Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2009-40

Bayesian Estimation and Stochastic Model Specification Search for Dynamic Survival Models

Helga Wagner

February 2009

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. In this paper a new auxiliary mixture sampler for Bayesian estimation of the model parameters is introduced. This sampler forms the basis of a model space MCMC method for stochastic model specification search in dynamic survival models, which involves selection of covariates to include in the model as well as specification of effects as time-varying or constant. The method is applied to two well-known data sets from the literature.

Some key words: piecewise exponential model, time-varying effects, non-centered parameterization, variable selection, model choice, Cox model

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 modelled as

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

Here $\mathbf{z}$ is a vector of covariates and $\boldsymbol{\beta}$ is the vector of unknown parameters. 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

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

This is an assumption which is not necessarily true in applications 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 covariate effects, see Verweij and van Houwelingen (1995) for a frequentist and Gamerman (1991), Arjas and Gasbarra (1994) and Sinha et al. (1999) for a Bayesian approach.

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 and a specification of the stochastic evolution over time. Whereas Gamerman (1991) used a random walk with process disturbances specified only up to their second moments, recently Hemming and Shaw (2002, 2005) considered the normal dynamic survival model with Gaussian process disturbances. Closely related to this model specification is the model of Hennerfeind et al. (2006) who use penalized splines for baseline log-hazard and covariate effects with a correlated normal prior process for the spline coefficients.

Estimation of the dynamic survival model was accomplished in Gamerman (1991) using linear Bayes approximation. In a fully Bayesian approach posteriors for all parameters of normal dynamic survival models can be obtained by MCMC methods which however require a Metropolis-Hastings-Algorithm, see Hemming and Shaw

(2005). In this paper a new approach for estimation of the normal dynamic survival model relying on data augmentation is proposed. This new auxiliary mixture sampler involves only draws from standard densities and needs no further tuning. By introducing two sequences of latent variables a representation of the normal dynamic survival model as a Gaussian state space model is obtained, where the multidimensional latent state vector consisting of baseline log-hazard and time-varying covariate effects can be sampled in one move. 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., 2007). Frühwirth-Schnatter and Wagner (2006b,a) introduced auxiliary mixture sampling for Bayesian analysis of parameter-driven models for count data based on the poisson distribution and Frühwirth-Schnatter and Frühwirth (2007) apply auxiliary mixture sampling to binary and multinomial logit models.

Dynamic survival models are very flexible but model specification is a complex task as one has to decide not only which covariates to include in the final model but also whether the effect of a certain covariate is constant or varies over time. A classical model choice and variable selection strategy for similar models based on the conditional AIC was only recently implemented in Hofner et al. (2008). In a Bayesian setting model selection can be accomplished by model space MCMC methods, e.g. the reversible jump algorithm (Green, 1995) or the stochastic variable selection approach (George and McCulloch, 1993, 1997). Konrath et al. (2008) consider Bayesian regularisation together with a hard shrinkage rule for variable selection.

The stochastic variable selection approach, usually applied for model selection in regression models, was recently extended to model selection in state space models by Frühwirth-Schnatter and Wagner (2008). Based on the data augmentation scheme introduced in this paper its implementation is feasible also for the normal dynamic survival model. Thus it is possible to start with a general model specification where all covariates are included with their effects specified as time-varying. By exploring the model space during MCMC covariate effects are identified as constant rather than time-varying or even as zero leading to a parsimonious model specification.

The rest of the paper is organized as follows: Section 2 describes the model specification. Data augmentation by auxiliary variables and the resulting sampling scheme are discussed in Section 3. In Section 4 the noncentered parameterization of the model is introduced and variable selection is incorporated. The methods are applied on two data sets from the literature in Section 5. Finally Section 6 concludes by summarizing the results and discussing possible extensions.

2 The Normal Dynamic Survival Model

2.1 Model specification

As 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 d_i, taking the value 1 for complete and 0 for censored observations, 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 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 model, as proposed by Gamerman (1991), is a piecewise exponential model for lifetimes, with correlated prior processes for the baseline log-hazard and 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 $\beta_k(t)$ is assumed to follow a random walk. Whereas Gamerman (1991) specified process disturbances only up to their second moments, recently Hemming and Shaw (2002) considered the normal dynamic survival model with Gaussian random walks,

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

The random walk priors are smoothness priors which penalize abrupt jumps of baseline log-hazard and covariate effects in subsequent intervals. Baseline log-hazard and time-varying effects can also be interpreted as a linear combination of J B-spline basis functions of degree zero with knots $\{s_0, s_1, \ldots, s_J\}$,

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

and thus the normal dynamic survival model is a special case of the model considered in Hennerfeind et al. (2006).

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. Hemming and Shaw (2002) use equally spaced time points and Gamerman (1991) and Hemming and Shaw (2005) use a data-dependent division where the division points are the observed failure times.

In this paper I consider estimation and model selection for the normal dynamic survival model where the piecewise constant hazard of subject i is defined by

$$\lambda(t|\mathbf{z}_i; t \in I_j) = \lambda_{ij} = \exp(\mathbf{z}_i'\boldsymbol{\beta}_j) \tag{2}$$

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j \qquad \boldsymbol{\omega}_j \sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})). \tag{3}$$

Here $\mathbf{z}_i = (1, z_{i1}, \ldots, z_{iK})$ is the vector of covariates and $\boldsymbol{\beta}_j = (\beta_{0j}, \ldots, \beta_{Kj})'$ denotes the effects in interval $I_j, j = 1, \ldots, J$, β_{0j} being the baseline log-hazard in I_j. The components of the state vector $\boldsymbol{\beta}$ are assumed to evolve independently, hence $\mathbf{Q}(\boldsymbol{\theta}) =$

diag$(\theta_0, \ldots, \theta_K)$. In this model an evolution variance $\theta_0 = 0$ implies a constant baseline hazard, and $\theta_k = 0$ a constant effect of covariate z_k. If all evolution variances $\theta_k, k = 0, \ldots, K$ are zero, the model reduces to the exponential regression model

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

with fixed covariate effects $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)$.

2.2 Likelihood

The likelihood of the normal dynamic survival model defined in equations (2) - (3) is the product of individual likelihood contributions $L(y_i, d_i|\mathbf{z}_i) = S(y_i|\mathbf{z}_i)\lambda(y_i|\mathbf{z}_i)^{d_i}$ where $S(y|.)$ and $\lambda(y|.)$ denote the survival and the hazard function. The survival function of an observation $y_i \in I_l$ can be expressed through conditional survival functions, see Hemming and Shaw (2005), as

$$S(y_i|\mathbf{z}_i) = \big[\prod_{j=1}^{l-1} S(s_j|T > s_{j-1}, \mathbf{z}_i)\big] S(y_i|T > s_{l-1}, \mathbf{z}_i)$$

Hence

$$L\big(y_i, d_i|\lambda_{i1}, \ldots, \lambda_{ij}\big) = \Big[\prod_{j=1}^{l-1} \exp\big(-\lambda_{ij}(s_j - s_{j-1})\big)\Big] \exp\big(-\lambda_{il}(y_i - s_{l-1})\big)(\lambda_{il})^{d_i}. \quad (4)$$

3 Auxiliary mixture sampling

Estimation of the unknown parameters of normal dynamic survival model, i.e. the state vector $\boldsymbol{\beta}$ and the process variances $\boldsymbol{\theta}$, is feasible by MCMC methods. Hemming and Shaw (2002) use Gibbs sampling with a single move random walk Metropolis-Hastings-step to sample the state vector from the conditional posterior $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{d}, \boldsymbol{\theta})$. In this paper a simple Gibbs scheme based on data augmentation and avoiding Metropolis-Hastings steps is proposed. Building on the ideas of Frühwirth-Schnatter and Wagner (2006a) and Frühwirth-Schnatter and Frühwirth (2007) two sequences of latent variables are introduced which lead to a representation of the normal dynamic survival model as a conditional Gaussian state space model where direct sampling of $\boldsymbol{\beta}$ in one move is possible.

