Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2006-13

Auxiliary mixture sampling for Dynamic Survival Models

Helga Wagner

June 2006

Abstract

Dynamic survival models are a useful extension of the popular Cox model as the effects of explanatory variables are allowed to change over time. The baseline log-hazard as well as covariate effects are assumed to be piecewise constant with a correlated prior process. These models can be estimated in a Bayesian framework by Markov chain Monte Carlo methods. We propose a new auxiliary mixture sampler based on data augmentation, which avoids Metropolis-Hastings steps and needs no further tuning. The sampler can be applied to more general models, e.g. models including frailty or spatial effects and for data observed under censoring schemes different from right censoring.

Some key words: Bayesian survival analysis, data augmentation, Gibbs sampling, piecewise exponential

1 Introduction

For survival data the predominately used model to assess the effect of explanatory variables is the Cox model, where the hazard function is modeled as

$$\lambda(t \,|\, \mathbf{z}) = \lambda(t)\ \exp(\mathbf{z}'\boldsymbol{\beta}).$$

Here $\mathbf{z}' = (z_1, \ldots, z_k)$ is a vector of covariates, $\boldsymbol{\beta}$ is the vector of unknown parameters, $\eta = \mathbf{z}'\boldsymbol{\beta}$ and is the linear predictor. The model is semi-parametric insofar as the baseline hazard function is not specified. The Cox model relies on the assumption that the hazard ratio for two individuals with covariate values $\mathbf{z}$ and $\mathbf{z}^*$ respectively is constant over time and only depends on the difference between their linear predictors

$$\frac{\lambda(t \,|\, \mathbf{z})}{\lambda(t \,|\, \mathbf{z}^*)} = \lambda(t)\ \exp\big((\mathbf{z} - \mathbf{z}^*)'\boldsymbol{\beta}\big).$$

This is an assumption which is not necessarily true in applied situations where the effect of covariates may vary over time, e.g. a certain treatment may have a positive short-term effect which vanishes in the long run. Thus models where the covariate effects are allowed to change over time are often more appropriate. A common approach to model time varying effects is by piecewise constant functions, as these are flexible enough to capture any shape of the baseline hazard or covariate effects, see Verweij and van Houwelingen (1995) for a frequentist and Gamerman (1991) for a Bayesian approach. In the last years different non- and semi-parametric Bayesian survival models have been suggested, for an overview see Ibrahim et al. (2001) and Sinha and Dey (1997). Most of these models use correlated prior processes for the coefficients of the piecewise functions, thus defining a stochastic evolution over time, where coefficients of adjacent intervals are expected to be close.

Models using correlated gamma increments are considered in Arjas and Gasbarra (1994) who use a first order autoregressive jump process for the hazard rate, not including covariates and in Nieto-Barajas and Walker (2002) who propose a dependent gamma process defined by two latent processes. Sinha et al. (1999) use an independent Gamma process for the baseline hazard and a correlated normal prior for the covariate effects. In the dynamic survival model, introduced by Gamerman (1991) the baseline log-hazard as well as the effects of covariates are modeled by

piecewise constant functions with a correlated gaussian prior process. Closely related is the model of Fahrmeir and Hennerfeind (2003) and Hennerfeind et al. (2006) who use penalized splines for modeling baseline hazard and covariate effects, with a correlated normal prior process for their coefficients. Gamerman's dynamic survival model can be seen as a special case of their model with penalized splines of degree zero.

Bayesian estimation of these models usually is accomplished by MCMC methods which however require a Metropolis-Hastings-Algorithm, see Gamerman (1991), Hemming and Shaw (2005) and Hennerfeind et al. (2006). Here we present a new approach for the dynamic piecewise exponential model of Gamerman (1991), using data augmentation. We develop an auxiliary mixture sampler which involves only draws from standard densities and needs no further tuning. Auxiliary mixture sampling was introduced for Bayesian analysis of stochastic volatility models by Shephard (1994) and was applied in this context to different models by a couple of authors (Kim et al., 1998; Chib et al., 2002; Omori et al., 2006). Frühwirth-Schnatter and Wagner (2005, 2006) introduced auxiliary mixture sampling for Bayesian analysis of parameter-driven models for count data based on the poisson distribution and Frühwirth-Schnatter and Frühwirth (2006) apply auxiliary mixture sampling to binary and multinomial logit models.

To extend auxiliary mixture sampling to dynamic survival models in a first data augmentation step auxiliary survival times are introduced as missing data. Thus a representation of the original model as a linear state space model with non-normal errors is achieved. The error distribution which turns out to be a type one extreme value distribution is then approximated by a mixture of normal components, as in Frühwirth-Schnatter and Wagner (2005, 2006) and Frühwirth-Schnatter and Frühwirth (2006). By introducing the component indicators of this normal mixture as a second sequence of missing data, a Gibbs sampling type algorithm is obtained which allows multi-move sampling of time varying effects.

The paper is organized as follows. The next section presents the model specification. Data augmentation steps and the auxiliary mixture sampling scheme for estimating model parameters are described in Section 3. In Section 4 we illustrate the method on simulated data and two well-known real data sets. In Section 5 we outline application of the sampler for models extending the linear predictor to include frailty effects and describe how the necessary steps to deal with data subject to interval censoring. The paper concludes with some discussion in Section 6.

2 The Dynamic Survival Model

2.1 Model specification

Survival data usually are subject to right-censoring. We assume that each individual i, $i = 1, \ldots, n$, has a survival time t_i and a censoring time c_i which are independent random variables. Observed data consist of the observation time $y_i = \min(t_i, c_i)$, a failure indicator δ_i and a vector of K covariates $(z_{i1}, \ldots, z_{iK})$. Extending Cox's proportional hazards model, not only the baseline hazard but also covariate effects

are assumed to be a functions of time, i.e.

$$\lambda(t|\mathbf{z}_i) = \exp\left(\beta_0(t) + \sum_{k=1}^{K} z_{ik}\beta_k(t)\right)$$

The dynamic survival models, as proposed by Gamerman (1991), is a piecewise exponential model for lifetimes, with correlated prior processes for the baseline log-hazard as well as the covariate effects. It is based on a given partition of the time axis $\mathcal{S} = \{s_0, s_1, \ldots, s_J\}$, $0 \leq s_0 < s_1 < \cdots < s_J$. These division points form J intervals $(s_0, s_1], \ldots, (s_{J-1}, s_J]$. The baseline log-hazard $\beta_0(t)$ as well as the covariate effects $\beta_k(t)$ are defined for $k = 0, \ldots, K$ via the piecewise constant functions

$$\beta_k(t) = \beta_{kj}, \qquad \text{for } t \in I_j = (s_{j-1}, s_j].$$

To model stochastic evolution each β_k is assumed to follow a random walk, i.e.

$$\beta_{kj} = \beta_{k,j-1} + w_{kj} \qquad w_{kj} \sim \text{Normal}\,(0, \theta_k)\,. \tag{1}$$

In this model an evolution variance $\theta_0 = 0$ would imply a constant baseline hazard, and $\theta_k = 0$ would imply a constant effect of covariate z_k. If all evolution variances $\theta_k, k = 0, \ldots, K$ were zero, the model would simply be an exponential regression model where

$$\lambda(t|\mathbf{z}) = \exp\left(\mathbf{z}'\boldsymbol{\beta}\right),$$

with $\mathbf{z} = (1, z_1, \ldots, z_K)$ being the vector of covariates and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)$.

