Nonparametric Bayesian hazard rate models based on penalized splines

Ludwig Fahrmeir and Andrea Hennerfeind

Department of Statistics, University of Munich

Abstract

Extensions of the traditional Cox proportional hazard model, concerning the following features are often desirable in applications: Simultaneous nonparametric estimation of baseline hazard and usual fixed covariate effects, modelling and detection of time–varying covariate effects and nonlinear functional forms of metrical covariates, and inclusion of frailty components. In this paper, we develop Bayesian multiplicative hazard rate models for survival and event history data that can deal with these issues in a flexible and unified framework. Some simpler models, such as piecewise exponential models with a smoothed baseline hazard, are covered as special cases. Embedded in the counting process approach, nonparametric estimation of unknown nonlinear functional effects of time or covariates is based on Bayesian penalized splines. Inference is fully Bayesian and uses recent MCMC sampling schemes. Smoothing parameters are an integral part of the model and are estimated automatically. We investigate performance of our approach through simulation studies, and illustrate it with a real data application.

1 Introduction

Cox's proportional hazard model is a benchmark method in survival and event history data analysis. In this model, the hazard rate is assumed as the product

$$\lambda(t; v) = \lambda_0(t) \exp(\gamma_1 v_1 + \ldots + \gamma_r v_r) = \lambda_0(t) \exp(v'\gamma). \tag{1}$$

The baseline hazard rate is unspecified, and, through the exponential link function, the covariates $v = (v_1, \ldots, v_r)$ act multiplicatively on the hazard rate. To keep notation simple, covariates are assumed to be time–independent in (1), but extensions to time–dependent covariates are possible under usual assumptions. In a number of applications there is a need for extending this basic model with respect to several aspects, such as simultaneous estimation of baseline hazards and covariate effects, allowing more flexible nonlinear functional forms for covariates, inclusion of time-varying effects, thereby dropping the proportional hazard assumption, and incorporation of unobserved heterogeneity or frailty. Increased flexibility becomes even more important for more complex life or event history data with recurrent events and multiple states. Various extensions have been suggested to deal with such issues. For example, the piecewise exponential model is a simple approach for estimating the baseline hazard jointly with covariates, Hastie and Tibshirani (1993) estimate nonlinear and time–varying covariate effects by smoothing splines via penalized partial likelihood, and Aalen's additive model (1989, 1993) focuses on time–varying effects using a martingale approach. Fahrmeir and Klinger (1998) develop full penalized likelihood inference for event history analyses, modelling time–varying effects by penalized step functions and/or smoothing splines.

In this paper, we propose nonparametric Bayesian hazard rate models that can deal with these issues in a flexible and unified framework. Inference uses the information from the full likelihood instead of a partial likelihood in combination with appropriate priors. It contains some previous approaches as special cases, and extensions to more complex event history data are conceptually easy. For survival data, we reparametrize the baseline hazard rate through $\exp\{f_0(t)\}$, $f_0(t) = log\{\lambda_0(t)\}$, and we extend model (1) to the nonparametric multiplicative model

$$\lambda(t; x, z, v) = \exp\{\eta(t; x, z, v)\} \tag{2}$$

with predictor

$$\eta(t; x, z, v) = f_0(t) + \sum_{j=1}^{p} f_j(t) z_j + \sum_{j=p+1}^{p+q} f_j(x_j) + v'\gamma. \tag{3}$$

The function $f_0(t)$ is the baseline effect, and a function $f_j(t)$ represents a time–varying effect of the covariate z_j, for example the time–varying effect of a therapy. The functions $f_1(x_1), \ldots, f_q(x_q)$ are possibly nonlinear effects of metrical covariates $x_1, \ldots, x_q$, and $v'\gamma$ is the usual linear part of the predictor. To account for unobserved heterogeneity or frailty, random intercepts and slopes can be incorporated in (3). A generalization to models for event history analysis is described in Section 2. Estimation of the unknown functions is based on penalized spline (P-spline) regression, introduced by Eilers and Marx (1996), Marx and Eilers (1998) for generalized additive models in a frequentist setting. We will use Bayesian versions (Lang and Brezger, 2002) as a basic building block. Basically, time t is treated in the same way as a metrical covariate x, but the degree and amount of smoothness may be different. For example, simple random walk priors for the baseline effect $f_0(t)$ in a piecewise exponential model are P-splines of degree zero. Inference is fully Bayesian and uses the information from the full likelihood, instead of a partial likelihood, in combination with appropriate priors. Posterior analysis is carried out with computationally efficient MCMC techniques. Some advantages are: Smoothing parameters are an integral part of the model and can be estimated jointly with unknown functions and other parameters; predictive hazard rates and survivor functions can be computed directly from the MCMC output instead of using plug-in estimates; and no asymptotic approximations or conjectures have to be made. Inferential procedures have been implemented in C++ as part of *BayesX* (Brezger, Kneib and Lang, 2002).

Nonparametric Bayesian survival models have become quite popular in recent years. Ibrahim, Chen and Sinha (2001) provide a very good introduction and overview. Some previous work deals with related, special cases of our approach. Joint estimation of the baseline hazard and usual linear covariate effects in the Cox model has been considered by several authors. Gamerman (1991) proposes a Gaussian random walk model for $log\{\lambda_0(t)\}$ in the piecewise exponential model, and Sinha (1993) suggests a joint Gaussian smoothness prior. Arjas and Gasbarra (1994) introduce a first order autoregressive gamma model for $\lambda_0(t)$, and Cai, Hyndman and Wand (2002) use a mixed model representation of linear regression splines to estimate the baseline hazard. Time–varying effects have been treated within a state space framework by Gamerman (1991) for the piecewise exponential model, and Fahrmeir (1994), Fahrmeir and Wagenpfeil (1996) for discrete time survival and competing risks models. In all these approaches, however, covariate effects are assumed to be of the usual linear form, and are mostly restricted to the important but special case of survival analysis.

