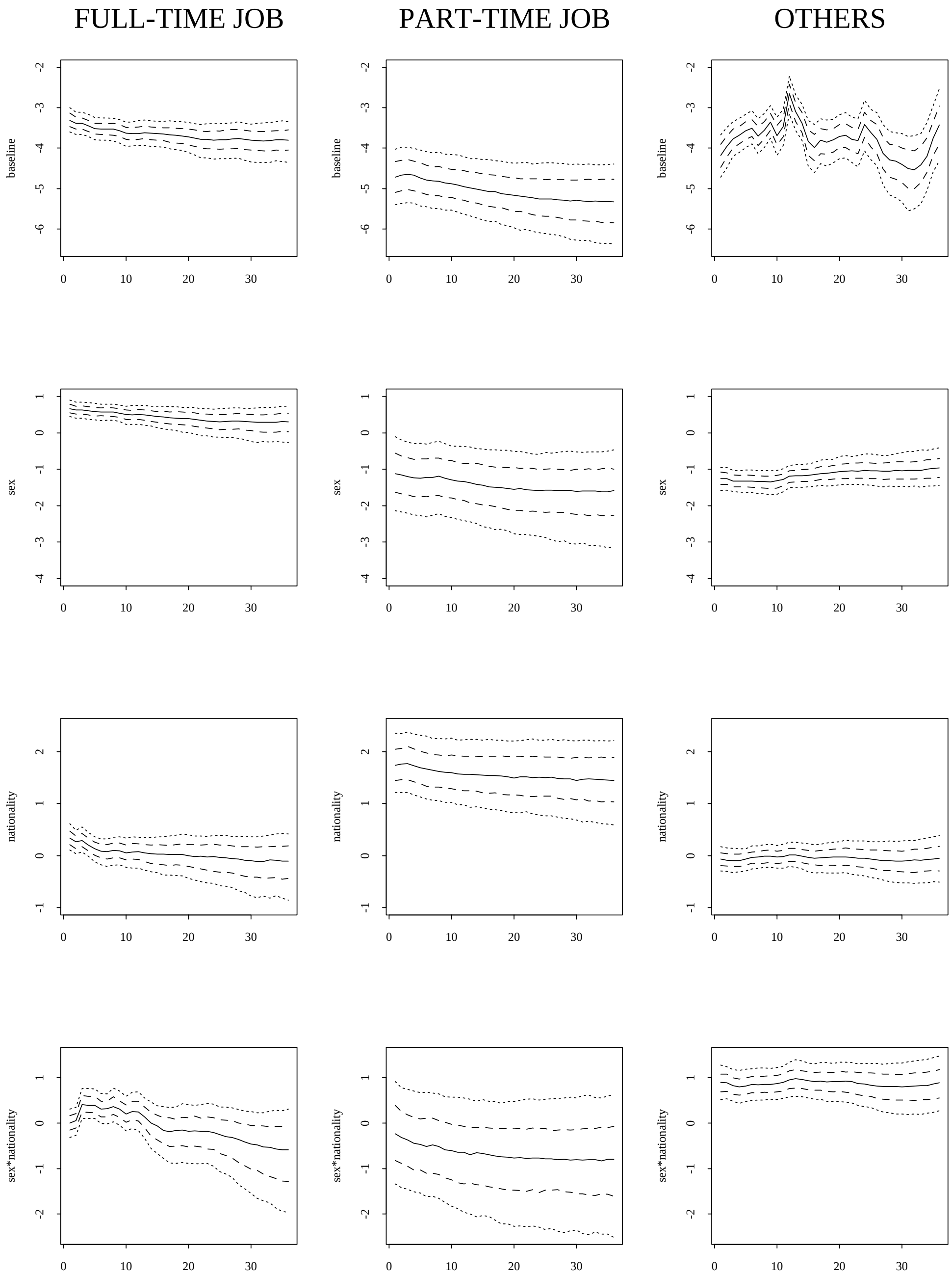

Figure 6: Covariate effects for different types of terminating events. Same credible regions as in Figure 3.