3.1 Splitting the observation time into episodes

The factorization of the likelihood into independent contributions from the time intervals I_j suggests to split an observation time $y_i \in I_l$ into episodes $u_{ij}, j = 1, \ldots, l$ experienced under the regime of the constant hazard λ_{ij}. The total observation time y_i is the sum of these episodes, $y_i = \sum_{j=1}^{l} u_{il}$. Consider e.g. an observation time $y_i \in I_2$ which is the sum of the time $u_{i1} = s_1 - s_0$ spent under hazard rate λ_{i1} and the time $u_{i2} = y_i - s_1$ spent under hazard rate λ_{i2}, see Figure 1 for illustration.

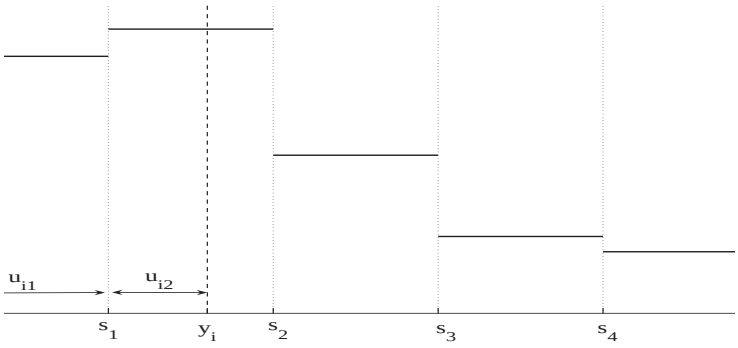

Figure 1: Hazard in the piecewise exponential model

Each episode u_{ij} except the last is just the length of the interval I_j, $u_{ij} = s_j - s_{j-1}, j = 1, \ldots, l-1$. The likelihood contributions of these episodes are of the form

$$\exp\big(-\lambda_{ij}u_{ij}\big), \quad j = 1, \ldots, l-1$$

and are equal to the likelihood contribution of a right-censored $\mathrm{Ex}(\lambda_{ij})$ observation. The last episode subject i experiences is $u_{il} = y_i - s_{l-1}$ with a likelihood contribution given as

$$\exp\big(-\lambda_{il}u_{il}\big)(\lambda_{il})^{d_i}. \tag{5}$$

Dependent on the censoring indicator d_i, (5) corresponds to the likelihood either of a complete (for $d_i = 1$) or a right-censored $\mathrm{Ex}(\lambda_{il})$ observation.

3.2 Data Augmentation

In the first data augmentation step all episodes not ending by the occurrence of the interesting event are interpreted as right-censored. For each censored episode u_{ij} an unobserved complete survival time τ_{ij} is introduced. The residual lifetime $\xi_{ij} = \tau_{ij} - u_{ij}$, conditional on $\{\tau_{ij} > u_{ij}\}$, follows the $\mathrm{Ex}(\lambda_{ij})$-distribution, due to the no-memory property of the exponential distribution. Therefore the complete auxiliary survival times τ_{ij} are given as

$$\tau_{ij} = u_{ij} + \xi_{ij}, \quad \xi_{ij} \sim \mathrm{Ex}(\lambda_{ij}) \qquad \text{for } j = 1, \ldots, l-1 \tag{6}$$

and

$$\tau_{il} = \begin{cases} u_{il} & \text{if } d_i = 1, \\ u_{il} + \xi_{il}, \quad \xi_{il} \sim \mathrm{Ex}(\lambda_{il}) & \text{if } d_i = 0. \end{cases} \tag{7}$$

Using the auxiliary survival times τ_{ij} the normal dynamic survival model defined in equations (2) and (3) can be represented as a dynamic generalized linear model (West et al., 1985) with exponentially distributed observations

$$\tau_{ij}|\boldsymbol{\beta}_j \sim \mathrm{Ex}\left(\exp\left(\mathbf{z}_i'\boldsymbol{\beta}_j\right)\right), \tag{8}$$

$$\boldsymbol{\beta}_j = \mathbf{F}\boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j, \quad \boldsymbol{\omega}_j \sim N\left(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})\right), \tag{9}$$

with $\mathbf{F} = \mathbf{I}$.

Taking logarithms the observation equation of this model is transformed into the linear model

$$-\ln \tau_{ij} = \mathbf{z}_i \boldsymbol{\beta}_j + \epsilon_{ij}$$

where the error ϵ_{ij} has a type I extreme value distribution. As shown in Frühwirth-Schnatter and Frühwirth (2007) the density of the type I extreme value distribution can be approximated very accurately by a mixture of ten normal components

$$p_\epsilon(\epsilon) = \exp(-\epsilon - e^{-\epsilon}) \approx \sum_{r=1}^{10} w_r f_{\mathrm{N}}(\epsilon; m_r, v_r). \tag{10}$$

In this approximation the weights w_r, means m_r and the variances v_r are fixed, therefore only the component indicators $r_{ij} \in \{1, \ldots, 10\}$ have to be introduced as a second sequence of latent variables to obtain the Gaussian state space model

$$-\ln \tau_{ij} = \mathbf{z}_i' \boldsymbol{\beta}_j + m_{r_{ij}} + \varepsilon_{r_{ij}}, \quad \varepsilon_{r_{ij}} \sim N\left(0, v_{r_{ij}}\right).$$

Let n_j denote the number of individuals at risk at the beginning of interval I_j, and assume that the observation times are arranged decreasingly, so that y_1 is the largest and y_n is the smallest observed time. Defining a multivariate observation vector $\mathbf{x}_j$ of dimension n_j as

$$\mathbf{x}_j = \begin{pmatrix} -\ln \tau_{1j} - m_{r_{1j}} \\ \vdots \\ -\ln \tau_{n_j,j} - m_{r_{n_j,j}} \end{pmatrix},$$

and $\boldsymbol{\varepsilon}_j = (\varepsilon_{r_{1j}}, \ldots, \varepsilon_{r_{n_j,j}})$ the model can be written in the following linear Gaussian state space form:

$$\mathbf{x}_j = \mathbf{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{\varepsilon}_j \sim N\left(\mathbf{0}, \mathbf{V}_j\right), \tag{11}$$

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j, \quad \boldsymbol{\omega}_j \sim N\left(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})\right), \tag{12}$$

where $\mathbf{V}_j = \mathrm{diag}(v_{r_{1j}}, \ldots, v_{r_{n_j,j}})$ and $\mathbf{Z}_j$ is an $n_j \times (K+1)$ matrix containing the design vectors $\mathbf{z}_1, \ldots, \mathbf{z}_{n_j}$ for all individuals at risk at time s_{j-1} as its rows

$$\mathbf{Z}_j = \begin{pmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_{n_j}' \end{pmatrix}.$$

Thus, instead of the original dynamic survival model we arrive at a partially Gaussian state space model, where the transition equation is the same as for the original model. The model for the log-hazard however is replaced by a Gaussian observation equation with a multivariate response vector $\mathbf{x}_j$, where $\mathbf{x}_j$ is determined from the auxiliary survival times of the risk population at the beginning of interval I_j.

