

A Bayesian Survival Model for Time-Varying Coefficients and Unobserved Heterogeneity

Knaus, Peter; Winkler, Daniel; Jomrich, Gerd

DOI:
[10.48550/arXiv.2206.11320](https://doi.org/10.48550/arXiv.2206.11320)

Published: 22/06/2022

[Link to publication](#)

Citation for published version (APA):

Knaus, P., Winkler, D., & Jomrich, G. (2022). *A Bayesian Survival Model for Time-Varying Coefficients and Unobserved Heterogeneity*. <https://doi.org/10.48550/arXiv.2206.11320>

A Bayesian Survival Model for Time-Varying Coefficients and Unobserved Heterogeneity

Peter Knaus Daniel Winkler Gerd Jomrich

June 24, 2022

Abstract

Dynamic survival models are a flexible tool for overcoming limitations of popular methods in the field of survival analysis. While this flexibility allows them to uncover more intricate relationships between covariates and the time-to-event, it also has them running the risk of overfitting. This paper proposes a solution to this issue based on state of the art global-local shrinkage priors and shows that they are able to effectively regularize the amount of time-variation observed in the parameters. Further, a novel approach to accounting for unobserved heterogeneity in the data through a dynamic factor model is introduced. An efficient MCMC sampler is developed and made available in an accompanying R package. Finally, the method is applied to a current data set of survival times of patients with adenocarcinoma of the gastroesophageal junction.

1 Introduction

The two most popular methods for analyzing survival data are the Cox proportional hazards model (Cox, 1972) and the Kaplan-Meier estimator (Kaplan and Meier, 1958). Despite their popularity, they both suffer from particular drawbacks. The Cox model assumes that the impact of covariates on the time-to-event does not change over time, while the Kaplan-Meier estimator is incapable of examining the effect of continuous covariates and suffers from the curse of dimensionality. An elegant solution in the Bayesian paradigm that aims to overcome these problems is proposed by Gamerman (1991) in the form of the piecewise exponential hazard model. The key idea behind this approach is to split the time axis into disjoint intervals, with a constant hazard rate within each interval. This allows it to flexibly adapt to any underlying functional form (given a fine enough grid of intervals), while still permitting a tractable likelihood at the level of the unit of observation.

Building on this approach, Wagner (2011) re-casts the model into a conditionally Gaussian state-space form using the ideas of Frühwirth-Schnatter and Frühwirth (2007a). The benefit of this is twofold: On the one hand, it allows for the use of established Markov chain Monte Carlo (MCMC) methods for estimation, side-stepping the need for more complex estimation strategies. On the other hand, it opens up the possibility to make use of the substantial literature on variance selection in time-varying parameter models to tackle a key challenge in such a highly parameterized model: overfitting.

Wagner specifically employs the stochastic variable selection approach (SSVS) (George and McCulloch, 1993, 1997) to select which covariates should have an effect that varies over time, in the spirit of the seminal paper by Frühwirth-Schnatter and Wagner (2010). An issue with this approach is that it can become computationally infeasible when the number of covariates becomes very large. In the context of time-varying parameter models, this limitation has been dealt with by making use of various different shrinkage

priors instead of SSVS (e.g. Belmonte et al. 2014, Hotz-Behofsits et al. 2018, Bitto and Frühwirth-Schnatter 2019, Cadonna et al. 2020, Huber et al. 2020).

Building on these advances, we propose a piecewise exponential hazard model using state of the art continuous shrinkage priors to regularize the amount of time variation we allow in the effects of the covariates. We also account for unobserved heterogeneity by adding a random-effect parameter in the form of a single latent factor, which also allows for time-variation. We complement the proposed method with the development of an efficient Gibbs sampler to estimate the model and implement it in a user-friendly R package, making estimation easily accessible to researchers looking for a flexible time-to-event model that requires little tuning. The package further provides the option to sample from the posterior predictive distribution of survival times given new covariates (which is side effect of being an absolute risk model and not a relative risk model), as well as the possibility to use time-varying covariates as inputs. Finally, we apply the model to a current data set of survival times of patients with adenocarcinoma of the gastroesophageal junction who underwent primary resection or were treated with neoadjuvant therapy before surgery and demonstrate that using our model can substantially improve the understanding of covariates’ effects on the time-to-event.