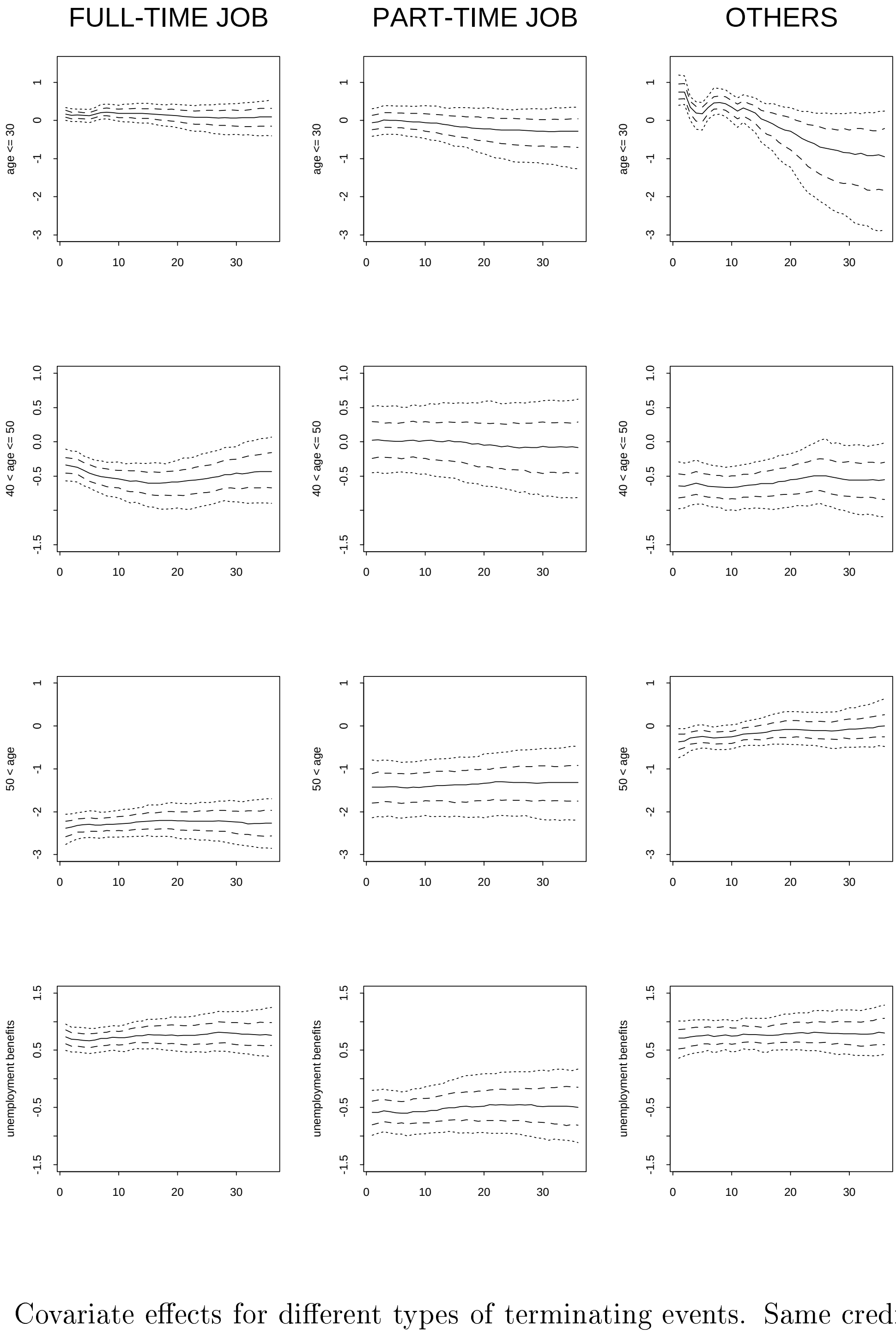

Figure 7: Covariate effects for different types of terminating events. Same credible regions as in Figure 3.