In the sense of Hennerfeind et al. (2006) in this dynamic survival model baseline log-hazard and time varying effects are modeled as a linear combination of J B-spline basis functions of degree zero with knots $\{s_0, s_1, \ldots, s_J\}$.

$$\beta_k(t) = \sum_{j=1}^{J} \beta_{kj} 1_{(s_{j-1}, s_j]}(t)$$

The random walk priors are smoothness priors penalizing abrupt jumps $\beta_{k,j} - \beta_{k,j-1}$.

In the following we consider a slight modification of Gamerman's dynamic survival model

$$\lambda(t|\mathbf{z}_i; t \in I_j) = \exp(\eta_{ij})$$

by defining the linear predictor through

$$\eta_{ij} = (\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_j \tag{2}$$

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j \qquad \boldsymbol{\omega}_j \sim \text{Normal}\,(\mathbf{0}, \text{diag}(\theta_0, \ldots, \theta_K)) \tag{3}$$

$\mathbf{z}^f$ is a vector of covariates with fixed effects $\boldsymbol{\alpha}$, whereas $\mathbf{z}^v = (1, z_1^v, \ldots, z_K^v)'$ is the vector of covariates with time-varying effects $\boldsymbol{\beta}$. $\boldsymbol{\beta}_j = (\beta_{0j}, \ldots, \beta_{Kj})$ denotes these effects in interval I_j, β_{0j} being the baseline log-hazard in I_j.

The division points $\{s_0, s_1, \ldots, s_J\}$ should be chosen fine enough to capture the shape of baseline hazard and time-varying effects. Usually s_J is taken to be the last observed failure or censoring time. Whereas Hennerfeind et al. (2006) use equally spaced time points, Gamerman (1991) and Hemming and Shaw (2005) also consider a data-dependent division where the division points are the observed death times.

2.2 Priors

For a fully Bayesian specification of the model, priors for the fixed effects $\boldsymbol{\alpha}$, the starting values of the random walks $\beta_{00}, \dots \beta_{0k}$ and the process variances $\boldsymbol{\theta} = (\theta_0, \dots, \theta_k)$ have to be specified. For the fixed effects either normal priors $\boldsymbol{\alpha} \sim \text{Normal}\,(\mathbf{a}_0, \mathbf{A}_0)$ or diffuse priors $\boldsymbol{\alpha} \propto const$ can be assumed.

For the starting points of the random walks $\beta_{k0}, k = 0, \dots, K$ we assume independent normal priors

$$\beta_{k0} \sim \text{Normal}\,(0, B_k)$$

and for their variances θ_k independent inverse Gamma priors

$$\theta_k \sim \text{InvGamma}\,(c_{k0}, C_{k0})\,, \quad k = 0, \dots, K.$$

In the applications we will use proper but uninformative priors for $\beta_{k0}, k = 0, \dots, K$ as well as for the evolution variances. Hennerfeind et al. (2006) give conditions for propriety of posteriors under partially improper priors on the parameter vectors $\beta_0 = (\beta_{00}, \dots, \beta_{K0})$.

Note that the processes $\beta_0, \dots, \beta_K$ are independent a priori and

$$p(\beta_{k0}, \dots, \beta_{kJ}) = p(\beta_{k0}) p(\beta_{k1}, \dots, \beta_{kJ} | \beta_{k0}).$$

Given the starting value β_{k0}, $(\beta_{k1}, \dots, \beta_{kJ})$ has a multivariate normal distribution

$$p(\beta_{k1}, \dots, \beta_{kJ} | \beta_{k0}, \theta_k) = \text{Normal}\,(\beta_{k0} \cdot \mathbf{1}, \theta_k \mathbf{C}\mathbf{C}')$$

where $\mathbf{1}$ is a column vector of J ones and $\mathbf{C}$ is the random walk generating matrix

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & \cdot & \cdots & 0 \\ 1 & 1 & & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

3 Auxiliary Mixture Sampling

In this section we develop a Markov Chain Monte Carlo sampling scheme for the dynamic survival model which relies on two steps of data augmentation. The goal of the first data augmentation step is to represent the original model as a state space model for complete exponentially distributed survival times. For these models, auxiliary mixture sampling as in Frühwirth-Schnatter and Wagner (2005) and Frühwirth-Schnatter and Wagner (2006) is feasible. There it is shown how by mixture approximation for the negative of the logarithm of an Exponential (1)-distribution, the original model can be represented as a partial Gaussian model.

We start with discussing auxiliary mixture sampling for a regression model of exponentially distributed survival times before we turn to the dynamic survival model.

3.1 Auxiliary mixture sampling for the exponential regression model

A simple model for survival data taking into account effects of covariates is the exponential regression model

$$t_i \sim \text{Exponential}\,(\lambda_i)\,, \quad \lambda_i = \exp(\mathbf{z}_i'\boldsymbol{\alpha}) \tag{4}$$

where $\mathbf{z}_i$ is the row vector of covariates for individual i and $\boldsymbol{\alpha}$ is the vector of unknown parameters.

Usually survival times are not completely observed but subject to right censoring. Let $(\mathbf{y}, \boldsymbol{\delta}) = (y_1, \delta_1), \ldots, (y_n, \delta_n)$ denote the data, where the observation time y_i is a survival time if $\delta_i = 1$ and a censoring time if $\delta_i = 0$. For censored observations the exact lifetime t_i is unobserved. (y_i, δ_i) however contains partial information on the lifetimes, namely $t_i > y_i$. Our first data augmentation step uses this information to generate complete auxiliary survival times.