3.3 Prior distributions

Priors have to be chosen for the initial value of the state vector $\boldsymbol{\beta}_0$ and the process variances. We assume independent normal priors for the elements of $\boldsymbol{\beta}_0$

$$\beta_{k0} \sim N\left(b_k, B_k\right)$$

and independent inverse Gamma priors for the process variances

$$\theta_k \sim \mathrm{IG}\left(c_{k0}, C_{k0}\right), \quad k = 0, \ldots, K.$$

3.4 Sampling scheme

With this prior choice a simple MCMC sampling scheme combining data augmentation with Bayesian estimation of Gaussian state space models as in Frühwirth-Schnatter (1994) can be implemented to sample jointly the latent process $\boldsymbol{\beta}$, the variances $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_K)$ and the auxiliary variables $\mathbf{x}_j$ and $\mathbf{R}_j = (r_{1j}, \ldots, r_{n_j,j})$ for $j = 1, \ldots, J$:

Select starting values for $\boldsymbol{\theta}$ and the augmented variables $\mathbf{x}_j$ and $\mathbf{R}_j$ for $j = 1, \ldots, J$ and repeat the following steps:

(a) Sample the latent states $\boldsymbol{\beta}$ conditional on the process variances $\boldsymbol{\theta}$ and the auxiliary variables $\mathbf{x}_j$ and $\mathbf{R}_j$: Sample the whole sequence $\boldsymbol{\beta}$ by forward-filtering backward sampling (FFBS, Frühwirth-Schnatter (1994); Carter and Kohn (1994); de Jong and Shephard (1995)) for the conditionally Gaussian state space form (11) and (12).

(b) Sample $\theta_k, k = 0, \ldots, K$ conditional on $\boldsymbol{\beta}$ from the inverse Gamma distribution $\mathrm{IG}\left(c_k, C_k\right)$, where

$$c_k = c_{k0} + J/2$$
$$C_k = C_{k0} + \sum_{j=1}^{J} (\beta_{kj} - \beta_{k,j-1})^2$$

(c) Sample the auxiliary variables $\mathbf{x}_j$ and $\mathbf{R}_j$ conditional on $\boldsymbol{\beta}$

(c1) Sample the auxiliary variables τ_{ij} conditional on the corresponding hazard $\lambda_{ij} = \exp(\mathbf{z}_i'\boldsymbol{\beta}_j)$, $i = 1, \ldots, n_j; j = 1, \ldots J$ as described in equations (6) - (7).

(c2) Sample the component indicators r_{ij} conditional on τ_{ij} and λ_{ij} from the following discrete distribution

$$\Pr\{r_{ij} = r^* | \tau_{ij}, \lambda_{ij}\} \propto w_{r^*}\, \varphi(\ln \tau_{ij} - \ln \lambda_{ij}; m_{r^*}, s^2_{r^*}), \quad r^* = 1, \ldots, 10.$$

Here $\varphi(x; \mu, \sigma^2)$ denotes the probability density function of the $N\left(\mu, \sigma^2\right)$ distribution at x, see Frühwirth-Schnatter and Wagner (2006a) for details.

Starting values for $\mathbf{x}_j$ and $\mathbf{R}_j$ can be obtained by performing sampling step (b) with a starting value λ^0_{ij} for the hazards. In the applications I used the same hazard rate for all subjects at risk in an interval, $\lambda^0_{ij} = \lambda^0_j$, where λ^0_j was determined as the number of failures in interval I_j divided by the sum of observation times spent in I_j.

4 Model Selection

Model selection in the normal dynamic survival model means not only to select which covariate to include in the model but also to decide whether the effect of a certain covariate is constant or varies over time. To perform model selection we use a non-centered parameterization of the augmented state space model defined in equations (11) and (12).

4.1 The noncentered parameterization

Let $\mathbf{b}_j = \mathbf{Q}^{-1/2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)$ denote the standardized latent process. Obviously $\mathbf{b}$ starts in $\mathbf{b}_0 = \mathbf{0}$ and its stochastic evolution is described by

$$\mathbf{b}_j = \mathbf{b}_{j-1} + \tilde{\boldsymbol{\omega}}_j \qquad \tilde{\boldsymbol{\omega}}_j \sim N(\mathbf{0}, \mathbf{I}) \tag{13}$$

Using the standardized state vector $\mathbf{b}_j$ observation equation (11) can be written as

$$\mathbf{x}_j = \mathbf{Z}_j\boldsymbol{\beta}_0 + \mathbf{Z}_j\mathbf{Q}^{1/2}\mathbf{b}_j + \boldsymbol{\varepsilon}_j, \quad \boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \mathbf{V}_j). \tag{14}$$

This noncentered parameterization is not identified as in the observation equation

$$x_{ij} = \sum_{k=0}^{K} z_{ik}\beta_{k0} + \sum_{k=0}^{K} z_{ik}b_{kj}(\pm\sqrt{\theta_k}) + \varepsilon_{ij} \tag{15}$$

for each $k = 0, \ldots, K$ the sign of $\sqrt{\theta_k}$ and the sequence $b_{kj}, j = 1, \ldots J$ may be changed without changing the likelihood.

As discussed in Frühwirth-Schnatter and Wagner (2008) the likelihood function of the noncentered parameterization of a state space model is bimodal in the direction of a process standard deviation if the respective component of the state vector is stochastic, and symmetric with a mode at 0 for a constant component. This means that $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{b})$ will be bimodal in the direction of $\sqrt{\theta_k}$ if the effect of covariate k varies over time and symmetric with a mode $\sqrt{\theta_k} = 0$ at if the effect is constant. Note that non-identifiability of the noncentered parameterization concerns only the sign of a component of the state vector and the corresponding process standard deviation: the effect of covariate k in each interval j given as $\beta_{k0} + b_{kj}(\pm\sqrt{\theta_k})$ is identified. However, this non-identifiability has to be taken into account in the MCMC sampling scheme to guarantee exploration of the whole posterior space.

4.2 The parsimonious normal dynamic survival model

In the observation equation of the noncentered parameterization (15) the mean of the auxiliary variable x_{ij} is equal to the log-hazard $\log\lambda_{ij}$ of subject i in interval I_j. This log hazard has two components, the first resulting from the initial covariate effects β_{k0}, the second from the modification of these initial effects in interval I_j, $b_{kj}(\pm\sqrt{\theta_k})$

To perform model selection we introduce for each covariate effect $k = 1, \ldots K$ two indicator variables δ_k and γ_k. The first of these indicators δ_k is 0 iff the initial effect $\beta_{0k} = 0$ and thus selects the effect of covariate z_k at time 0. The second indicator γ_k is defined as $\gamma_k = 0$ iff $\theta_k = 0$ and selects effects that vary over time.