For a parametric reanalysis, we first fitted a model with baseline effects specified as low order polynomials, but kept all covariate effetcs time–constant. The shape of the estimated baseline effects was in agreement with our nonparametric analysis, although the peaks in the baseline effect for "others" could not be reproduced. Attemps to model time–varying structures in more detail by inclusion of time–varying effects, led to divergence of ML estimates due to the large number of parameters involved. This illustrates that nonparametric approaches, imposing appropriate smoothness restrictions, are useful tools for refined and flexible analyses.

5 Conclusions

We conclude with some discussion of topics not treated in detail in the main text, e.g. other estimation approaches, multiple time scales and unobserved heterogeneity.

In this paper, we propose a certain type of Metropolis–Hastings algorithm. In our context, it has distinct advantages compared to a Gibbs step combined with rejection sampling based on knowledge of envelope functions and log–concavity of conditionals (e.g. Gilks and Wild 1992). Other MCMC sampling schemes for dynamic models have been suggested by Gamerman (1995) and Shephard and Pitt (1995), but they require distinctly more computation time per iteration. The multi move schemes of Carter and Kohn (1994) and Frühwirth–Schnatter (1994), designed for observation models with errors from normal or mixtures of normals, cannot be extended to discrete observation models.

As a conceptually simpler alternative that avoids MCMC at all, posterior mode estimation – obtained by maximizing the unnormalized posterior – has been considered in Fahrmeir (1994, for $m = 1$) and Fahrmeir and Wagenpfeil (1996). This approach can be viewed as an empirical Bayes method, since the matrix Q is treated as fixed and unknown, not as a random variable with some prior distribution. Posterior mode estimation has also a non–Bayesian interpretation, being equivalent to maximization of a penalized likelihood. Efficient estimation can be carried out by iterative Kalman filtering and smoothing.

However, posterior mode estimation suffers from some disadvantages: Duration data usually becomes sparse towards the end of the observation period, so inference based on approximate posterior normality will be questionable. Also, like for any empirical Bayes approach, the uncertainty associated with estimates of $\{\alpha_t\}$ is underestimated, since no allowance is made for the uncertainty associated with Q. Furthermore, estimation of functionals of α, such as hazard or survival functions, has to be based on further approximations like the Delta method. Posterior mode estimation is, nevertheless, useful as an ingredient of a fully Bayesian approach: It provides an initial solution for a refined analysis and can be used to check convergence behavior of simulation–based Monte Carlo methods.

Our concepts and also our applications focused on one time dimension, i.e. that of duration in a certain state, regarding other time scales such as calendar time, age or cohort as methodologically secondary. Although the models of this paper allow the inclusion of other time scales through covariates, they are not built to deal with multiple time scales in a symmetric way. Here we outline how this could be achieved. For simplicity, let us only consider the case of two time scales: duration time t and calendar time u, with one terminating event. The hazard function, now depending on t, u and covariates, may be modelled by

$$\lambda(t, u|x) = h(\gamma_t + \theta_u + \theta_u^s + x'\beta_t),$$

where x is a, possibly time–dependent, covariate vector, γ_t is the baseline effect for duration time t, θ_u is a trend component in calendar time, e.g. a random walk of first order, and θ_u^s may be a monthly seasonal component, e.g. $\theta_u^s + \cdots + \theta_{u-11}^s = v_u \sim N(0, \sigma_u^2)$, in calendar time. Multiple time scales models require appropriate conditional independence assumptions leading to modified full conditionals for MCMC simulation. In principle real–time effects θ_u, θ_u^s can be updated using additional Metropolis–Hastings steps analogous to those in Section 3. However, the simple structure of conditional likelihoods is destroyed, leading to a considerably increasing amount of implementation and computation requirements. We plan to consider dynamic models for multiple time scales and to develop efficient MCMC methods for such models in future research. Close in spirit is the work by Berzuini and Clayton (1994), who discuss survival models with multiple time scales and time–constant covariate effects.