The rest of the paper is organized as follows. In Section 2 we describe models, likelihoods, and priors for unknown functions and parameters. Inference is outlined in Section 3. Performance is studied in Section 4 through simulation studies. The application in Section 5 illustrates the method. The concluding section contains some proposals for future research.

2 Models, likelihoods and priors

Consider n individuals and let N_{hi}, for $h = 1, ..., k$, $i = 1, ..., n$, denote the individual counting processes for events of type h, where $N_{hi}(t)$ is the number of observed type h events experienced by the ith individual up to time t. We assume that individual intensity processes exist and have multiplicative structure:

$$\alpha_{hi}(t) = Y_{hi}(t)\lambda_{hi}\{t; z_{hi}(t), x_{hi}(t), v_{hi}(t)\} \tag{4}$$

where $Y_{hi}(t)$ are left–continuous 1-0 processes indicating whether or not individual i is at risk of experiencing a type h event just before time t. The individual type h hazard or transition rate λ_{hi} in (4) depends on t and on possibly type–specific and time–dependent covariates. As in (3), the covariate vector $z_{hi}(t)$ is assumed to have time–varying effects, $x_{hi}(t)$ consists of metrical covariates with possibly nonlinear effects, and $v_{hi}(t)$ comprises covariates with linear effects.

Right censored survival data with lifetimes T_i, independent censoring times C_i, $i = 1, ..., n$, observed lifetimes $t_i = min(T_i, C_i)$, and censoring indicators δ_i are a special case with $h = 1$, $N_i(t) = I(T_i \leq t, \delta_i = 1)$, $Y_i(t) = I(t_i \geq t)$ and $\lambda_i(t)$ as in (1) or (2) and (3). The hazard rate $\lambda_{hi}(t)$ for individual i is assumed to follow a multiplicative model

$$\lambda_{hi}(t) := \lambda_{hi}(t; z_{hi}(t), x_{hi}(t), v_{hi}(t)) = \exp(\eta_{hi}(t)), \tag{5}$$

with the general form of the predictor given by

$$\eta_{hi}(t) = f_0^h(t) + \sum_{j=1}^{p} f_j^h(t) z_{hij}(t) + \sum_{j=p+1}^{p+q} f_j^h(x_{hij}(t)) + v'_{hi}(t)\gamma_h. \tag{6}$$

Here $f_0^h(t) = log\{\lambda_0^h(t)\}$ is the baseline effect, $f_j^h(t)$ are time–varying effects of covariates $z_{hj}(t)$, $f_j^h(x_{hj}(t))$ is the nonlinear effect of $x_{hj}(t)$, γ_h is the vector of usual linear fixed effects. As a further extension, i.i.d. frailty effects and random slopes could be introduced in (6), but we omit this here.

For given predictors $\eta = \{\eta_{hi},\ h = 1, ..., k,\ i = 1, ..., n\}$, the likelihood is given by

$$L(\eta) = \prod_{i=1}^{n} \prod_{h=1}^{k} \int_0^\infty \lambda_{hi}(s) dN_{hi}(s) \cdot \exp\left\{ - \int_0^\infty Y_{hi}(s)\lambda_{hi}(s) ds \right\}. \tag{7}$$

For survival data with noninformative right censoring, the likelihood (7) reduces to the well–known form

$$\begin{aligned} L &= \prod_{i=1}^{n} \lambda_i(t)^{\delta_i} \cdot \exp\left(-\int_0^{t_i} \lambda_i(u)du\right) \\ &= \prod_{i=1}^{n} \lambda_i(t)^{\delta_i} \cdot S_i(t)\,, \end{aligned} \tag{8}$$

Note that the first term in (7) is always a sum because $N_{hi}(s)$ is a step function. Numerical problems arise in the evaluation of the second integral in (7) and (8). Only if time–varying functions in the predictor are step functions, this integral also reduces to a sum. A prominent case is the piecewise exponential model, where the baseline hazard in (1), or (2) and (3) is assumed as a step function, see also Section 3.

The Bayesian model formulation is completed by assumptions about priors for parameters and functions. Because the priors do not depend on the type h of events we omit the index h to simplify notation. For fixed effect parameters γ we assume diffuse priors $p(\gamma) \propto const$. A weakly informative normal prior would be another choice.

For unknown functions f_j, we assume Bayesian P–spline priors as in Lang and Brezger (2002). Random walk priors, which have been suggested in Fahrmeir and Lang (2001) and may be used as smoothness priors for the baseline effect and dynamic effects in a piecewise exponential model, appear as a special case. The basic idea of P-spline regression (Eilers and Marx, 1996) is to approximate a function $f_j(x)$ as a linear combination of B-spline basis functions B_m, i.e.

$$f_j(x) = \sum_{m=1}^{M_j} \beta_{jm} B_m(x).$$

The basis functions B_m are B–splines of degree l defined over a grid of equally spaced knots $x_{j\,min} = \xi_0 < \xi_1 < \ldots < \xi_s = x_{j\,max}$, $M_j = l + s$. The number of knots is moderate, but not too small, to maintain flexibility, but smoothness of $f(x)$ is encouraged by difference penalties for neighboring coefficients in the sequence $\beta_j = (\beta_{j1}, \ldots, \beta_{jM_j})'$. The Bayesian analogue are first or second order random walk smoothness priors

$$\beta_{jm} = \beta_{j,m-1} + u_{jm} \qquad or \qquad \beta_{jm} = 2\beta_{j,m-1} - \beta_{j,m-2} + u_{jm} \tag{9}$$

with i.i.d. Gaussian errors $u_{jm} \sim N(0, \tau_j^2)$ and diffuse priors $p(\beta_{j1}) \propto const$, or $p(\beta_{j1})$ and $p(\beta_{j2}) \propto const$, for initial values. A first order random walk penalizes abrupt jumps $\beta_{jm} - \beta_{j,m-1}$, and a second order random walk penalizes deviations from a linear trend. The amount of smoothness or penalization is controlled by the variance τ_j^2, which acts as a smoothness parameter. Note that B–splines of order $l = 0$ are 0-1 functions, and $f_j(x)$ is a step function with value β_{jm} in interval m. Then (9) is a random walk smoothness prior for the function values itself. An important special case are piecewise exponential models with a random walk prior for the log–baseline hazard.