The remainder of the paper is structured as follows: Section 2 provides a discussion of the different components of the model and outlines the MCMC scheme used for estimation. Section 3 gives a brief overview of some of the functionality of the R package `shrinkDSM`. Section 4 presents an application to a clinical dataset, highlighting the benefits of the proposed approach, in particular vis-à-vis the Cox proportional hazards model. Finally, section 5 concludes.

2 Model specification and estimation

2.1 The normal dynamic survival model

A defining characteristic of survival data is that it is often subject to right censoring, meaning that the failure time of an individual is not observed. To account for this, each individual $i = 1, \dots, N$ is assumed to have both a survival time t_i and a censoring time c_i , which are independent random variables. The observed data of an individual then consists of the observed survival time $y_i = \min(c_i, t_i)$, coupled with a failure indicator d_i , which is equal to 0 if the observation is censored and 1 otherwise, and a vector of K covariates (z_{i1}, \dots, z_{iK}) . Gamerman (1991) extended the popular Cox proportional hazards model (Cox, 1972) by making both the baseline hazard β_0 as well as the effect of the covariates functions of time:

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

The benefit of this is twofold: the first and obvious advantage is that the effect of the covariates can change throughout the observed survival time. A second, somewhat less obvious benefit, is that the model changes from a relative risk model to an absolute risk model, allowing for the estimation of a baseline hazard as well as the prediction of survival times given new covariate values.

As proposed by Gamerman, the dynamic survival model is a piecewise exponential model for lifetimes. The underlying assumption is that a given set $\mathcal{S} = \{s_0 = 0, s_1, \dots, s_J\}$, $s_0 < s_1 < \dots < s_J$ partitions the time axis into J intervals $(s_0, s_1], \dots, (s_{J-1}, s_J]$, where the hazard within an interval is constant. More formally, the baseline log-hazard and co-

variate effects $\beta_k(t)$, for $k = 0, \dots, K$, are defined by the piecewise constant functions

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

This, in turn, implies that the hazard also has a piecewise constant form:

$$\lambda(t|\mathbf{z}_i; t \in (s_{j-1}, s_j]) = \lambda_{ij} = \exp \left(\beta_{0j} + \sum_{k=1}^K z_{ijk} \beta_{kj} \right). \quad (1)$$

In principle, as long as there is a value for each covariate for each interval and individual, the covariates can also vary over time, i.e.:

$$\lambda(t|\mathbf{z}_{ij}; t \in (s_{j-1}, s_j]) = \lambda_{ij} = \exp \left(\beta_{0j} + \sum_{k=1}^K z_{ijk} \beta_{kj} \right). \quad (2)$$

While the software introduced in section 3 supports this, nothing fundamental about the model itself changes and therefore, to simplify notation, it will be omitted.

An important component of the model is the choice of stochastic process that governs the evolution of the β_{kj} 's from one time interval to the next. Following Hemming and Shaw (2002), this paper considers the normal dynamic survival model with Gaussian random walks, i.e.:

$$\beta_{kj} = \beta_{k,j-1} + w_j, \quad w_j \sim \mathcal{N}(0, \theta_k), \quad (3)$$

where $\beta_{k0} \sim \mathcal{N}(\beta_k, \theta_k)$. The random walk specification assumes that neighboring intervals will not have state values that are too far from each other, thereby imposing smoothness on the evolution.