Our dynamic model specifications in Section 2 allow rather flexible modeling of time–varying hazards and covariate effects. However, they do not explicitly take into account unobserved heterogeneity or frailty. For example, differences between short– and longterm unemployment might be considered as a potential source of unobserved heterogeneity that is not or insufficiently explained by observed variables. The effect of neglected heterogeneity on estimation of hazards and covariate effects in duration models with fixed parameters has been studied by a number of authors, see e.g. Lancaster (1990). The most important consequence of neglecting unobserved heterogeneity is that it may appear as spurious duration dependence.

The conventional procedure to account for heterogeneity is to introduce unit–specific parameters, say θ_i, in the linear predictor and to assume that they are random effects, distributed according to some mixing distribution f_θ. Two main approaches to the modeling of this mixing distribution have been proposed. The first assumes a parametric form, e.g. a log–Gamma or a normal density, for f_θ. For discrete–time duration models with fixed effects and a single terminating event ($m = 1$), one treatment is to extend the linear predictor η_{it} additively to

$$\tilde{\eta}_{it} = \eta_{it} + \theta_i, \quad \theta_i \quad \text{i.i.d} \sim N(0, \sigma^2),$$

and to carry out inference by MCMC, see Raftery, Lewis and Aghajanian (1995). Clayton (1991) uses a log–Gamma distribution instead in so called frailty models. This approach can be combined with dynamic models by extending the linear predictor to

$$\eta_{it} = \alpha_t + x_{it}'\beta_t + \theta_i, \quad \theta_i \quad \text{i.i.d} \sim N(0, \sigma^2)$$

and to add a further full conditional for θ_i in the MCMC updating steps. For panel data with many repeated events, such mixed dynamic models have been successfully implemented and applied by Knorr–Held (1995). For duration models, without repeated events, there is some evidence given in the literature that estimates can be very sensitive to the choice of the mixing distribution, see e.g. Meyer (1990). The likelihood of observations becomes rather flat, so that the prior has much influence on the posterior. This is also to be expected for the second approach, where a discrete distribution, typically with small number of mass points, is chosen for f_θ (Heckman and Singer 1984). In addition, the effect of heterogeneity decreases with flexible models for baseline hazards (Narendranathan and Stewart 1993) and may be even less serious if time–varying covariate effects are introduced. For duration models with several terminating events ($m > 1$), these problems become even more evident, since the extension to this case is accompanied by additional prior assumptions. It is therefore likely that misspecification of the mixing distribution can be worse than omitting heterogeneity. Therefore, and since our interest here lies in allowing flexibility in form of time–varying effects of duration dependence, we have restricted attention to models without heterogeneity. This has to be kept in mind for a careful interpretation of the results in Section 4. For example, the time–varying effect of nationality in Figure 3c) reflects differences in short–term and long–term unemployment between Germans and non–Germans. Concerning short unemployment, Germans have better chances for leaving unemployment, but this effect vanishes for longterm unemployment. Thus, time–varying effects may be interpreted as caused by unobserved heterogeneity.

Other interesting extensions, where our approach should be useful, are dynamic continuous–time duration models, e.g. the dynamic piecewise exponential model development by Gamerman (1991), with an application to unemployment data in Gamerman and West (1987), and event history models for multiple cycles and states e.g. semi–Markov models.

Obviously, a large number of possible models raise questions about model determination and validation, that are beyond the scope of this paper. Bayesian model choice via MCMC is currently an intensive research area; promising solutions are based on Bayes factors (Lewis and Raftery 1994, Raftery 1996) or on predictive distributions; see Gelfand (1996).

References

Aalen, O. O. 1989. "A linear regression model for the analysis of life–times." *Statistics in Medicine 8*: 907–925.

Aalen, O. O. 1993. "Further results on the non–parametric linear regression model in survival analysis." *Statistics in Medicine 12*: 1569–1588.