The joint prior of the regression parameters β_j is Gaussian and can be easily computed as a product of conditional densities defined by (9) as

$$\beta_j \mid \tau_j^2 \propto \exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right).$$

The penalty matrix K_j is of the form $K_j = D'D$, where D is a first or second order difference matrix. For second order random walks, for example, K_j is given by

$$K_j = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}$$

with zero elements outside the second off–diagonals.

The band structure of K_j is very useful for computationally efficient MCMC updating schemes.

A common choice for approximating smooth curves are quadratic or cubic B-splines. Computationally, linear splines are simpler. The simplest choice are B–splines of degree zero, i.e. $B_m(x) \equiv 1$ over the m-th interval, and $B_m(x) \equiv 0$ elsewhere. Then $f(x)$ is approximated by a piecewise constant function, and the function values follow a random walk model as in Fahrmeir and Lang (2001). This special choice, with time t as covariate, is the easiest way to smooth the baseline in the piecewise exponential model; moreover the integral in the likelihood (8) reduces to a sum, see the next section. With P–splines of higher degree, however, estimation of smooth baseline effects is improved considerably in terms of MSE's, see Section 4.

Variances τ_j^2 follow inverse Gamma priors $IG(a_j; b_j)$. The hyperparameters a_j, b_j are chosen such that this prior is weakly informative. We routinely use $a_j = b_j = 0.001$ as a standard choice. For moderate to large data sets, results are rather insensitive to choice of a_j and b_j. For smaller data sets, a sensitivity analysis is useful.

The Bayesian model specification is completed by assuming that priors $p(\beta_j \mid \tau_j^2)$, $j = 0, ..., p+q$, are conditionally independent, and that all priors are mutually independent.

3 Markov chain Monte Carlo inference

In what follows, let $\beta = (\beta_0, ..., \beta_p, \beta_{p+1}, ..., \beta_{p+q})'$ denote the vector of all B–spline regression coefficients, γ the vector of fixed effects, and $\tau^2 = (\tau_0^2, ..., \tau_{p+q}^2)$ the vector of all variance components. Full Bayesian inference is based on the entire posterior distribution

$$p(\beta, \gamma, \tau^2 \mid data) \propto L(\beta, \gamma, \tau^2)\, p(\beta, \gamma, \tau^2).$$

Due to the (conditional) independence assumptions, the joint prior factorizes into

$$p(\beta, \gamma, \tau^2) = \left\{ \prod_{j=0}^{p+q} p(\beta_j \mid \tau_j^2) p(\tau_j^2) \right\} p(\gamma),$$

where the last factor can be omitted for diffuse fixed effect priors.

The likelihood $L(\beta, \gamma, \tau^2)$ is given by inserting (5), (6) into (7) or (8), but the integral requires integration over all terms depending on survival time t, i.e. terms of the form

$$\int_0^{t_i} \exp\left(f_0(u) + \sum_{j=1}^{p} f_j(u) z_{ij}(u) \right) du,$$

where $f_j(t) = \sum \beta_{jm} B_m(t)$. Apart from B–splines $B_m(t)$ of degree zero, i.e. random walk models, and linear B–splines, these integrals are not available in closed form. The first case leads to the piecewise exponential model: The time axis is divided into a grid

$$0 = \xi_0 < \xi_1 < ... < \xi_{t-1} < \xi_t < ... < \xi_s = t_{max},$$

and $f_j(t)$ is assumed to be a piecewise constant function, i.e.

$$f_j(t) = \beta_{jt}$$

in time interval $(\xi_{t-1}, \xi_t]$, $t = 1, ..., s$. In this case, the integral reduces to a sum, and, after some simple calculations, the likelihood can be rewritten in the form of a Poisson–likelihood, with the predictor η_{it} containing an additional offset term, see Fahrmeir and Tutz (2001, Section 9.1) or Ibrahim, Chen and Sinha (2001, Section 3.1) for details.

For linear B–splines, the integrals can still be solved analytically, but expressions are rather messy and the computational effort is quite high, see Cai et al. (2002, Appendix). Following their suggestion, we use simple numerical integration in form of the trapezoidal rule for linear B–splines as well as the commonly used cubic B–splines, where analytical integration is not possible anyway.

Full Bayesian inference via MCMC simulation is based on updating full conditionals of single parameters or blocks of parameters, given the rest of the data.

For updating the parameters β_j, $j = p+1, ..., p+q$, which correspond to the time–independent functions $f_j(x_j)$, as well as fixed effects γ , we use an MH–algorithm based on iteratively weighted least squares (IWLS) proposals, developed for fixed and random effects by Gamerman (1997) and adapted to generalized additive mixed models in Brezger and Lang (2003). Suppose we want to update β_j, with current value β_j^c of the chain. Then a new value β_j^p is proposed by drawing a random vector from a (high–dimensional) multivariate Gaussian proposal distribution $q(\beta_j^c, \beta_j^p)$, which is obtained from a quadratic approximation of the log–likelihood by a second order Taylor expansion with respect to β_j^c, in analogy to IWLS iterations in generalized linear models. The proposed vector β_j^p is accepted as the new state of the chain with probability

$$\alpha(\beta_j^c, \beta_j^p) = min\left(1, \frac{p(\beta_j^p \mid \cdot) q(\beta_j^p, \beta_j^c)}{p(\beta_j^c \mid \cdot) q(\beta_j^c, \beta_j^p)}\right)$$

where $p(\beta_j \mid \cdot)$ is the full conditional for β_j (i.e. the conditional distribution of β_j given all other parameters and the data).