2.1.1 Variable and variance selection under the random walk

Three interesting special cases arise out of the random walk specification, depending on whether β_k and/or θ_k are equal to zero. The baseline is the case where both $\beta_k \neq 0$ and $\theta_k \neq 0$. This implies that the covariate associated with this parameter has an effect on the hazard and that this effect varies over time. The second case is when $\beta_k \neq 0$ but $\theta_k = 0$, resulting in a random walk where $\beta_{kj} = \beta_{k,j-1} = \dots = \beta_k$ with probability 1. This means that the corresponding covariate has an effect on the hazard rate, however this effect remains constant at β_k over time. The third and final interesting case is when both $\beta_k = 0$ and $\theta_k = 0$. In this case $\beta_{k0} = 0$ and all subsequent values $\beta_{kj} = \beta_{k0} = 0$ with probability 1. Thus the covariate has no effect on the hazard and this does not change over time, effectively selecting this covariate out of the model. A visual example of each of these cases is given in Figure 1. Given a tool that can automatically differentiate between these cases leads to a setup that can choose which covariates to include and which to allow to vary over time. This paper uses global-local shrinkage priors to this effect, in the spirit of Bitto and Frühwirth-Schnatter (2019) and Cadonna et al. (2020), among others. The exact priors used are discussed in section 2.3.

2.2 Likelihood

Under this model, the likelihood contribution of individual i can be expressed through conditional survival functions, see Hemming and Shaw (2005). This equates to factorizing the total contribution of individual i into the likelihood contributions from all periods survived fully, as well as the observed time survived in the period in which they exit the sample, beyond which they no longer contribute to the likelihood. Let u_{ij} denote the length of an interval survived fully by individual i (meaning it is equal

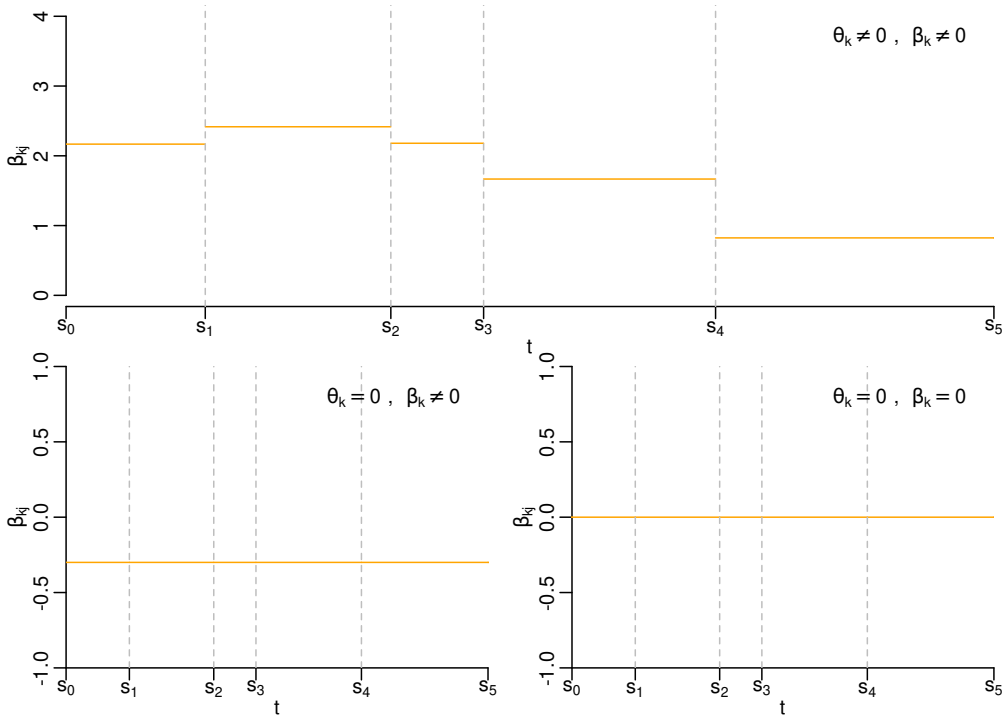


Figure 1: Exemplary evolution of the effect of three covariates on the hazard rate. The top figure is an example of a parameter β_{kj} associated with a covariate that has a time-varying effect, while the bottom left and bottom right display β_{kj} 's where the covariate's effect is significant and non time-varying and insignificant and non time-varying, respectively.

to $u_{ij} = s_j - s_{j-1}$) and u_{il} denote the amount of time the individual is observed in the interval they exit from the sample (making it equal to $u_{il} = y_i - s_{l-1}$). Clearly, the total observed survival time y_i can be reconstructed as the sum of these intervals $y_i = \sum_{j=1}^l u_{ij}$, see Figure 2 for a visual example.