We will include a constant baseline hazard effect in each model, hence β_{00} is not selected. However an additional indicator γ_0, defined as $\gamma_0 = 0$ iff $\theta_0 = 0$, is introduced to select between a constant and a time-varying baseline hazard. The final model for variable selection is then given as:

$$x_{ij} = \beta_{00} + \sum_{k=1}^{K} \delta_k z_{ik} \beta_{k0} + \sum_{k=0}^{K} \gamma_k z_{ik} b_{kj} (\pm\sqrt{\theta_k}) + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N\left(0, v_{r_{ij}}\right) \tag{16}$$

$$b_{kj} = b_{k,j-1} + \tilde{\omega}_{kj}, \qquad \tilde{\omega}_{kj} \sim N(0,1). \tag{17}$$

The variable selection model can be written as the following state space model

$$\mathbf{x}_j = \mathbf{Z}_j(\boldsymbol{\delta})\boldsymbol{\beta}_0 + \mathbf{H_j}(\boldsymbol{\gamma})\mathbf{b}_j + \boldsymbol{\varepsilon}_j, \qquad \boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \mathbf{V}_j) \tag{18}$$

$$\mathbf{b}_j = \mathbf{b}_{j-1} + \tilde{\boldsymbol{\omega}}_j, \qquad \tilde{\boldsymbol{\omega}}_l \sim N(\mathbf{0}, \mathbf{I}) \tag{19}$$

where $\mathbf{Z}_j(\boldsymbol{\delta})$ and $\mathbf{H}_j(\boldsymbol{\gamma})$ depend on the indicators and are given as

$$\mathbf{Z}_j(\boldsymbol{\delta}) = \mathbf{Z}_j \operatorname{diag}(1, \boldsymbol{\delta}) \qquad \mathbf{H}_j(\boldsymbol{\gamma}) = \mathbf{Z}_j \mathbf{Q}^{1/2} \operatorname{diag}(\boldsymbol{\gamma}).$$

4.3 Priors

To complete model specification prior distribution for all unknown model parameters $(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\beta}_0, \boldsymbol{\theta})$ have to be chosen.

4.3.1 Prior distribution for the indicators

In the variable selection model (16) – (17) there is a choice between two modeling options for the baseline log-hazard namely constant, $\gamma_0 = 0$, or time-varying, $\gamma_0 = 1$. For each covariate $k = 1, \ldots, K$ there are three different modeling options which are interesting from a practical point of view: no effect of the covariate, a constant effect or a time-varying effect. We identify these cases by the values $\zeta_1 = (0,0)$, $\zeta_2 = (1,0)$ and $\zeta_3 = (1,1)$ for the indicator pair (δ_k, γ_k).

Assuming independence of indicators for $k = 0, 1, \ldots, K$ leads to a prior

$$p(\boldsymbol{\delta}, \boldsymbol{\gamma}) = p(\gamma_0) \prod_{k=1}^{K} p(\delta_k, \gamma_k).$$

Let $p(\gamma_0 = 1) = \eta_0$ and $p((\delta_k, \gamma_k) = \zeta_l) = \eta_l$ for $l = 1, \ldots, 3$. As a first prior, we consider fixed prior probabilities $\eta_0 = \frac{1}{2}$ and $\eta_l = \frac{1}{3}, l = 1, \ldots, 3$. This prior, denoted as prior 1 in the following, assigns equal probability $p = 0.5 \cdot 3^{-K}$ to all models under consideration.

A second prior is obtained by putting a hyper-prior on the inclusion probabilities as in Smith and Kohn (2002) and Frühwirth-Schnatter and Tüchler (2008). For model selection in regression models Ley and Steel (2007) showed that a prior where the inclusion probability is random clearly outperforms priors with fixed inclusion probabilities. Conjugate priors are a Beta prior for η_0 and a Dirichlet prior for (η_1, η_2, η_3). For $\eta_0 \sim \text{Beta}(1,1)$ and $(\eta_1, \eta_2, \eta_3) \sim \text{Dirichlet}(1,1,1)$ the resulting prior is

$$p(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \frac{1}{2} \cdot \frac{2 \prod_{l=1}^{3} \Gamma(h_l + 1)}{\Gamma(K+3)}$$

where $h_0 = \gamma_0$ and $h_l = \sum_{k=1}^{K} I_{\{(\delta_k,\gamma_k)=\zeta_l\}}$. This prior is uniform on (h_0, h_1, h_2, h_3) as

$$p(h_0, h_1, h_2, h_2) = \binom{K}{h_1\ h_2\ h_3} p(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \frac{1}{(K+1)(K+2)}.$$

As prior 3 we specify a prior of the form $p(\delta_k, \gamma_k) = p(\delta_k)p(\gamma_k|\delta_k)$ where $\Pr\{\gamma_k = 1|\delta_k = 0\} = 0$. This prior somehow reflects the strategy of Hofner et al. (2008), who in a first step select covariates into the model and in a second step for each of the selected covariate choose a modeling alternative (fixed or time-varying effect). A flexible prior is obtained by assuming that

$$\begin{aligned}
&\Pr\{\delta_k = 1|\eta_\delta\} = \eta_\delta, \quad k = 1, \dots K \\
&\Pr\{\gamma_k = 1|\delta_k = 1, \eta_\gamma\} = \eta_\gamma, \quad k = 1, \dots K \\
&\Pr\{\gamma_0 = 1|\eta_\gamma\} = \eta_\gamma.
\end{aligned}$$

If both hyper-parameters η_δ and η_γ are iid Uniform on [0,1], the resulting prior is $p(\boldsymbol{\delta}, \boldsymbol{\gamma}) = p(\boldsymbol{\delta})p(\boldsymbol{\gamma}|\boldsymbol{\delta})$ where

$$\begin{aligned}
p(\boldsymbol{\delta}) &= B(1 + \sum_{k=1}^{K} \delta_k, 1 + K - \sum_{k=1}^{K} \delta_k) \\
p(\boldsymbol{\gamma}|\boldsymbol{\delta}) &= B(1 + \gamma_0 + \sum_{k:\delta_k=1} \gamma_k, 1 + (1 - \gamma_0) + \sum_{k:\delta_k=1} (1 - \gamma_k)),
\end{aligned}$$

if $B(\cdot, \cdot)$ is the Beta function. Note that this prior leads to a uniform distribution over the number of covariates included for δ

$$p(\sum_{k=1}^{K} \delta_k = h_\delta) = \frac{1}{K+1}.$$

The resulting conditional prior for $p(\boldsymbol{\gamma}|\boldsymbol{\delta})$ is a uniform distribution over the number of regressors with potentially time-varying effect. As the log-baseline hazard is included in each model, the number of these regressors is $\sum \delta_k + 1$ and

$$p(\gamma_0 + \sum_{k:\delta_k=1} \gamma_k = h_\gamma|\boldsymbol{\delta}) = \frac{1}{\sum_{k=1}^{K} \delta_k + 2}.$$

4.3.2 Prior distributions for the effects

Conditional on the state vector, the observation equation (16) of the variable selection model defines a Gaussian regression model with heteroscedastic errors and known error variances. Denoting by $\boldsymbol{\alpha}$ the parameter vector

$$\boldsymbol{\alpha} = (\beta_{00}, \dots, \beta_{0K}, \pm\sqrt{\theta_0}, \dots, \pm\sqrt{\theta_K}),$$

this regression model can be written as

$$x_{ij} = (\mathbf{w}_{ij}^{\boldsymbol{\delta},\boldsymbol{\gamma}})'\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, v_{r_{ij}}\right) \tag{20}$$

If all indicators take the value 1 then $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}} = \boldsymbol{\alpha}$ and the design vector $\mathbf{w}_{ij}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ is of the form

$$\mathbf{w}_{ij}' = \Big(1, z_{i1}, \ldots, z_{iK}, b_{0j}, z_{i1}b_{1j}, \ldots, z_{iK}b_{Kj}\Big).$$

Otherwise the restricted parameter vector $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ and the predictor vector $\mathbf{w}_{ij}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ contain only those elements for which the corresponding indicator is equal to 1.