If y_i is a censored observation, i.e. $\delta_i = 0$, then $y_i = c_i = \min(t_i, c_i)$ where $t_i \sim \text{Exponential}\,(\lambda_i)$. Conditionally on the event $\{t_i > c_i\}$ the – unobserved – residual survival time $\xi_i = t_i - c_i$ is distributed as

$$P(t_i - c_i \geq t | t_i > c_i) = \frac{e^{-\lambda_i(t+c_i)}}{e^{-\lambda_i c_i}} = e^{-\lambda_i t}.$$

Thus in a first data augmentation step the unobserved residual survival time ξ_i can be generated as an $\text{Exponential}\,(\lambda_i)$ distributed random variable. The (partly unobserved) auxiliary survival times τ_i

$$\tau_i = \begin{cases} y_i & \text{if} \quad \delta_i = 1 \\ y_i + \xi_i & \text{if} \quad \delta_i = 0 \end{cases}$$

are complete observations and follow the exponential regression model defined in equation (4). By taking logarithms and multiplying by minus one, this nonlinear and non-normal model is transformed into the linear model

$$-\ln(\tau_i | \mathbf{z}_i) = \mathbf{z}_i'\boldsymbol{\alpha} + \varepsilon_i$$

where the error term ε_i has a type I extreme value distribution. To obtain a model that is conditionally Gaussian, this non-normal density can be approximated by a mixture of ten normal components, as in Frühwirth-Schnatter and Wagner (2006):

$$p_\varepsilon(\varepsilon) = \exp(-\varepsilon - e^{-\varepsilon}) \approx q_{R,\varepsilon}(\varepsilon) = \sum_{r=1}^{10} w_r f_{\text{N}}(\varepsilon; m_r, s_r^2), \tag{5}$$

where m_r and s_r are the mean and the variance of component r and $f_{\text{N}}(\varepsilon; m_r, s_r^2)$ denotes the Gaussian density, see Table 1. For more details on how this mixture approximation is obtained see Frühwirth-Schnatter and Frühwirth (2006).

The second step of our data augmentation scheme for each ε_i introduces the latent component indicator r_i as missing data. Let $\mathcal{R} = \{r_i, i = 1, \ldots, n\}$. Then,

Table 1: Ten component normal mixture approximation for the density of $-\ln \xi$, where $\xi \sim$ Exponential (1).

r	1	2	3	4	5	6	7	8	9	10
w_r	0.00397	0.0396	0.168	0.147	0.125	0.101	0.104	0.116	0.107	0.088
m_r	5.09	3.29	1.82	1.24	0.764	0.391	0.0431	-0.306	-0.673	-1.06
s_r^2	4.50	2.02	1.10	0.422	0.198	0.107	0.0778	0.0766	0.0947	0.146

conditional on $\boldsymbol{\tau} = \{\tau_i, i = 1, \ldots, n\}$ and $\mathcal{R}$, the non-normal, nonlinear model (4) reduces to a linear, Gaussian model

$$-\ln \tau_i | \boldsymbol{\alpha}, r_i = \mathbf{z}_i' \boldsymbol{\alpha} + m_{r_i} + \varepsilon_i, \quad \varepsilon_i | r_i \sim \text{Normal}\left(0, s_{r_i}^2\right).$$

Consequently, the conditional posterior $p\big(\boldsymbol{\alpha} | \boldsymbol{\tau}, \mathcal{R}, (\mathbf{y}, \boldsymbol{\delta})\big)$ is given by

$$p(\boldsymbol{\alpha} | \boldsymbol{\tau}, \mathcal{R}, (\mathbf{y}, \boldsymbol{\delta})) \propto p(\boldsymbol{\alpha}) \prod_{i=1}^{n} \text{Normal}\left(-\ln \tau_i; \mathbf{z}_i' \boldsymbol{\alpha} + m_{r_i}, s_{r_i}^2\right), \tag{6}$$

which is proportional to a multivariate normal density.

Thus for the exponential regression model containing censored observations the following two-block Gibbs sampler can be applied. After selecting starting values for $\boldsymbol{\alpha}$ and $\mathcal{R}$ repeat the following steps:

1. Sample auxiliary survival times for censored observations and introduce the component indicator as a second sequence of missing data.
   (a) Sample the auxiliary survival time τ_i of a censored observation y_i as $\tau_i = y_i + \xi_i$, where $\xi_i \sim$ Exponential (λ_i).
   (b) Sample the component indicators r_i.
2. Sample $\boldsymbol{\alpha}$ from a multivariate normal density.

The component indicators are sampled from a discrete distribution with 10 categories

$$\Pr\{r_i = k | \tau_i, \boldsymbol{\alpha}\} \propto p(\tau_i | r_i = k, \boldsymbol{\alpha}) w_k,$$

where

$$p(\tau_i | r_i = k, \boldsymbol{\alpha}) \propto \frac{1}{s_k} \exp\left\{-\frac{1}{2}\left(\frac{-\ln \tau_i - \ln \lambda_i - m_k}{s_k}\right)^2\right\},$$

for details see Frühwirth-Schnatter and Wagner (2006).

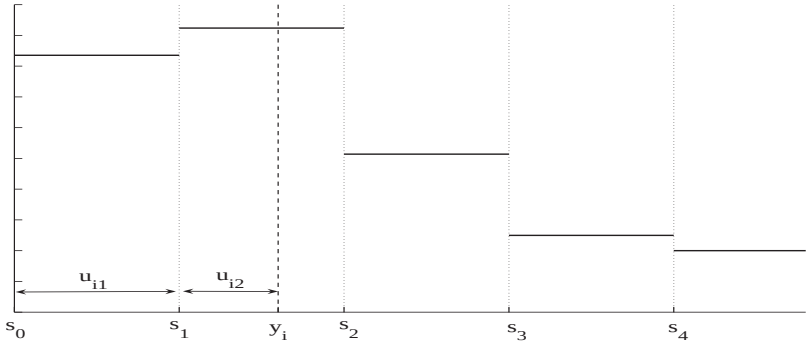

Figure 1: Hazard in the piecewise exponential model

3.2 Auxiliary mixture sampling for the dynamic survival model

Also for dynamic survival models the goal of the first data augmentation step is to represent the original model for the hazard as a linear model for suitably chosen auxiliary variables. Due to the piecewise constant structure of baseline log-hazard and time-varying effects each individual experiences not just one, but several hazards over the observation time – one in each time interval I_l. Therefore, for each of these hazards, a separate auxiliary survival time is introduced. This data augmentation step can be motivated from a factorization of the complete data likelihood over individual contributions and time intervals, which is discussed in the next subsection.