For a fast implementation, we use the fact that the precision matrices of the Gaussian proposal distributions are banded, so that Cholesky decompositions can be performed efficiently.

For the parameters $\beta_0, ..., \beta_p$ corresponding to the functions $f_0(t), ..., f_p(t)$ depending on time t, the IWLS–MH algorithm requires considerably more computational effort, because the integrals in the log–likelihood as well as first and second derivatives are involved now. Therefore, we adopt a computationally faster MH–algorithm based on conditional prior proposals first developed by

Knorr–Held (1999) for state space models and extended for generalized additive mixed models in Fahrmeir and Lang (2001). It requires only evaluation of the log–likelihood, not of derivatives. However, draws are not performed for the entire vector β_j, but iteratively for blocks of subvectors, see Fahrmeir and Lang (2001) for details.

The full conditionals for the variance parameters τ_j^2 are inverse gamma with parameters

$$a_j' = a_j + \frac{1}{2}rank(K_j) \quad \text{and} \quad b_j' = b_j + \frac{1}{2}\beta_j' K_j \beta_j$$

and updating can be done by simple Gibbs steps, drawing random numbers directly from the inverse gamma densities.

For model comparison, we suggest to use the Deviance Information Criterion (DIC) developed in Spiegelhalter et al. (2002). Let $\theta = (\beta, \gamma)$ denote the vector of all parameters of interest defining the predictor η. Then the DIC is based on the Bayesian deviance

$$D(\theta) = -2logL(\theta) + 2logL(data), \tag{10}$$

where $L(\theta)$ is the likelihood (7) or (8) of a specific survival or event history model, given $\eta = \eta(\theta)$, and $L(data)$ is taken as the likelihood of some saturated model. For a Cox model, where the baseline hazard $\lambda_0(t)$ is a nuisance parameter, only the effects γ in (1) are the parameters of interest. In this case, a saturated term for $D(\theta)$ can be defined as in Fleming and Harrington (1991, p.168). In our models, however, the baseline hazard and other nonparametric functions are parameters of interest. In this case it is not clear what a saturated model should be. Therefore, we drop the second term in (10), which is of no relevance for model comparison. Based on the deviance $D(\theta)$, Spiegelhalter et al. (2002) define

$$p_D = \overline{D(\theta)} - D(\overline{\theta})$$

as the effective number of parameters in the model. Here $\overline{D(\theta)}$ is the posterior mean of the deviance, and $D(\overline{\theta})$ is $D(\theta)$ evaluated at $\theta = \overline{\theta}$, the posterior mean of the parameters. The Deviance Information Criterion is then defined as

$$DIC = \overline{D(\theta)} + p_D.$$

Properties of the DIC have been mainly explored for exponential family regression. With the piecewise exponential model, which leads to a Poisson–type likelihood, we are still within this framework, but we leave it with more general P–spline priors for functions. To be on the safe side, we recommend to run models always with piecewise constant functions and alternatively with B–splines of higher degree, and to compare results. More empirical experience with simulated and real data has to be gained before making a general recommendation.

4 Simulation studies

We investigate performance through simulation studies for three related survival models. For the basic Model 1, life times T_i, $i = 1, ..., 1000$, were generated according to the hazard model

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \exp(\gamma d_i + f_2(x_i)) \\ &= \exp\left(log(3t^2) - 0.3 d_i + sin(x_i)\right), \end{aligned} \tag{11}$$

In this model, the baseline hazard rate $\lambda_0(t)$ is set to $3t^2$, which is a Weibull hazard rate, so that $f_0(t) = log(3t^2)$. The covariate d is binary, with the d_i´s randomly drawn from a Bernoulli $B(1; 0.5)$ distribution, and the covariate x is continuous, with the x_i´s randomly drawn from a uniform $U[-3, 3]$ distribution. Censoring variables C_i, $i = 1, ..., 1000$, were generated as i. i. d. draws from a uniform $U[0, 5]$ distribution, resulting in a proportion of 15–20 percent of censored observations.

Model 2 extends Model 1 to a model with time–varying effect $f_1(t) = log\frac{5t^2}{3}$ of the covariate d, i.e.,

$$\lambda_i(t) = \exp\left(log(3t^2) + log\left(\frac{5t^2}{3}\right) d_i + sin(x_i)\right).$$

Then the hazard rate is of Weibull form for $d_i = 0$ as well as for $d_i = 1$, and life time T_i can be generated by drawing randomly from corresponding Weibull distributions. In Model 3, we add random effects b_g, i.i.d. $\sim N(0; 0.5^2)$, $g = 1, ..., 10$, to the basic Model 1, leading to hazard rates

$$\lambda_i(t) = \exp\left(log(3t^2) - 0.3 d_i + sin(x_i) + b_{gi}\right).$$

Covariates d_i and x_i, $i = 1, ..., 1000$, were generated as before.

Keeping the predictor fixed 100 replications $\{T_i^{(r)}, C_i^{(r)}, i = 1, ..., 1000\}$ resp. $\{(t_i^{(r)}, \delta_i^{(r)}), i = 1, ..., 1000\}$, $r = 1, ..., 100$ of censored survival times were generated for each simulation study.

The log–baseline hazard $f_0(t)$ and the time–varying effect $f_1(t)$ were modelled by second order random walk priors, corresponding to a piecewise exponential model (with grid length $\triangle = 0.1$) and – alternatively – as cubic P–splines, with 20 knots. A cubic P–spline prior with 20 knots was chosen for $f_2(x) = sin(x)$. Hyperparameters of inverse gamma priors for variance components were set to $a = b = 0.001$, the standard choice.