Let $\mathbf{x}$, $\mathbf{R}$ and $\boldsymbol{\varepsilon}$ denote the vectors obtained by stacking the vectors $\mathbf{x}_j$, $\mathbf{R}_j = (r_{1j}, \ldots, r_{n_j,j})$ and $\boldsymbol{\varepsilon}_j$, $j = 1, \ldots, J$, $\mathbf{V}$ the diagonal matrix containing the variances of the error terms and $\mathbf{W}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ the regressor matrix with rows equal to $(\mathbf{w}_{ij}^{\boldsymbol{\delta},\boldsymbol{\gamma}})'$ in appropriate order. Using this notation the regression model for all $N = \sum_{j=1}^{J} n_j$ auxiliary variables can be written as

$$\mathbf{x} = \mathbf{W}^{\boldsymbol{\delta},\boldsymbol{\gamma}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \mathbf{V}).$$

Conditional on the indicators $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ we specify a fractional prior with fraction $b = N$ for $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$, as in Frühwirth-Schnatter and Tüchler (2008):

$$p(\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}|\mathbf{b}, \mathbf{x}, \mathbf{R}) = N\left(\mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}, (\mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}})\frac{1}{b}\right), \tag{21}$$

where

$$(\mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}})^{-1} = (\mathbf{W}^{\boldsymbol{\delta},\boldsymbol{\gamma}})'\mathbf{V}^{-1}\mathbf{W}^{\boldsymbol{\delta},\boldsymbol{\gamma}} \tag{22}$$

$$\mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}} = \mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}(\mathbf{W}^{\boldsymbol{\delta},\boldsymbol{\gamma}})'\mathbf{V}^{-1}\mathbf{x} \tag{23}$$

4.4 Sampling scheme

The noncentered parameterization together with the prior choices allows for a simple MCMC sampling scheme to sample jointly the indicators $(\boldsymbol{\delta}, \boldsymbol{\gamma})$, the unrestricted elements of the parameter vector $\boldsymbol{\alpha}$, the state process $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_J)$ and the auxiliary variables $\mathbf{x}$ and $\mathbf{R}$.

The sampling scheme consists of the following steps:

(a) Sample the indicator pair (δ_k, γ_k) from

$$p(\delta_k, \gamma_k|\mathbf{b}, \mathbf{x}, \mathbf{R}, \boldsymbol{\delta}_{\backslash k}, \boldsymbol{\gamma}_{\backslash k}) \propto p(\mathbf{x}|\boldsymbol{\delta}, \boldsymbol{\gamma}, \mathbf{b}, \mathbf{R})p(\delta_k, \gamma_k, \boldsymbol{\delta}_{\backslash k}, \boldsymbol{\gamma}_{\backslash k})$$

where $\boldsymbol{\delta}_{\backslash k}$ and $\boldsymbol{\gamma}_{\backslash k}$ denote the elements of the indicator vectors except δ_k and γ_k.

(b) Sample all *unrestricted* elements of the initial values of $\boldsymbol{\beta}_0$ and all *unrestricted* variance parameters $\sqrt{\theta_k}$ jointly from the multivariate normal distribution posterior $N\left(\mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}, \mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}\right)$ conditional on the state vector $\mathbf{b}$ and the auxiliary variables $(\mathbf{x}, \mathbf{R})$.

Set all remaining initial values of $\boldsymbol{\beta}_0$ and all remaining variances equal to 0.

(c) Sample $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_J)$ from the state space form $p(\mathbf{b}|\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{x}, \mathbf{R})$ given in equations (18) and (19).

(d) For each k=0, ..., K: Perform a random sign switch for $\sqrt{\theta_k}$ and $b_{kj}, j = 1, \ldots, J$

(e) Sample the auxiliary variables $\mathbf{x}_j$ and $\mathbf{R}_j$ conditional on $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ and $\mathbf{b}_j$ as in subsection 3.4.

Details on sampling steps (a) and (c) are given below.

4.4.1 Sampling the indicators and the unrestricted parameters

Sampling the indicators $(\boldsymbol{\delta}, \boldsymbol{\gamma})$ requires to marginalize over the parameters which are subject to model selection, see George and McCulloch (1997) for a full account. This is feasible for the noncentered parameterization of the normal dynamic survival model as the auxiliary mixture sampling scheme leads to a conditionally Gaussian regression model, where the marginal likelihood can be derived analytically.

To sample the indicators marginally with respect to $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ we combine the prior (21) with the remaining $(1-b)$ proportion of the likelihood $p(\mathbf{x}|\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}, \mathbf{b}, \mathbf{R})^{(1-b)}$. Integration with respect to $\boldsymbol{\alpha}^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ yields the marginal likelihood

$$p(\mathbf{x}|\boldsymbol{\delta}, \boldsymbol{\gamma}, \mathbf{b}, \mathbf{R}) = b^{(q_\delta + q_\gamma)/2} \left(\frac{|\mathbf{V}|^{-1}}{(2\pi)^N} \right)^{(1-b)/2} \cdot \exp\left(-\frac{(1-b)}{2} \mathbf{S} \right),$$

where

$$\mathbf{S} = \mathbf{x}' \mathbf{V}^{-1} \mathbf{x} - (\mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}})' (\mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}})^{-1} \mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}),$$

q_δ and q_γ are the number of nonzero elements in $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ respectively and $\mathbf{a}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ and $\mathbf{A}_N^{\boldsymbol{\delta},\boldsymbol{\gamma}}$ are given in (23) and (22).

For sampling the indicators a random order for updating the indicator pairs (δ_k, γ_k) is used.

4.4.2 Sampling the state vector

In sampling step (c) forward-filtering-backward-sampling (FFBS, Frühwirth-Schnatter (1994); Carter and Kohn (1994); de Jong and Shephard (1995)) is used to sample the state vector $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_J)$. If at least one indicator $\gamma_k = 0$ a reduced state space form is used. As the observation equation is independent of any component of the state vector $\mathbf{b}$ where the corresponding $\gamma_k = 0$, FFBS is applied to the reduced state vector of the components where $\gamma_k = 1$ and all other components are sampled from the prior (17).

5 Case studies

5.1 Gastric cancer data

As a first application I use 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 was applied. Overall 10 observation times were right censored.