Allison, P. D. 1982. “Discrete–time Methods for the Analysis of Event Histories.” In *Sociological Methodology 1982*, edited by S. Leinhardt, 61–89. San Francisco: Jossey–Bass.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. 1993. *Statistical models based on counting processes.* New York: Springer Verlag.

Arjas, E., and Haara P. 1987. “A Logistic Regression Model for Hazard: Asymptotic Results.” *Scandinavian Journal of Statistics 14*: 1–18.

Arjas, E., and Liu, L. 1995. “Assessing the losses caused by an intervention: A hierarchical Bayesian approach.” *Applied Statistics 44*: 357-368.

Berzuini, C., and Clayton, D. 1994. “Bayesian Analysis of Survival on Multiple Time Scales.” *Statistics in Medicine 13*: 823–838.

Berzuini, C, and Larizza, C. 1996. “A Unified Approach for Modelling Longitudinal and Failure Time Data, with Application in Medical Monitoring.” *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*: 109–123.

Besag, J. E. 1974. “Spatial interaction and the statistical analysis of lattice systems.” *Journal of the Royal Statistical Society B 36*: 192–236.

Besag, J. E., Green, P. J., Higdon, D., and Mengersen, K. 1995. “Bayesian computation and stochastic systems (with discussion).” *Statistical Science 10*: 3–66.

Besag, J. E., and Kooperberg, C. 1995. “On conditional and intrinsic autoregressions.” *Biometrika 82*: 733–746.

Blossfeld, H.–P., Hamerle, A., and Mayer, K. U. 1989. *Event History Analysis.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Carter, C. K., and Kohn, R. 1994. “On Gibbs sampling for state space models.” *Biometrika 81*: 541–553.

Casella, G., and George, E. I. 1992. “Explaining the Gibbs sampler.” *The American Statistician 46*: 167–174.

Chib, S., and Greenberg, E. 1995. “Understanding the Metropolis–Hastings Algorithm.” *The American Statistician 49*: 327–335.

Clayton, D. G. 1991. “A Monte Carlo Method for Bayesian Inference in Frailty Models.” *Biometrics 47*: 467–485.

Cowles, M. K., and Carlin, B. P. 1996. “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.” *Journal of the American Statistical Association 91*: 883–904.

Fahrmeir, L. 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. 1996. “Smoothing Hazard Functions and Time–Varying Effects in Discrete Duration and Competing Risks Models.” *Journal of the American Statistical Association*, to appear. [1]

Frühwirth–Schnatter, S. 1994. “Data augmentation and dynamic linear models.” *Journal of Time Series Analysis 15*: 183–202.

Gamerman, D. 1991. “Dynamic Bayesian Models for Survival Data.” *Applied Statistics 40*: 63–79.

Gamerman, D. 1995. “Monte Carlo Markov Chains for Dynamic Generalized Linear Models.” Unpublished manuscript. Instituto de Matematica, Universidade Federal do Rio de Janerio.

Gamerman, D., and West, M. 1987. “An Application of Dynamic Survival Models in Unemployment Studies.” *The Statistician 36*: 269–274.

[1] Available under `http://www.stat.uni-muenchen.de/sfb386/publikation.html`

Gelfand, A. E. 1996. "Model determination using sampling–based methods." In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter, 145–161. London: Chapman & Hall.

Gelman, A. 1996. "Inference and monitoring convergence." In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter, 131–143. London: Chapman & Hall.

Gilks, W. R., Richardson, S, and Spiegelhalter, D. J. 1996. "Introducing Markov Chain Monte Carlo." In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter, 1–19. London: Chapman & Hall.

Gilks, W. R., and Wild, P. 1992. "Adaptive rejection sampling for Gibbs sampling." *Applied Statistics 41*: 337–348.

Hamerle, A., and Tutz, G. 1989. *Diskrete Modelle zur Analyse von Verweildauern und Lebenszeiten.* New York: Campus.