For each replication $r = 1, ..., 100$, we computed the mean square errors

$$MSE_r(f_k) = \frac{1}{1000}\sum_{i=1}^{1000}(\widehat{f}_k^{(r)}(t_i^{(r)}) - f_k(t_i^{(r)}))^2, \quad k = 0, 1$$

for the log–baseline hazard $f_0(t)$ and – for Model 2 – the time–varying effect $f_1(t)$, and

$$MSE_r(f_2) = \frac{1}{1000}\sum_{i=1}^{1000}(\widehat{f}_2^{(r)}(x_i) - f_2(x_i))^2$$

for $f_2(x) = sin(x)$, where $\widehat{f}_k^{(r)}$, $k = 0, 1, 2$, are posterior mean estimates for simulation run r.

For Model 1 and 3, the $MSE(\gamma)$ was computed in the usual way.

Table 1 summarizes the results, displaying $\overline{MSE} = \frac{(\sum_{r=1}^{100} MSE_r)}{100}$ as well as $min_r MSE_r$ and $max_r MSE_r$ in each cell.

As was to be expected, the P–spline model has smaller MSE´s for f_0 and f_1 when compared to the piecewise exponential model, although the difference is smaller for $MSE(f_1)$ of the time–varying effect f_1. Interestingly, the MSE´s for $\gamma = -0.3$ and $f_2(x)$ are more or less unaffected by the choice of the smoothness prior for time–varying functions $f_0(t)$ and $f_1(t)$.

Figures 3–5 in the appendix show some selected posterior mean estimates together with the true function for the piecewise exponential model and the P–spline model.

5 Application: Long Term Care Insurance

As an illustration, we analyze data on survival time after entering long term care insurance (LTC) from a German private insurance company. The data was recorded between April 1, 1995, when compulsory LTC insurance was introduced by the German government, and December 31, 1998. It contains information on 5603 recipients of benefits from LCT insurance. This data set has already been analyzed by Czado and Rudolph (2002), and more details on the data set are given

	Model 1	Model 2	Model 3
piecewise exponential model	$\overline{MSE}(f_0) = 0.176$	$\overline{MSE}(f_0) = 0.183$	$\overline{MSE}(f_0) = 0.189$
	$minMSE(f_0) = 0.044$	$minMSE(f_0) = 0.053$	$minMSE(f_0) = 0.029$
	$maxMSE(f_0) = 0.524$	$maxMSE(f_0) = 0.450$	$maxMSE(f_0) = 0.441$
	$\overline{MSE}(\gamma) = 0.005$	$\overline{MSE}(f_1) = 0.359$	$\overline{MSE}(\gamma) = 0.005$
	$minMSE(\gamma) = 1.5e^{-7}$	$minMSE(f_1) = 0.095$	$minMSE(\gamma) = 8.5e^{-10}$
	$maxMSE(\gamma) = 0.043$	$maxMSE(f_1) = 0.805$	$maxMSE(\gamma) = 0.045$
	$\overline{MSE}(f_2) = 0.007$	$\overline{MSE}(f_2) = 0.007$	$\overline{MSE}(f_2) = 0.008$
	$minMSE(f_2) = 0.001$	$minMSE(f_2) = 0.0004$	$minMSE(f_2) = 0.001$
	$maxMSE(f_2) = 0.025$	$maxMSE(f_2) = 0.025$	$maxMSE(f_2) = 0.026$
P–spline model	$\overline{MSE}(f_0) = 0.143$	$\overline{MSE}(f_0) = 0.151$	$\overline{MSE}(f_0) = 0.156$
	$minMSE(f_0) = 0.032$	$minMSE(f_0) = 0.037$	$minMSE(f_0) = 0.017$
	$maxMSE(f_0) = 0.462$	$maxMSE(f_0) = 0.391$	$maxMSE(f_0) = 0.408$
	$\overline{MSE}(\gamma) = 0.005$	$\overline{MSE}(f_1) = 0.341$	$\overline{MSE}(\gamma) = 0.005$
	$minMSE(\gamma) = 2e^{-8}$	$minMSE(f_1) = 0.071$	$minMSE(\gamma) = 4.5e^{-9}$
	$maxMSE(\gamma) = 0.043$	$maxMSE(f_1) = 0.753$	$maxMSE(\gamma) = 0.045$
	$\overline{MSE}(f_2) = 0.007$	$\overline{MSE}(f_2) = 0.007$	$\overline{MSE}(f_2) = 0.008$
	$minMSE(f_2) = 0.001$	$minMSE(f_2) = 0.0004$	$minMSE(f_2) = 0.001$
	$maxMSE(f_2) = 0.027$	$maxMSE(f_2) = 0.025$	$maxMSE(f_2) = 0.025$

Table 1: Summary of MSE´s

there. In a first step, they analyzed the data with a conventional Cox model. After careful model diagnosis, they extended it to a model with time–varying effects, which were modelled through 0-1 step functions. Our analysis is based on their final model (3.2), - CR model for short - with results displayed in their Table 5. The following covariates are included:

age (in years),

sex (1=female, 0=male),

and the time–dependent covariates

$$\begin{aligned}
\text{nh(t)} &= \begin{cases} 1 & \text{care in a nursing home at time t} \\ 0 & \text{care at home} \end{cases} \\
\text{level 2(t)} &= \begin{cases} 1 & \text{care at level 2 at time t} \\ 0 & \text{else}\,, \end{cases} \\
\text{level 3(t)} &= \begin{cases} 1 & \text{care at level 3 at time t} \\ 0 & \text{else}\,, \end{cases}
\end{aligned}$$

with level 1(t) as the reference category. The three levels of care (and benefits) are defined as follows:

Level 1: The LTC–claimant needs at least 90 minutes help per day to manage his/her activities of daily living (like going to bed, washing, eating),

Level 2: The LTC–claimant needs at least 180 minutes help per day to manage his/her activities of daily living,

Level 3: The LTC–claimant needs at least 300 minutes help per day to manage his/her activities of daily living.