3.2.1 Likelihood factorization

In the dynamic survival model defined in Section 2 the hazard $\lambda(t|\mathbf{z}_i)$ for each individual $i = 1, \ldots, n$ is constant within the intervals I_l but changes at each division point s_l, see Figure 1. Individual observation times $y_i \in I_j$ therefore can be split into periods $u_{il}, l = 1, \ldots, j$ observed under the regime of hazards $\lambda_{il}, l = 1, \ldots, j$ with

$$\lambda_{il} = \lambda(t|\mathbf{z}_i, t \in I_l) = \exp\left((\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l\right).$$

Consider e.g. an observation time $y_i \in I_2$, which is composed of the time $u_{i1} = \Delta_1 = s_1 - s_0$ spent under the regime of hazard rate λ_{i1} and the time $u_{i2} = y_i - s_1$ spent under the regime of hazard rate λ_{i2}, see Figure 1 for illustration. In time point s_1, where the hazard rate changes, subject i is still alive. The survival time this subject would experience under hazard rate λ_{i1}, is not completed, but right censored at s_1.

More generally, for $y_i \in I_j$, $y_i = \sum_{l=1}^{j} u_{il}$. The observed time under hazard λ_{il} is $u_{il} = \Delta_l = s_l - s_{l-1}$, for $l = 1, \ldots, j-1$. For $l < j$, each u_{il} can be regarded as a right-censored observation, i.e. $u_{il} = \min(\tau_{il}, \Delta_l)$ where τ_{il} is the unobserved auxiliary survival time. As the hazard is constant in each interval, this auxiliary survival time τ_{il} is exponentially distributed, $\tau_{il} \sim \text{Exponential}(\lambda_{il})$. If $y_i \in I_j$, the time spent in the last interval is $u_{ij} = y_i - s_{j-1}$ and for $\delta_i = 1$, this is an uncensored, otherwise a censored observation from an Exponential (λ_{ij})-distribution.

Due to the no-memory property of the exponential distribution the random variables $u_{il}, l = 1, \ldots, j$ are independent.

Splitting the observed time y_i into periods u_{il} can be motivated more formally from factorizing the likelihood contribution of an individual observation $(y_i, \delta_i) \in I_j$ over contributions of time intervals $I_l, l = 1 \ldots, j$.

Generally the likelihood of an observation time y and a censoring indicator δ is

$$p(y, \delta|.) = S(y|.)\lambda(y|.)^{\delta}$$

where $S(y|.)$ and $\lambda(y|.)$ denote the appropriate survival and hazard function.

The survival function in the piecewise exponential model can be expressed through conditional survival functions as

$$S(y_i|\mathbf{z}_i) = \big(\prod_{l=1}^{j-1} S(s_l|T > s_{l-1}, \mathbf{z}_i)\big) S(t|T > s_{j-1}, \mathbf{z}_i) = \Big(\prod_{l=1}^{j-1} \exp\big(-\lambda_{il}(s_l - s_{l-1})\big)\Big) \exp\big(-\lambda_{ij}(y_i - s_{j-1})\big)$$

and hence

$$p\big(y_i, \delta_i|\lambda_{i1}, \ldots, \lambda_{ij}\big) = \Big(\prod_{l=1}^{j-1} \exp\big(-\lambda_{il}\Delta_l\big)\Big) \exp\big(-\lambda_{ij}(y_i - s_{j-1})\big)(\lambda_{ij})^{\delta_i}.$$

Obviously contributions of the intervals $l = 1, \ldots, j$ are independent. The likelihood contribution of interval I_l, for $l = 1, \ldots, j-1$ is $\exp\big(-\lambda_{il}\Delta_l\big)$, which corresponds to that of a random variable distributed as Exponential (λ_{il}), which is right-censored at Δ_l. The likelihood contribution of the last interval I_j, $\exp\big(-\lambda_{ij}(y_i - s_{j-1})\big)(\lambda_{ij})^{\delta_i}$ is equal to that of a right-censored or a complete exponential survival time, depending on the censoring indicator δ_i.

3.2.2 Data augmentation for the dynamic survival model

If instead of $\big((y_{i1}, \delta_{i1}), \ldots, (y_{ij}, \delta_{ij})\big)$ the complete survival times $(\tau_{i1}, \ldots \tau_{ij})$ were available, using the auxiliary mixture sampler would be feasible, as then the model could be represented as a normal state space model.

The basic idea of the first data augmentation step therefore is to introduce unobserved exponentially distributed auxiliary survival times under the regime of each hazard rate λ_{il}. To achieve this, first note, that conditionally on $y_i \in I_j$ the unobserved survival time $\tau_{il}, \quad l = 1, \ldots, j-1$ has an exponential distribution with mean $1/\lambda_{il}$, left-truncated at Δ_l, i.e. the residual lifetime under hazard rate λ_{il} is

$$\xi_{il} = \tau_{il} - \Delta_l \quad \sim \text{Exponential}\,(\lambda_{il})\,.$$

For $y_i \in I_j$ the observation time in the last interval $\tau_{ij} = y_i - s_{j-1}$ is an observation from the Exponential (λ_{ij}) distribution, either complete or right censored.

Thus, for any observation time $(y_i, \delta_i) \in I_j$ in the first data augmentation step the complete auxiliary survival times $\tau_{i1}, \ldots, \tau_{ij}$ can be generated as

$$\tau_{il} = \Delta_l + \xi_{il}, \qquad \text{for } l = 1, \ldots, j-1,$$

$$\tau_{ij} = \begin{cases} y_i - s_{j-1}, & \text{if } \delta_i = 1, \\ y_i - s_{j-1} + \xi_{ij}, & \text{if } \delta_i = 0. \end{cases}$$

where $\xi_{il} \sim \text{Exponential}\,(\lambda_{il})$ with $\lambda_{il} = \exp((\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l)$ for $l = 1, \ldots, j$.

In terms of the auxiliary survival times the dynamic survival model (2) and (3) is special case of the dynamic generalized linear model considered by West et al. (1985), for exponentially distributed observations

$$\tau_{il}|\boldsymbol{\alpha}, \boldsymbol{\beta}_l \sim \text{Exponential}\left(\exp((\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l)\right), \tag{7}$$

$$\boldsymbol{\beta}_l = \mathbf{F}\boldsymbol{\beta}_{l-1} + \boldsymbol{\omega}_l,\ \boldsymbol{\omega}_l \sim \text{Normal}\,(0, \mathbf{Q}), \tag{8}$$

where the matrices $\mathbf{F}$ and $\mathbf{Q}$ ar known or may depend on unknown model parameters $\boldsymbol{\theta}$. In our model, $\mathbf{F}$ is the identity matrix and $\mathbf{Q} = \text{diag}(\theta_0, \ldots, \theta_K)$. For the negative of the logarithm of the auxiliary survival times this is a linear state space model

$$\begin{aligned} -\ln \tau_{il} &= (\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l + \varepsilon_{il} \\ \boldsymbol{\beta}_l &= \boldsymbol{\beta}_{l-1} + \boldsymbol{\omega}_l, \quad \boldsymbol{\omega}_l \sim \text{Normal}\,(0, \mathbf{Q}) \end{aligned}$$

where the error ε_{il} has a type I extreme value distribution. Using the mixture approximation and introducing the component indicator r_{il} for each auxiliary survival time τ_{il} in the second data augmentation step – as described in Section 3.1 – we arrive at a representation as a partially Gaussian state space model as in Shephard (1994).