Zooming in on the likelihood contribution of a fully survived time period for individual i , the probability of surviving said period is given by $1 - F(u_{ij})$, where $F(x)$ is the cumulative distribution function of the exponential distribution with rate λ_{ij} evaluated at u_{ij} . On the other hand, the likelihood contribution in the final period of individual i depends on the indicator d_i . If the observation is right censored (meaning failure is not observed, $d_i = 0$), then the contribution in this interval is of the same form as in the previous ones, i.e. $1 - F(u_{il})$. If, however, failure is observed, then the contribution comes in the form of the probability density function of an exponential distribution with parameter λ_{il} evaluated at u_{il} . Thus, the likelihood contribution of the final period can be written as

$$p(u_{il}, d_i | \lambda_{il}) = \exp(-\lambda_{il} u_{il}) \lambda_{il}^{d_i}.$$

The likelihood contribution of individual i is then equal to the product of the contributions in each interval:

$$L(y_i, d_i | \lambda_{i1}, \dots, \lambda_{il}) = \left[\prod_{j=1}^{l-1} \exp(-\lambda_{ij} u_{ij}) \right] \exp(-\lambda_{il} u_{il}) \lambda_{il}^{d_i}.$$

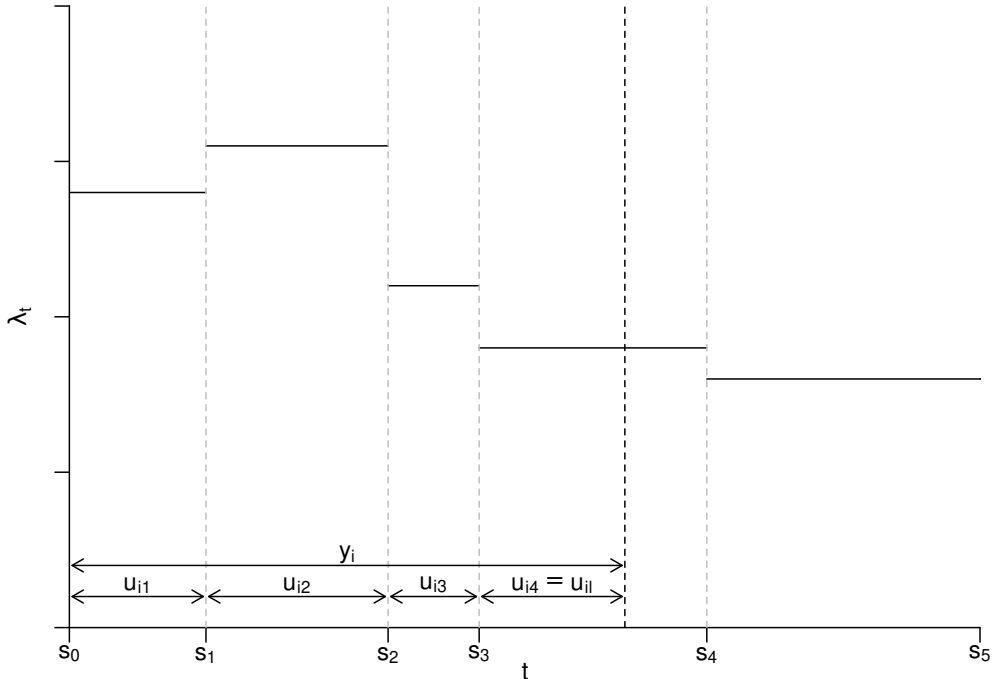


Figure 2: Hazard in the piecewise exponential model. y_i is the observed survival time of individual i , while u_{ij} is that individuals' observed survival in each interval. u_{il} denotes the time survived in the interval in which individual i exits the sample. In this example, the individual survives up until time period 4.