About 60 per cent of the observations are right censored.

We apply a Bayesian multiplicative hazard rate model $\lambda(t) = \exp(\eta(t))$ with predictor

$$\begin{aligned}
\eta(t) &= f_0(t) + f_{age}(age) + f_{sex}(age)sex + f_{l2}(t)level2(t) \\
&\quad + f_{l3}(t)level3(t) + \gamma_1 nh(t) + \gamma_2(sex \cdot nh(t)) \\
&\quad + \gamma_3(level2(t) \cdot nh(t)) + \gamma_4(level3(t) \cdot nh(t))
\end{aligned}$$

	mean	std. dev.
nh	0.775	0.170
sex·nh	-0.376	0.108
level2·nh	-0.410	0.181
level3·nh	-0.722	0.178

Table 2: Posterior mean estimates and standard deviations for the fixed effects

Compared to the CR model we omitted interactions between age and survival time t and between age and $nh(t)$, which had very small effects. The log–baseline effect $f_0(t)$, the main effect $f_{age}(age)$ of age, the age–dependent effect $f_{sex}(age)$ of sex, and the time–varying effects $f_{l2}(t)$, $f_{l3}(t)$ of levels 2 and 3 were estimated nonparametrically with cubic P–splines with 20 knots.

Results for fixed effects are given in Table 2. Care in a nursing home seems to increase the hazard rate; a plausible reason for this effect is that individuals in a nursing home are generally less healthy than individuals who still receive care at home. This main effect decreases for females and for individuals who receive more care. Figure 2 shows the main effect f_{age} and its interaction with sex. The main effect of age increases almost linearly, while the interaction with sex is quite small. The latter is comparable with the fixed effect interaction of CR only for age between 50 and 90 years.

The log–baseline hazard rate $f_0(t)$ and the effect $f_{l2}(t)$ of level2(t) in Figure 1 are more or less time–constant. Note that $f_0(t)$ is centered around zero, while $f_{l2}(t) \approx const. = 0.9$. The time–variation in the effect of level2(t) of the CR model cannot be detected. The increased hazard for level3(t) for smaller t corresponds to a similar finding of the initial CR. Later on, this effect becomes approximately constant (ca. 1.5). A possible interpretation is that this effect is caused by individuals which are already in a bad health state and therefore need level 3 care immediately at the beginning of LCT.

Conclusions

Nonparametric Bayesian hazard rate modeling with P–splines offers a very flexible tool for simul-

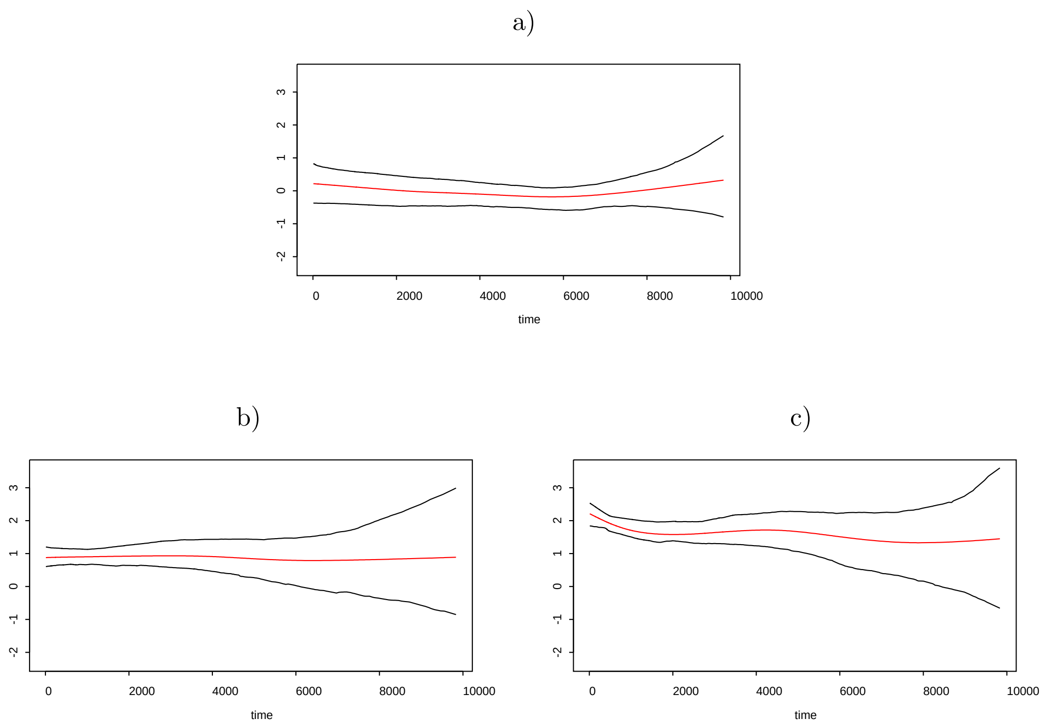

Figure 1: Posterior mean estimates and 95% credible intervals for time–dependent effects: a) log–baseline effect, b) time–varying effect of level2, c) time–varying effect of level3

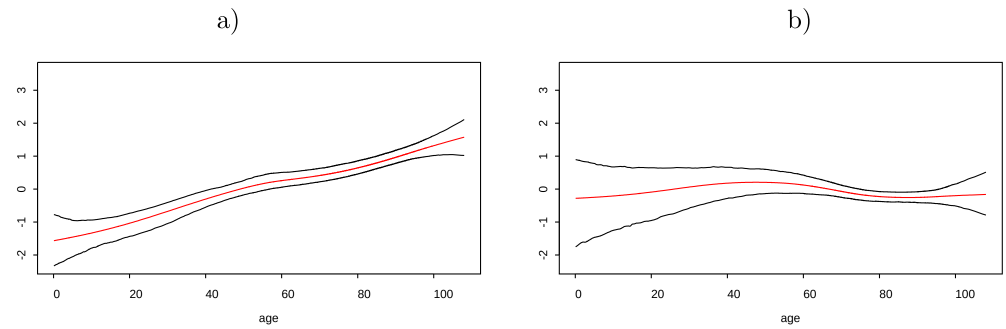

Figure 2: Posterior mean estimates and 95% credible intervals for nonparametric effects: a) main effect of age, b) effect of sex varying over age

taneous analysis of baseline hazard rates, time–varying effects and nonlinear effects of metrical covariates together with usual linear effects of further covariates.