Conditioning on all auxiliary survival times and the component indicators, the observation equation can be written as

$$-\ln \tau_{il}|\lambda_{il}, r_{il} = (\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l + m_{il} + \varepsilon_{il}, \quad \varepsilon_{il} \sim \text{Normal}\left(0, s_{r_{il}}^2\right).$$

Let n_l denote the number of individuals at risk at the beginning of interval I_l, and assume that the observation times are arranged decreasingly, so that y_1 is the largest and y_n is the smallest observed time. If we define a multivariate observation vector $\tilde{\mathbf{y}}_l$ of dimension n_l as

$$\tilde{\mathbf{y}}_l = \begin{pmatrix} -\ln \tau_{1l} - m_{r_{1l}} \\ \vdots \\ -\ln \tau_{n_l,l} - m_{r_{n_l,l}} \end{pmatrix},$$

the model may be written in the following linear Gaussian state space form:

$$\tilde{\mathbf{y}}_l = \tilde{\mathbf{Z}}_l^f\boldsymbol{\alpha} + \tilde{\mathbf{Z}}_l^v\boldsymbol{\beta}_l + \boldsymbol{\varepsilon}_l,\ \boldsymbol{\varepsilon}_l \sim \text{Normal}\,(0, \mathbf{V}_l), \tag{9}$$

$$\boldsymbol{\beta}_l = \boldsymbol{\beta}_{l-1} + \boldsymbol{\omega}_l, \quad \boldsymbol{\omega}_l \sim \text{Normal}\,(0, Q), \tag{10}$$

where $\mathbf{V}_l = \text{Diag}\left(s_{r_{1l}}^2, \ldots, s_{r_{n_l,l}}^2\right)$. $\tilde{\mathbf{Z}}_l^f$ and $\tilde{\mathbf{Z}}_l^v$ are matrices with n_l rows, containing the design vectors $\mathbf{z}_1^f, \ldots, \mathbf{z}_{n_l}^f$, and $\mathbf{z}_1^v, \ldots \mathbf{z}_{n_l}^v$ for all individuals at risk at s_{l-1}:

$$\tilde{\mathbf{Z}}_l^f = \begin{pmatrix} (\mathbf{z}_1^f)' \\ \vdots \\ (\mathbf{z}_{n_l}^f)' \end{pmatrix}, \qquad \tilde{\mathbf{Z}}_l^v = \begin{pmatrix} (\mathbf{z}_1^v)' \\ \vdots \\ (\mathbf{z}_{n_l}^v)' \end{pmatrix}.$$

Thus, instead of the original dynamic survival model we have a partially Gaussian state space model, where the transition equation is the same as for the original model but the model for the log-hazard is replaced by a Gaussian observation equation with a multivariate observation vector $\tilde{\mathbf{y}}_l$, determined from the auxiliary survival times of the individuals at risk at the beginning of interval I_l.

3.2.3 The sampling scheme

Auxiliary mixture sampling can be carried out by adding a further step to the auxiliary mixture sampler described in Section 3.1. Select starting values for $\theta_0, \ldots, \theta_K$, the component indicators $\mathcal{R}$ and λ_{il} and repeat the following steps:

1. Carry out multi-move sampling for the whole sequence $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by forward-filtering backward sampling as in Frühwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995), or Durbin and Koopman (2002) for the conditionally Gaussian state space form (9) and (10).

2. Sample $\boldsymbol{\theta}$ conditional on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$ and $\mathcal{R}$ from the conditionally Gaussian state space form (9) and (10).

3. For each observation $(y_i, \delta_i) \in I_j$ sample the unobserved auxiliary survival times $\tau_i = (\tau_{il}; \quad l = 1, \ldots, j)$ and the component indicators r_{il} for each $\tau_{il}, l = 1, \ldots, j$ by sampling independently from a discrete distribution with 10 categories.

In the dynamic survival model $\mathbf{Q} = \text{diag}(\theta_0, \ldots, \theta_k)$ and thus the process variances $\theta_0, \ldots, \theta_k$ are sampled independently from inverse Gamma distributions.

We have not yet commented on how to obtain starting values for λ_{il}. One choice would be to draw values for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ from their respective prior distributions. An alternative which was used in the applications in the next section, is to use the same value for all auxiliary survival times in interval I_j, $\lambda^0_{ij} = \lambda^0_j$, which implies setting the starting values for all covariate effects to zero. For λ^0_j we used the ML estimator for the log-hazard $\hat{\lambda}_j$, which is the number of failures in interval I_j divided by total observed time in I_j.

4 Applications

4.1 Simulated data

For illustration of the sampler we first consider a set of $n = 200$ survival times, generated from a dynamic survival model, without covariates and a baseline log-hazard β evolving as a random walk. The starting point for the random walk was $\beta_0 = -5$ and the process variance θ_0 was set to 0.3. The time axes from 0 to 20 was divided into 20 intervals of length 1, and a last interval starting at $s = 20$ and ending at the largest survival time was added.

Data were analyzed using the complete survival times and under right-censoring for two different censoring schemes. Censoring times were generated as $c_i \sim U([0, 40])$ and $c_i \sim$ Exponential (10) leading to 21.5% respectively 50% censored observation. The largest observed time was 29.5610 for the uncensored data set and under censoring scheme 2 and 26.1537 for the first censoring scheme.

The auxiliary mixture sampler was run 25000 times with a burnin of 5000. We used 4 different choices for the prior distribution, which all are rather uninformative: $c_0 = C_0 = 0$, $c_0 = C_0 = 0.00001$, $c_0 = C_0 = 0.01$, $c_0 = 0.1, C_0 = 0.01$. Though $c_0 = C_0 = 0$ is an improper prior, it worked well for this example.