Hastie, T., and Tibshirani, R. 1993. "Varying–coefficient Models." *Journal of the Royal Statistical Society B 55*: 757–796.

Hastings, W. K. 1970. "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika 57*: 97–109.

Heckman, J. J., and Singer, B. L. 1984. "Econometric duration analysis." *Journal of Econometrics 24*: 63–132.

Hill, D. H., Axinn, W. G., and Thornton, A. 1993. "Competing Hazards with Shared Unmeasured Risk Factors." In *Sociological Methodology 1993*, edited by P. V. Marsden, 245–277. Oxford: Basil Blackwell.

Huffer, F., and Mc Keague, I. 1991. "Weighted least squares estimation for Aalen's additive risk model." *Journal of the American Statistical Association 86*: 114–129.

Kalbfleisch, J. D., and Prentice, R. L. 1980. *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons.

Knorr–Held, L. 1995. “Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models.” Discussion paper Nr. 8, Sonderforschungsbereich 386, Universität München. [2]

Knorr–Held, L. 1996. “Conditional Prior Proposals in Dynamic Models.” Discussion paper Nr. 36, Sonderforschungsbereich 386, Universität München. [3]

Lancaster, T. 1990. *The Econometric Analysis of Transition Data.* Cambridge, Cambridge University Press.

Lewis, S. M., and Raftery, A. 1994. “Estimating Bayes Factors via Posterior Simulation with the Laplace-Metropolis Estimator.” Technical Report, University of Washington.

Meyer, B. 1990. “Unemployment insurance and unemployment spells.” *Econometrica 58*: 757–782.

Narendranathan, W., and Stewart, W. M. 1993. “Modelling the probability of leaving unemployment: competing risks models with flexible base–line hazards.” *Applied Statistics 42*: 63–83.

Portugal, P., and Addison, J. T. 1995. “Short– and long–term unemployment: a parametric model with time–varying effects.” *Oxford Bulletin of Economics and Statistics 57*: 205–227.

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.

Raftery, A. E. 1996. “Hypothesis testing and model selection.” In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter, 163–187. London: Chapman & Hall.

[2] Available under `http://www.stat.uni-muenchen.de/sfb386/publikation.html`

[3] Available under `http://www.stat.uni-muenchen.de/sfb386/publikation.html`

Raftery, A. E., and Banfield, J. D. 1991. “Stopping the Gibbs sampler, the Use of Morphology, and Other Issues in Spatial Statistics.” *Annals of the Institute of Statistical Mathematics: 43*: 32–43.

Raftery, A. E., and Lewis, S. M. 1996. “Implementing MCMC” In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D. J. Spiegelhalter, 115–130. London: Chapman & Hall.

Raftery, A. E., Lewis, S. M., and Aghajanian, A. 1995. “Demand or Ideation? Evidence from the Iranian Marital Fertility Decline.” *Demography 32*: 159–182.

Shephard, N. & Pitt, M. K. 1995. “Parameter–driven exponential family models.” Unpublished manuscript. Nuffield College, Oxford, UK.

Thompson, W. A., Jr. 1977. “On the Treatment of Grouped Observations in Life Studies.” *Biometrics 33*:463–470.

Tierney, L. 1994. “Markov Chains for Exploring Posterior Distributions (with discussion).” *Annals of Statistics 22*: 1701–1762.

Tutz, G. 1995. “Dynamic modelling of discrete duration data: a local likelihood approach.” Report 95-15, Tech. Univ. Berlin, Institut für Quantitative Methoden.

Yamaguchi, K. 1993. “Modeling Time–Varying Effects of Covariates in Event–History Analysis Using Statistics from the Saturated Hazard Rate Model.” In *Sociological Methodology 1993*, edited by P. V. Marsden, 279–317. Oxford: Basil Blackwell.

Zimmermann, K. F. 1993. “Labor Responses to Taxes and Benefits in Germany.” In *Incentives, A North European Perspective*, edited by Atkinson A. B. and Mogenson, G. V., 192–240, Oxford: Clarendon Press.