Further research is necessary to investigate DIC as a model comparison tool, in particular for more complex models of event history analysis. From a computational point of view, further algorithmic improvements in numerical evaluations of the likelihood are essential to speed up computation times.

Acknowledgement

We thank Andreas Brezger and Stefan Lang for helpful discussions and computational assistance with *BayesX* and Claudia Czado, Florian Rudolph and Susanne Gschloessl for helpful comments on the LTC data. Financial support of the German Science Foundation DFG, Sonderforschungsbereich 386 "Statistische Analyse Diskreter Strukturen" is gratefully acknowledged.

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.

Arjas, E. and Gasbarra, D., 1994: Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* 4, 505–524.

Brezger, A., Kneib, T. and Lang, S., 2002: BayesX - Software for Bayesian Inference based on Markov Chain Monte Carlo simulation techniques, open domain software available from `http://www.stat.uni-muenchen.de/~lang/`.

Brezger, A. and Lang, S., 2003: Generalized structured additive regression based on Bayesian P-splines. Discussion Paper 321, SFB 386.

Cai, T., Hyndman, R., and Wand, M., 2000: Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics*, 11, 784–798.

Czado, C. and Rudolph, F., 2002: Application of survival analysis methods to long term care insurance. *Mathematics and Economics*, to appear.

Eilers, P.H.C. and Marx, B.D., 1996: Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statistical Science*, 11 (2), 89–121.

Fahrmeir, L., 1994: Dynamic Modelling and Penalized Likelihood Estimation for Discrete Time Survival Data. *Biometrika*, 81, 317–330.

Fahrmeir, L. and Klinger, A., 1998: A nonparametric multiplicative hazard model for event history analysis, *Biomerika*, 85(3), 581–592.

Fahrmeir, L. and Lang, S., 2001: Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Journal of the Royal Statistical Society C (Appl. Stat.)*, 50, 201-220.

Fahrmeir, L. and Tutz, G., 2001: *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer–Verlag, New York.

Fahrmeir, L. and Wagenpfeil, S., 1996: Smoothing Hazard Functions and Time-Varying Effects in Discrete Duration and Competing Risk Models. *Journal of the American Statistical Association*, 91 (436), 1584–1594.

Fleming, T. R. and Harrington, D. P., 1991: *Counting Processes and Survival Analysis*, John Wiley, New–York.

Gamerman, D., 1991: Dynamic Bayesian models for survival data. *Applied Statistics*, 40, 63–79.

Gamerman, D., 1997: Efficient Sampling from the Posterior Distribution in Generalized Linear Models. *Statistics and Computing*, 7, 57–68.

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

Ibrahim, J., Chen, M. H. and Sinha, D., 2001: *Bayesian Survial Analysis*. Springer Series in Statistics, New York.

Knorr–Held, L., 1999: Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics*, 26, 129–144.

Lang, S. and Brezger, A., 2002: Bayesian P-splines. *Journal of Computational and Graphical Statistics*, to appear.

Marx, B. D. and Eilers, P., 1998: Direct Generalized Additive Modeling with Penalized Likelihood. *Computational Statistics and Data Analysis*, 28, 193–209.

Sinha, D., 1993: Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 92, 1195–1212.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., 2002: Bayesian measures of model complexity and fit (with discussion and rejoinder). *J. Roy. Statist. Soc., Ser. B*, 64, 583–639.

Appendix

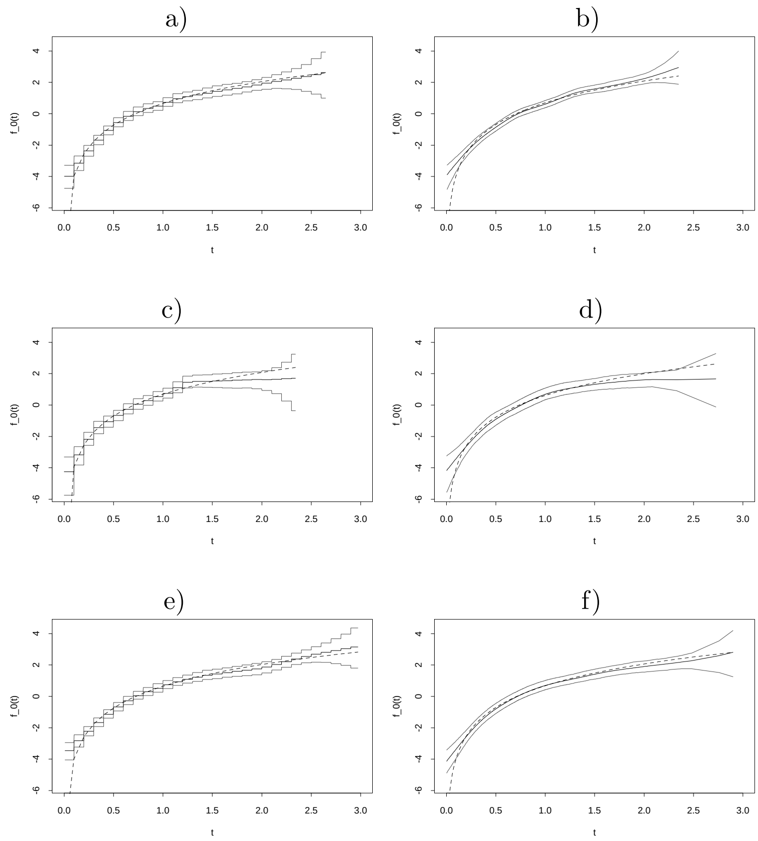

Figure 3: (log–)baseline effects for the various model specifications; displayed are posterior mean estimates and 95% credible intervals of run r, with r chosen such that MSE_r is the median of $MSE_1, \ldots, MSE_{100}$ (solid lines), and the true function (dashed line).