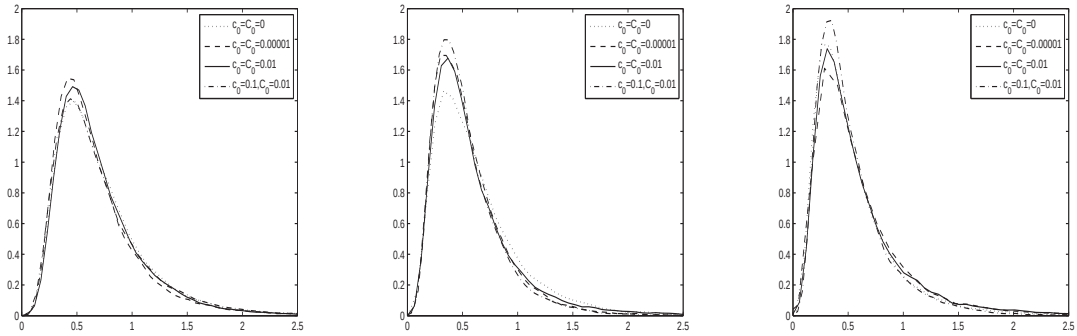

Figure 2: Kernel density estimates for posterior density of θ for complete survival times (left) , censored data 1 (middle), censored data 2 (right)

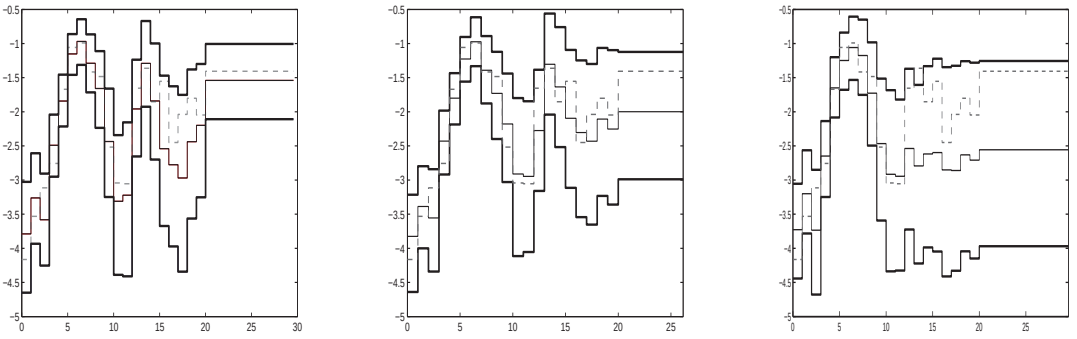

Figure 3: Estimated (full) and true (dashed) baseline log-hazard within 95%-credible regions for complete survival times (left), data under censoring scheme 1 (middle), and under censoring 2 (right)

Figure 2 shows the kernel density estimates for the posterior distribution of θ obtained from the sampling output. Obviously there is not much difference in the posterior distributions for θ for these 4 priors. Results were very similar for these priors with regard to the estimated baseline log-hazard as well. In Figure 3 the baseline log-hazard is compared to the true values β_0 used for generating the data for $c_0 = C_0 = 0.01$ The true value lies in the 95%-point-wise credible interval for any time interval except one for the second censoring scheme.

4.2 Gastric cancer data

As a second illustration we applied the auxiliary mixture sampler on a data set of patients with gastric cancer, analyzed previously by Gamerman (1991) and Hemming and Shaw (2005).

The data are survival times of 90 patients, randomly allocated to a therapy. Treatment was chemotherapy in the first group, and in the second group a combination of chemotherapy and radiation. Overall 10 observation times were right

Table 2: Estimated process variances for gastric cancer data

Parameter	Mean	Std.dev.	95%H.P.D. regions
θ_1	0.0245	0.0220	[0.0016 0.0705]
θ_2	0.0553	0.0534	[0.0033 0.1718]

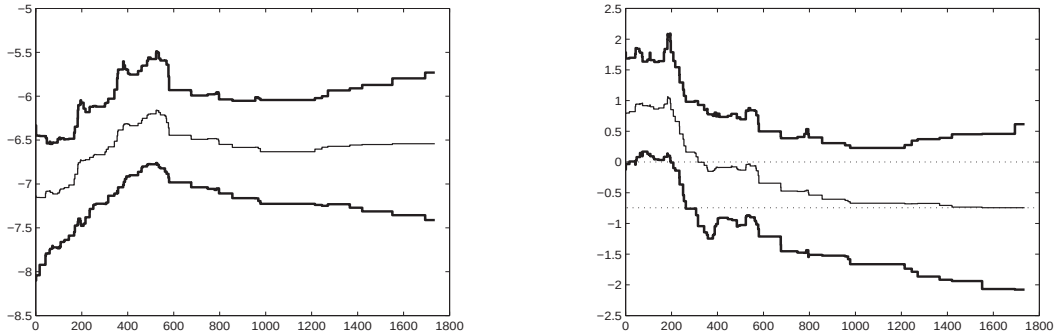

Figure 4: Gastric cancer data: Posterior means and 95% credible regions for the baseline log-hazard (left) and the effect of combined treatment (right)

censored.

In this setting it is of interest whether there is a treatment effect of combined therapy as compared to chemotherapy alone and whether this effect varies over time. The covariate vector is $\mathbf{z}_i^v = (1, z_{1i})$ where z_{1i} indicates whether patient i underwent combined therapy or not. The parameter vector $\boldsymbol{\beta}$ therefore has 2 components.

In their analysis Gamerman (1991) as well as Hemming and Shaw (2005) used the failure times in the data as division points of the time axis. We applied the auxiliary mixture sampler with the same intervals with 25000 iterations and a burnin of 5000. For the process variances we used an inverse Gamma prior with parameters $c_{0k} = C_{0k} = 0.01, k = 0, 1$. Prior moments for starting values $\boldsymbol{\beta}_0$ were chosen to be $\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{B}_0 = 100\,\mathbf{I}_2$.

Table 2 reports point estimates, standard errors as well as 95%-highest posterior density regions for the process variances θ_0 and θ_1. Figure 4 shows the estimated baseline log-hazard and the effect of combined therapy. Both plots are similar to those of Hemming and Shaw (2005).The baseline log-hazard increases early followed by a sharp decline, but as the credible intervals include a straight line, a constant baseline-hazard, as assumed in Gamerman (1991) cannot be ruled out. The estimated treatment effect varies with time declining from a significantly positive to a negative effect in the long run. Risk of death is thus higher for patients treated with combined therapy during the first 200 days but lower than for those treated only with chemotherapy later on. Evidence is clear in favor of a dynamic effect.