Table 1: Gastric data: Estimated process variances (centered parameterization)

Parameter	Mean	Std.dev.	95%H.P.D. intervals
θ_0	0.0240	0.0251	[0.0020 0.0651]
θ_1	0.0536	0.0506	[0.0030 0.1514]

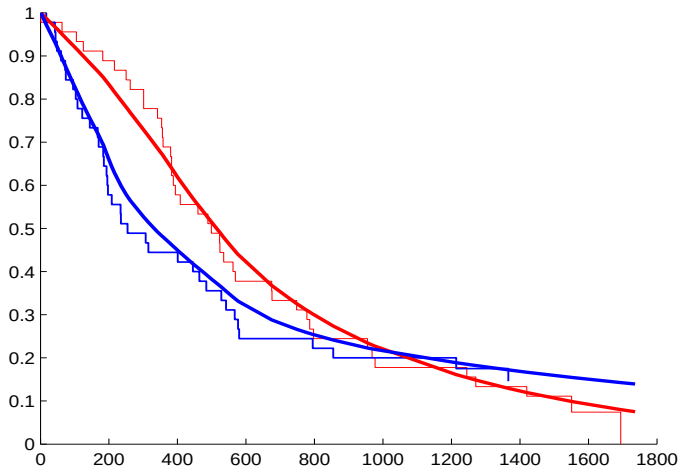

Figure 2: Gastric data: Estimated survival function and Kaplan-Meier-estimates for the treatment group (blue) and the control group (red)

In this setting it is of interest whether there is a treatment effect of combined therapy as compared to chemotherapy alone (which serves as control) and whether this effect varies over time. The covariate vector is $\mathbf{z}_i = (1, z_{1i})$ where z_{1i} indicates whether patient i underwent combined therapy or not. The parameter vector $\boldsymbol{\beta}$ therefore has 2 components. The fact that the estimated survival functions for treatment and control group cross (see Figure 2), indicates that a Cox model with constant treatment effect might not be appropriate.

5.1.1 Fitting a dynamic survival model

To fit a dynamic survival model, the failure times in the data were used as division points of the time axis, as in Gamerman (1991) and Hemming and Shaw (2005). The auxiliary mixture sampler of section 3.4 was run for 50000 iterations after a burn-in of 20000. Prior moments for the initial values $\boldsymbol{\beta}_0$ were $b_k = 0$ and $B_k = 100, k = 0, 1$. For the process variances proper, but uninformative inverse Gamma priors with parameters $c_{0k} = C_{0k} = 0.01, k = 0, 1$ were chosen.

Table 1 reports point estimates and standard errors as well as 95%-highest posterior density intervals for the process variances θ_0 and θ_1 of baseline log-hazard and treatment effect. As already noted in previous analyses the process variance of the treatment effect is higher than for the baseline log-hazard.

The estimated survival functions for the fitted normal dynamic survival model

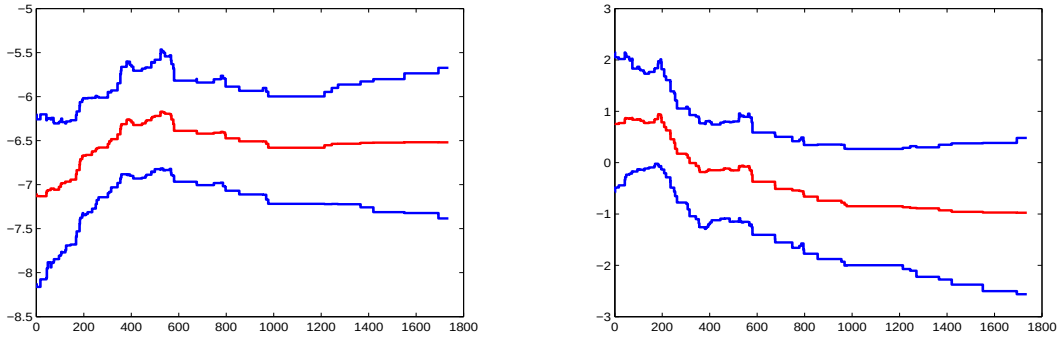

Figure 3: Gastric data: Posterior means and 95% credible regions for the baseline log-hazard (left) and the effect of combined treatment (right)

.

are compared to the Kaplan-Meier estimates in Figure 2.

The estimated baseline log-hazard and the effect of combined therapy are shown in Figure 3. The baseline log-hazard increases early followed by a sharp decline. The estimated treatment effect declines from an effect with positive sign to a negative sign 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. As the credible intervals include a straight line, a constant baseline hazard, as assumed in Gamerman (1991) and a constant treatment effect cannot be ruled out.

5.1.2 Unrestricted noncentered parameterization

A fit of the unrestricted noncentered parameterized model provides a useful tool for exploratory analysis as the posterior of process standard deviations gives insight whether an effect is time-varying or not. To fit the noncentered parametrized model sampling steps (b) - (e) of the sampler described in section 4.4 were run for 50000 iterations after a burnin of 20000. As convergence of the sampler was rather slow, a coarser division of the time axis was used, where every fifth failure time defines an interval endpoint. Histograms of the posterior densities for the process standard deviations $\pm\sqrt{\theta_i}, i = 0, 1$ are plotted in Figure 4. Both posteriors are bimodal, pointing to a model where both baseline log-hazard and treatment effect vary over time.

5.1.3 Stochastic Model Specification Search

Stochastic model specification search was carried out using the fractional prior and three different priors for the model indicators discussed in section 4.3.1. MCMC sampling was carried out for $M = 100000$ draws after a burn-in of 20000 draws. The first 10000 draws of the burn-in were drawn from the unrestricted model, model selection began after these first 10000 draws.

Results of the variable selection procedure are summarized in Table 2 and 3. Note that prior 1 and prior 2 in this case where only the effect of one covariate is

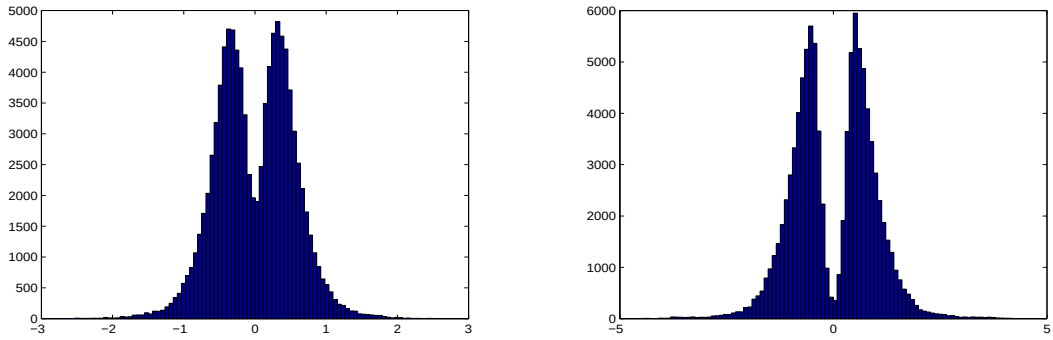

Figure 4: Gastric data: posterior densities of $\pm\sqrt{\theta_0}$ (left) and $\pm\sqrt{\theta_1}$ (right)

Table 2: Gastric data: number of draws from each model

model	δ_1	γ_0	γ_1	prior 1	prior 3
1	0	0	0	5534	10300
2	0	1	0	6992	10089
3	1	0	0	2218	2215
4	1	0	1	25236	13163
5	1	1	0	2739	1565
6	**1**	**1**	**1**	**57281**	**62668**

considered coincide and all 6 models considered have equal prior probability. This is not the case for prior 3, which assigns a prior probability of 1/4 to models 1 and 2, 1/6 to models 3 and 6 and 1/12 to models 4 and 5. The most frequently visited model in Table 2 is robust against the prior choice, but the frequency with which this model is selected varies, in particular models 4 and 5 are visited less frequently under prior 3.

If the selected model is defined as comprising all parameters where the posterior mean of the corresponding indicator is greater than 0.5, the same model, namely the model where both log-baseline and treatment effect are time-varying, results for both priors.