a) p.e.m., Model 1, r=15, MSE=0.159

b) P–spline model, Model 1, r=11, MSE=0.119

c) p.e.m., Model 2, r=5, MSE=0.176

d) P–spline model, Model 2, r=3, MSE=0.141

e) p.e.m., Model 3, r=100, MSE=0.173

f) P–spline model, Model 3, r=81, MSE=0.138

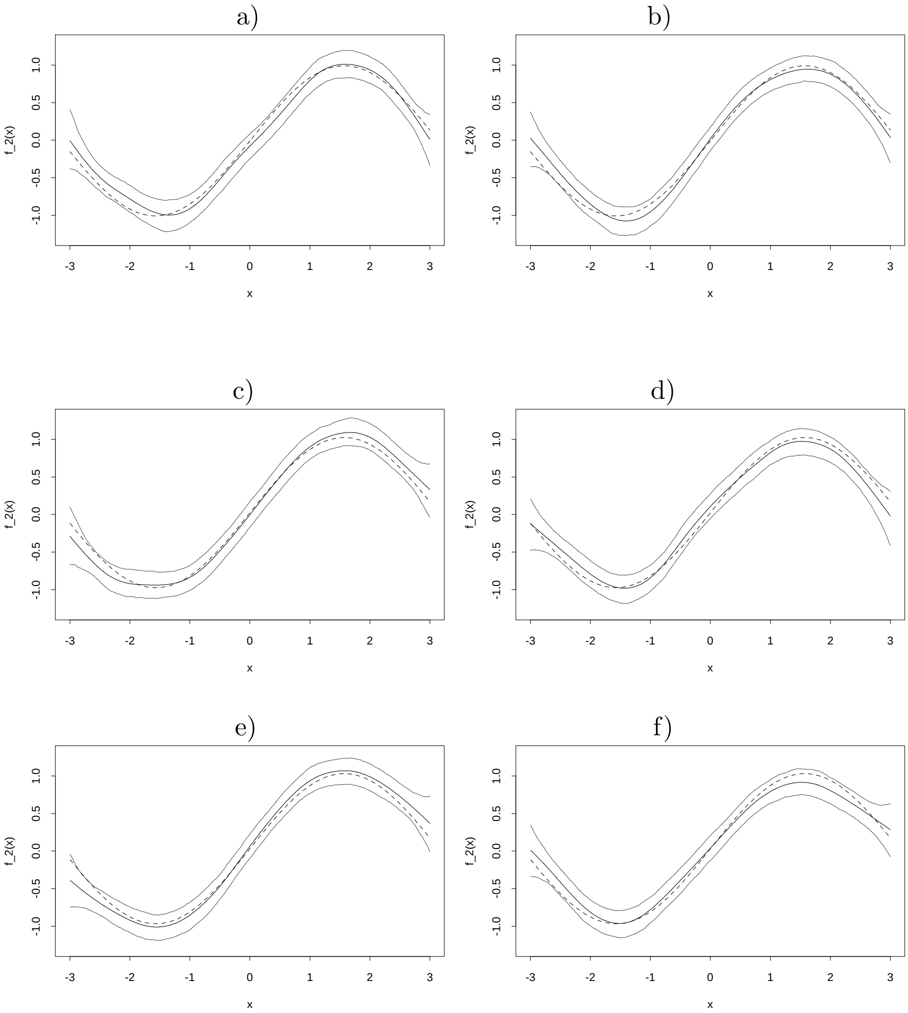

Figure 4: Nonparametric effects for the various model specifications; displayed are posterior mean estimates and 95% credible intervals of run r, with r chosen such that MSE_r is the median of $MSE_1, \ldots, MSE_{100}$ (solid lines), and the true function (dashed line).

a) p.e.m., Model 1, r=89, MSE=0.006

b) P–spline model, Model 1, r=93, MSE=0.006

c) p.e.m., Model 2, r=50, MSE=0.006

d) P–spline model, Model 2, r=61, MSE=0.006

e) p.e.m., Model 3, r=66, MSE=0.007

f) P–spline model, Model 3, r=34, MSE=0.007

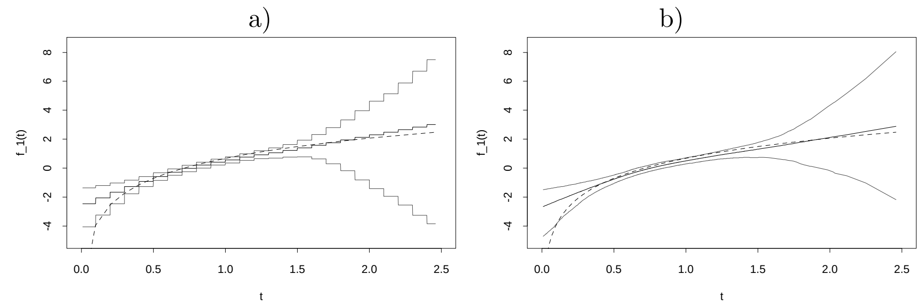

Figure 5: Time–varying effects of d for Model 2; displayed are posterior mean estimates and 95% credible intervals of run r, with r chosen such that MSE_r is the median of $MSE_1, \ldots, MSE_{100}$ (solid lines), and the true function (dashed line).

a) p.e.m., r=34, MSE=0.349

b) P–spline model, r=34, MSE=0.331