4.3 Worcester heart attack data

As a further application we analyzed the data of the Worcester heart attack study used in Hosmer and Lemeshow (1999). The main goal of the study was to describe

Table 3: Description of variables from the Worcester Heart Attack Study

Variable	Description
`age`	age at hospital admission in years
`sex`	0=male, 1=female
`sho`	cardiogenic shock complications (0=no, yes=1)
`cpk`	peak cardiac enzyme measured in international units (IU)
`chf`	left heart failure complications(0=no, yes=1)
`miorder`	myocardial infection order (0=first, 1=recurrent)
`mitype`	myocardial infection type (0=Q wave, 1=not Q wave or indeterminate)

trends over time in the incidence and survival rates following hospital admission for acute myocardial infarction. The data set provided by Hosmer and Lemeshow (1999) is a sample of the main data set with information on 481 patients. Additionally to length of follow-up, defined as days from hospital admission and status of last follow-up (dead or alive), several covariates were available which are described in Table 3.

In a preliminary analysis we estimated a Cox proportional hazards model and investigated for each of the seven covariates whether the proportional hazards assumption was appropriate. Plots of smoothed Schoenfeld residuals indicate that the proportional hazards assumption is not violated for covariates `age`, `sex` and `sho`, whereas for the remaining covariates time varying effects seem to be more appropriate. Therefore we finally fitted a dynamic survival model with fixed effects for three covariates (`age`, `sex` and `sho`) and time-varying effects for the remaining four covariates (`cpk`, `chf`, `miorder` and `mitype`). Most events occur early, therefore the time axis was partitioned into intervals of different length, starting with intervals of length 30 (corresponding to months) until day 1800, followed by intervals of length 90 (corresponding to quarters) until day 3600 and finally intervals of length 360.

As priors for fixed effects and starting values for the random walks we used independent standard normal distributions; for the inverse gamma prior for the process variances we chose $c_0 = 0.1$ and $C_0 = 0.01$. The auxiliary mixture sampler was run with 25 000 iterations and a burnin of 5000.

Estimation results for fixed effects and process variances are presented in Table 4, Figure 5 shows the estimated time varying effects. Whereas sex has no significant effect, age and cardiogenic shock complications (`sho`) increase the risk. The effect of peak cardiac enzyme (`cpk`) is significantly positive in the first month and negative in the long run. Higher values of `cpk` significantly decrease the risk during the time period of month 46 to month 70 after acute myocardial infection. Left heart failure complications (`chf`) significantly increase risk during the first year, but this effect vanishes in the long run. Recurrent myocardial infection (`miorder`) has a significant positive effect only in the first month, declining to a slightly negative long run effect. The effect of myocardial infection type is positive early after acute myocardial infection, but close to significance only in the first month, and decreases to a negative long run effect.

Table 4: Estimated parameters for the WHAS data

fixed effects	Mean	Std.dev.	95%H.P.D. regions
`age`	0.0173	0.0029	[0.0115 0.0226]
`sex`	0.0031	0.1347	[-0.2520 0.2433]
`sho`	1.7858	0.2024	[1.3886 2.1781]
process variances	Mean	Std.dev.	95%H.P.D. regions
θ_1 (`baseline`)	0.4984	0.1545	[0.2150 0.7946]
θ_2 (`cpk`)	0.0269	0.0212	[0.0037 0.0691]
θ_3 (`chf`)	0.0308	0.0324	[0.0020 0.1020]
θ_4 (`miorder`)	0.0149	0.0113	[0.0017 0.0371]
θ_5 (`mitype`)	0.0240	0.0287	[0.0012 0.0865]

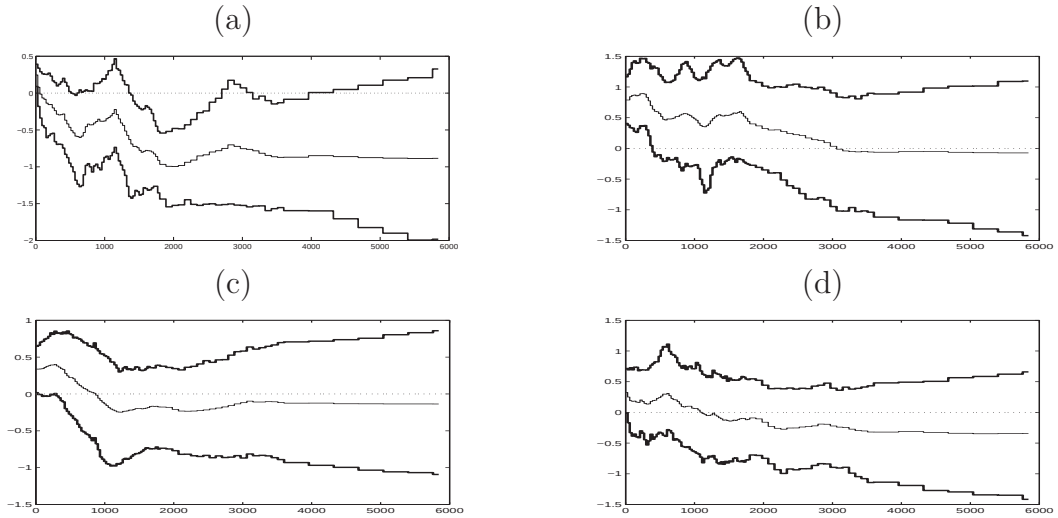

Figure 5: WHAS data: Posterior means and 95% credible regions for the effect of `cpk` (a), `chf` (b), `miorder` (c) and `mitype` (d)

5 Extensions

5.1 Extending the linear predictor

The auxiliary mixture sampler has a wider application than for the dynamic survival model described above. Through the mixture approximation it can deal with more general models, where further components are added to linear predictor, as long as the corresponding normal model can be estimated by Gibbs sampling.

Consider, for example, a dynamic survival model with random effects, i.e. frailties which are useful either to model correlation of survival times or unobserved heterogeneity, see Therneau and Grambsch (2000). To model correlation within groups of individuals group-specific frailties are used, whereas unobserved heterogeneity is captured by individual frailties. As frailties are supposed to act multiplicatively on the hazard, for an individual i with frailty u_i the hazard is

$$\lambda(t|\mathbf{z}_i, u_i, t \in I_j) = u_i \ \exp((\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_j)$$

A common choice for the frailty distribution is the Gamma distribution, in particular for exponential or Weibull survival data. With the auxiliary mixture sampler however it is more convening to assume a log-normal distribution, i. e. $\ln u_i \sim \text{Normal}\,(0, D_0)$. The partial Gaussian representation of the model for the auxiliary survival times conditioning on the component indicators r_{il} is

$$-\ln \tau_{il}|\lambda_{il} = (\mathbf{z}_i^f)'\boldsymbol{\alpha} + (\mathbf{z}_i^v)'\boldsymbol{\beta}_l + \ln u_i + m_{il} + \varepsilon_{il}, \quad \varepsilon_{il} \sim \text{Normal}\left(0, s^2_{r_{il}}\right),$$

which is a linear mixed model with random effect $\ln u_i$. Just one additional step has to be added to the auxiliary mixture sampler to draw the random effects $\ln u_i, i = 1, \ldots, n$ from their joint posterior which is a multivariate normal distribution.