5.2 Worcester heart attack data

As a further application I analyzed the data of the Worcester heart attack study described in Hosmer and Lemeshow (1999). The main goal of the study was to describe trends over time in the incidence and survival rates following hospital ad-

Table 3: Gastric data: posterior means of indicators

Prior	δ_1	γ_0	γ_1
1	0.8747	0.6701	0.8252
3	0.7961	0.7432	0.7583

Table 4: Description of variables in the Worcester heart attack study data

Variable	Description
`age`	age at hospital admission in years (centered at 67 years)
`gender`	0=male, 1=female
`sho`	cardiogenic shock complications (0=no,1=yes)
`cpk`	peak cardiac enzyme measured in international units (IU)
`chf`	left heart failure complications (0=no, 1=yes)
`miorder`	myocardial infection order (0=first, 1=recurrent)
`mitype`	myocardial infection type (0=Q wave, 1=not Q wave or indeterminate)

Table 5: WHAS data: time intervals and number of events in month 1.

j	1	2	3	4	5	6	7	8	9
interval endpoint s_j (in days)	1	2	3	4	7	10	14	21	30
number of events	8	16	10	7	7	14	12	12	7

mission 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 4.

5.2.1 Fitting a dynamic survival model in centered parameterization

As some of the covariates are related to the myocardial infarction, a time-varying effect for these covariates might be more plausible than a simple proportional hazards model. In a first analysis a dynamic survival model with time-varying effects for all covariates was fitted. As many deaths occur early, the time axis was partitioned into intervals of increasing lengths, starting with a finer division of one day length, followed by intervals of half a week, a week, one month, 3 months, half a year to intervals of one year length (from year 3 to year 10). The number of events in these time intervals ranges from 3 (in year 10) to 18 (in year 6). Table 5 summarizes the division of the time axis and the number of events for the first month.

For the initial values of the random walks (for baseline log-hazard and covariate effects), independent normal distributions with mean 0 and prior variance 100 were chosen. For the process variances inverse Gamma priors with parameters $c_0 = 0.1$ and $C_0 = 0.01$ were used. The auxiliary mixture sampler for the centered parameterization was run for 50 000 iterations after a burn-in of 20 000.

Posterior means of log-baseline and covariate effects are displayed in Figure 5 with 95% credible regions. Baseline log-hazard and most of the covariate effects show large changes in the beginning, many of them a sharp decline, and level off in the long run. Due to the general model specification credible intervals are rather wide making it hard to assess whether an effect is constant over time or not.

The effect of age is fairly constant with positive sign which means a negative effect, i.e. risk of death after myocardial infarction increases with age. Risk is higher for female than for male patients during the first 3 months, but approximates those of men later on. The adverse effect of cardiogenic shock complications (`sho`) shows

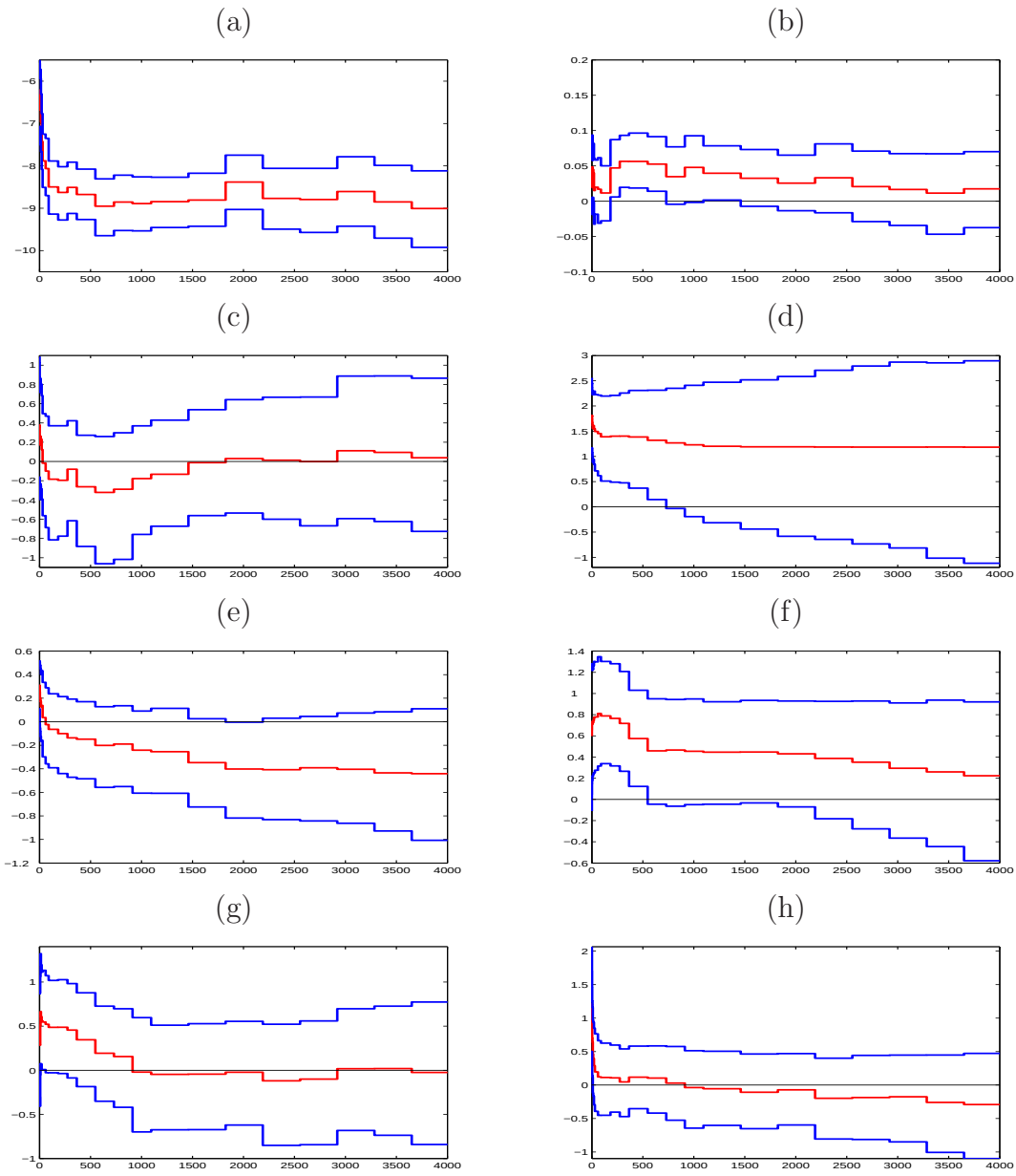

Figure 5: WHAS data: Posterior means and 95% credible regions for the baseline log-hazard (a) and the effects of `age` (b), `gender` (c), `sho` (d), `cpk` (e), `chf` (f), `miorder` (g) and `mitype` (h).