Adding multidimensional random effects or unstructured spatial effects with normal priors is straightforward. If spatially structured effects with Markov random field smoothing priors are included in the model, the efficient sampling algorithms for normal response models, presented Rue and Held (2005) can be used.

5.2 Different censoring schemes

As the model is parametric, data augmentation to generate auxiliary survival times is feasible also for missing information different from right-censoring. Consider for example an interval censored observation, where the survival time is known to lie in the interval $(t_L, t_R]$. Conditional on the available information, that $t_i \in (t_L, t_R]$, the survival function is

$$S^*(t) = P(t_i > t|t_i \in (t_L, t_R]) = \frac{S(t)}{S(t_L) - S(t_R)} \quad \text{for } t \in (t_L, t_R].$$

Complete auxiliary survival times can easily be obtained by first sampling the number v_i of the interval where failure for individual i occurs and then generating auxiliary survival times for each interval, $l = 1, \ldots, v_i$.

Let j_L and j_R be the number of intervals containing t_L and t_R respectively. Then v_i has a discrete distribution, given as

$$p(v_i = j) = \begin{cases} \frac{S(t_L)-S(s_j)}{S(t_L)-S(t_R)} & j = j_L \\ \frac{S(s_{j-1})-S(s_j)}{S(t_L)-S(t_R)} & j_L < j < j_R \\ \frac{S(s_{j-1})-S(t_R)}{S(t_L)-S(t_R)} & j = j_R \end{cases}.$$

Conditional on knowing $v_i = j$, all auxiliary survival times except the last, i.e. $\tau_{il}, l < j$ are known to be right censored. Auxiliary residual survival times ξ_{il} and hence auxiliary survival times τ_{il}, can be generated as described in Section 3.2. The last auxiliary survival time τ_{ij} is known not to exceed $\xi_{max} = s_j - s_{j-1} = \Delta_{ij}$ for $j < j_R$, respectively $\xi_{max} = t_R - s_{j-1}$ for $j = j_R$ and corresponds to an observation ξ_{ij} from the right truncated Exponential (λ_{ij})-Distribution. ξ_{ij} thus can be generated by sampling a uniform $U[0, 1]$ random variable and inverting the survival function

$$S^*(x) = \frac{\exp(-\lambda_{ij}x)}{1 - \exp(-\lambda_{ij}\xi_{max})}.$$

The auxiliary survival times then are defined as

$$\tau_{il} = \begin{cases} \Delta_{il} + \xi_{il} & \text{for } l < j \\ \xi_{ij} & \text{for } l = j \end{cases}$$

6 Discussion

Dynamic survival models are a useful alternative to the Cox model for analyzing survival data by allowing time-varying effects of covariates. We proposed a new auxiliary mixture sampler for these models. The convenience of this sampler results from the representation of the model for the log-hazard as a partial Gaussian model for auxiliary survival times. This allows to deal with any form of the linear predictor where Gibbs sampling for the equivalent model with Gaussian errors is feasible. In particular sampling algorithms for Gaussian models are easily adapted to survival models with the same linear predictor by adding the two steps of data augmentation described above. Thus inclusion of unstructured or structured spatial effects or nonlinear effects of covariates modelled by P-splines as in Hennerfeind et al. (2006) is straightforward. The key property that has to be maintained for application of the auxiliary mixture sampler is the piecewise constant structure of the log-hazard as a function of time. Models where the evolution of log-hazard in time is described by B-Splines of a higher degree, as in Hennerfeind et al. (2006) cannot be dealt with.

However the sampler can be easily adapted for models including time-dependent covariates, if these change their values only at some distinct time points. Let $t_C \in I_j$ denote the change-point of a covariate, then the interval I_j is split into two periods $(s_{j-1}, t_C]$ and $(t_C, s_j]$, which are observed under different hazards λ and λ^*. By introducing complete auxiliary survival times and component indicators for both of these observation times as described in Section 3.2 the representation as a partial gaussian state space model is maintained.

A further advantage of the dynamic survival model is, as it is a parametric model, that – by data augmentation – allows to deal with missing information different from right-censoring, details have been given for interval censoring.

Many useful generalizations of the model, as e.g. modeling the linear predictor by mixtures of linear models with random effects as in Lenk and DeSarbo (2000), or models allowing process variances to depend on the length of the time interval or varying over time are potential topics for further research.

References

Arjas, E. and Gasbarra, D. (1994). Semiparametric Bayesian analysis of survival data. *Statistica Sinica*, 4:505–524.

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

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:281–316.

de Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82:339–350.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89:603–15.

Fahrmeir, L. and Hennerfeind, A. (2003). Nonparametric Bayesian hazard rate models based on penalized splines. Working paper. LMU München.

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

Frühwirth-Schnatter, S. and Frühwirth, R. (2006). Auxiliary mixture sampling with applications to logistic models. IFAS research report, `http://www.ifas.jku.at`.

Frühwirth-Schnatter, S. and Wagner, H. (2005). Data augmentation and Gibbs sampling for regression models of time series of small counts. *Student*, 5:207–220.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. To appear in Biometrika.

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

Hemming, K. and Shaw, J. E. H. (2005). A class of parametric dynamic survival models. *Lifetime Data Analysis*, 11:81–98.

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). Geoadditive survival models. To appear in Journal of the American Statistical Association.

Hosmer, D. W. and Lemeshow, S. (1999). *Applied Survival Analysis*. John Wiley & Sons.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer Verlag.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility:likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393.

Lenk, P. J. and DeSarbo, W. S. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65:93–119.

Nieto-Barajas, L. E. and Walker, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, 29:413–424.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2006). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics*, pages xx–xx.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*. Taylor & Francis.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika*, 81:115–131.

Sinha, D., Chen, M. H., and Ghosh, S. K. (1999). Bayesian analysis and model selection for interval censored data. *Biometrics*, 55:585–590.

Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212.

Therneau, T. M. and Grambsch, P. M. (2000). *Modelling Survival data*. Springer Verlag.

Verweij, P. J. M. and van Houwelingen, H. C. (1995). Time-dependent effects of fixed covariates in cox regression. *Biometrics*, 51:1550–1556.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (C/R: p84-97). *Journal of the American Statistical Association*, 80:73–83.