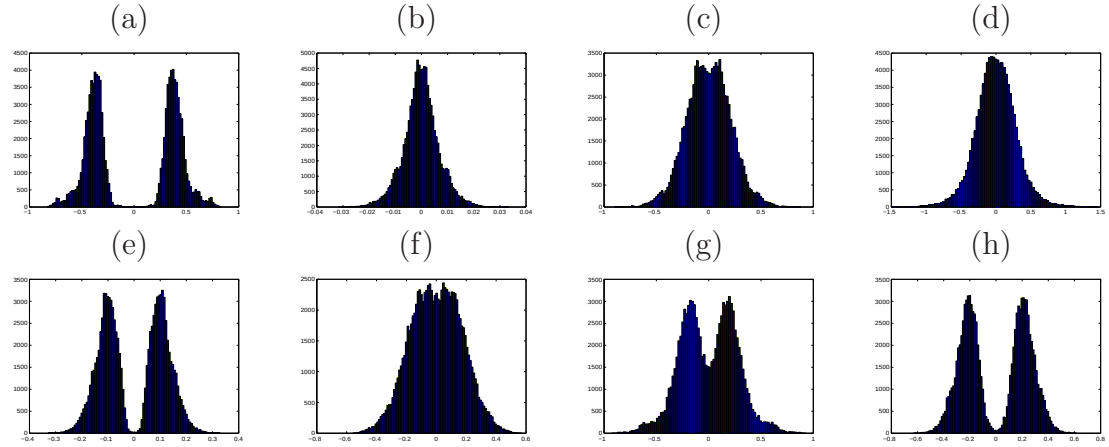

Figure 6: WHAS data: posterior densities of the process standard deviations $\pm\sqrt{\theta}$ for baseline log-hazard (a) and the effects of `age` (b), `gender` (c), `sho` (d), `cpk` (e), `chf` (f), `miorder` (g) and `mitype` (h).

a decline in the beginning but remains high even in the long run and is significantly positive until year 3 after infarction. For cardiac enzyme (`cpk`) there is evidence for a time-varying effect: higher values lead to a higher risk during the first 4 days, but a lower risk for those surviving more than 5 years with even a beneficial effect in the long run. Left heart failure complications (`chf`) have a negative effect (i.e. higher risk) during the first two years: Risk increases at the beginning, reaching its maximum in the third month, then declines. Persons experiencing a recurrent infarction `miorder` have a higher risk during the first three months, but this effect wears off in the long run. Finally for q-wave infarctions `mitype` there is sharp decline of the risk during the first 3 months which gives evidence to a time-varying effect.

5.2.2 Model Selection

As a next step the unrestricted model in the noncentered parametrization was fit under a fractional prior using steps (b)-(e) of the MCMC sampler described in section 4.4 with all indicators fixed to one. Histograms of the posterior densities (based on 100000 iterations after a burnin of 20000) for the process standard deviations $\pm\sqrt{\theta_k}, k = 0, \ldots, 7$ are shown in Figure 6.

The posterior of the process standard deviation is bimodal with two clearly separated modes for log-baseline, `cpk` and `mitype`. The posterior ordinate at $\pm\sqrt{\theta_k} = 0$ is practically zero, indicating a time-varying baseline and time-varying effects of `cpk` and `mitype`. For `miorder`, `gender` and `chf` the posterior for the process standard deviation has also two modes, but values of $\pm\sqrt{\theta}$ close to zero have a positive posterior probability. For `age` and `sho` the posterior is unimodal indicating a constant effect of these covariables.

For stochastic model specification search the MCMC sampling scheme of section 4.4 was run for $M = 200000$ draws after a burn-in of 20000 draws. The first 10000 draws of the burn-in were drawn from the unrestricted model, model selection began after these first 10000 draws. Results of the variable selection procedure under the

Table 6: WHAS data: posterior means of indicators

Prior		baseline	age	gender	sho	cpk	chf	miorder	mitype
1	δ		**1.00**	0.19	**1.00**	**0.93**	**1.00**	**0.77**	**0.90**
	γ	**1.00**	0.23	0.06	0.12	**0.90**	0.35	0.38	**0.81**
2	δ		**1.00**	0.46	**1.00**	**0.99**	**1.00**	**0.90**	**0.95**
	γ	**1.00**	0.28	0.18	0.16	**0.97**	0.38	**0.50**	**0.85**
3	δ		**1.00**	0.43	**1.00**	**0.90**	**0.99**	**0.89**	**1.00**
	γ	**1.00**	0.35	0.21	0.25	**0.87**	**0.51**	**0.60**	**0.97**

fractional prior and three different choices for the prior on the indicators discussed in Subsection 4.3.1 are summarized in Table 6.

For all priors the selected model (based on indicators with posterior mean $>$ 0.5) includes a time-varying log-baseline hazard and time-varying effects of `cpk` and `mitype`, whereas the effects of `age` and `sho` are selected as constant and the effect of `gender` is not selected at all. For `chf` and `miorder` the resulting specification is different for the three priors: under prior 1 a constant effect is selected for both variables, prior 2 selects a constant effect for `chf`, but a time-varying effect for `miorder`. Under prior 3 both effects are time-varying.

6 Concluding remarks

In this paper a new auxiliary mixture sampler for dynamic survival models is proposed, which is easy to implement and needs no tuning. Its convenience results from representing the log-hazard model as a partial Gaussian model for auxiliary variables which allows to deal with any form of the linear predictor where Gibbs sampling for the equivalent model with Gaussian errors is feasible. Sampling algorithms for Gaussian models are easily adapted to survival models with the same linear predictor by adding two steps of data augmentation. This is demonstrated for the noncentered parameterization of the log-hazard model, by implementing stochastic model specification search for state space models as proposed in Frühwirth-Schnatter and Wagner (2008). For survival data stochastic model specification search is attractive as additionally to variable selection the form of each effect, constant vs. time-varying has to be specified. Three different priors for the model indicators were proposed and investigated in the applications.

The auxiliary mixture sampler has a wider application, as e.g. inclusion of normal frailties, unstructured or structured normal 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. Also model specification search for random effects described for normal and logit models in Frühwirth-Schnatter and Tüchler (2008) and Tüchler (2008) could be incorporated easily for dynamic survival models.

As a further extension missing information different from right-censoring, e.g. interval interval censoring, can be dealt with by introducing complete auxiliary

survival times conditional on the available information.

Acknowledgements

I thank Sylvia Frühwirth-Schnatter for helpful discussions and comments and Christoph Pamminger for critical reading of the manuscript.

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.

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

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis*, 51:3509–3528.

Frühwirth-Schnatter, S. and Tüchler, R. (2008). Bayesian parsimonious covariance estimation for hierarchical linear mixed models. *Statistics and Computing*, 18:1–13.

Frühwirth-Schnatter, S. and Wagner, H. (2006a). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. *Biometrika*, 93:827–841.

Frühwirth-Schnatter, S. and Wagner, H. (2006b). Data augmentation and Gibbs sampling for regression models of time series of small counts. *Student*, 5:221–234.

Frühwirth-Schnatter, S. and Wagner, H. (2008). Stochastic model specification search for Gaussian and non-Gaussian models. IFAS research report, `http://www.ifas.jku.at`.

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

George, E. and McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

George, E. I. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.

Hemming, K. and Shaw, J. E. H. (2002). A parametric survival model applied to breast cancer survival times. *Applied Statistics*, 51:421–435.

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. *Journal of the American Statistical Association*, 101:1065–1075.

Hofner, B., Kneib, T., Hartl, W., and Küchenhoff, H. (2008). Model choice in Cox-type additive hazard regression modles with time-varying effects. Working paper. LMU München.

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

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

Konrath, S., Kneib, T., and Fahrmeir, L. (2008). Bayesian regularisation in structured additive regression models for survival data. Technical report.

Ley, E. and Steel, M. F. J. (2007). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. Technical report. Technical Report CRiSM Working Paper 07-08, University of Warwick.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics*, 140:425–449.

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.

Smith, M. and Kohn, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *Journal of the American Statistical Association*, 97:1141–1153.

Tüchler, R. (2008). Bayesian variable selection for logistic models using auxiliary mixture sampling. *Journal of Computational and Graphical Statistics*, 17:76–94.

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.