The likelihood for all observations is then simply the product over the individual likelihood contributions.

2.3 Prior specification

2.3.1 Priors on innovation variances and initial value means

As discussed in section 2.1.1, values of β_k and θ_k that are close to zero give rise to interesting special cases with regard to which covariate effects are included in the model and which are allowed to vary over time. Frühwirth-Schnatter and Wagner (2010) used this fact in the context of Gaussian state space models to perform model space MCMC for β_k and θ_k . As this becomes computationally burdensome in higher dimensions, the approach was adapted to employ global-local shrinkage priors instead of the spike-and-slab prior by Belmonte et al. (2014), Bitto and Frühwirth-Schnatter (2019) and Cadonna et al. (2020), among others. The possibility to use priors that were originally designed for *variable selection* for *variance selection* in the Gaussian state space context stems from an insight gleaned by Frühwirth-Schnatter and Wagner (2010), where it is shown that such state space models can be written in a non-centered parametrization in which $\sqrt{\theta_k} \in \mathbb{R}$. In this alternate parametrization, any prior defined for variable selection can be used to perform variance selection.

Following Cadonna et al. (2020), the triple gamma prior and its many special and limiting cases are considered in this paper. Note that, when used for variable selection, it is also known as the normal-gamma-gamma prior (Griffin et al., 2017). It has been shown

to have favorable theoretical properties and subsumes many popular prior choices, such as the horseshoe (Carvalho et al., 2009), the Bayesian Lasso (Park and Casella, 2008), the normal-gamma (Brown and Griffin, 2010) and a simple Gaussian prior. Interested readers are referred to Cadonna et al. (2020) for a more in-depth discussion, in particular due to the focus on state space models. For variance selection, the triple gamma prior placed on θ_k can be represented as a compound distribution consisting of three gamma distributions

$$\theta_k | \xi_k^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2\xi_k^2}\right), \quad \xi_k^2 | a^\xi, \kappa_k^2 \sim \mathcal{G}\left(a^\xi, \frac{a^\xi \kappa_k^2}{2}\right), \quad \kappa_k^2 | c^\xi, \kappa_B^2 \sim \mathcal{G}\left(c^\xi, \frac{c^\xi}{\kappa_B^2}\right),$$

depending on the global hyperparameters a^ξ , c^ξ and κ_B^2 . Similarly, a normal-gamma-gamma prior is placed on the mean of the initial value of the covariate effects,

$$\beta_k | \tau_k^2 \sim \mathcal{N}(0, \tau_k^2), \quad \tau_k^2 | a^\tau, \lambda_k^2 \sim \mathcal{G}\left(a^\tau, \frac{a^\tau \lambda_k^2}{2}\right), \quad \lambda_k^2 | c^\tau, \lambda_B^2 \sim \mathcal{G}\left(c^\tau, \frac{c^\tau}{\lambda_B^2}\right),$$

with hyperparameters a^τ , c^τ , and λ_B^2 . Letting c^ξ and c^τ go to infinity results in a double gamma prior on θ_k and a normal-gamma prior on β_k :

$$\begin{aligned} \theta_k | \xi_k^2 &\sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2\xi_k^2}\right), \quad \xi_k^2 | a^\xi, \kappa_B^2 \sim \mathcal{G}\left(a^\xi, \frac{a^\xi \kappa_B^2}{2}\right), \\ \beta_k | \tau_k^2 &\sim \mathcal{N}(0, \tau_k^2), \quad \tau_k^2 | a^\tau, \lambda_B^2 \sim \mathcal{G}\left(a^\tau, \frac{a^\tau \lambda_B^2}{2}\right). \end{aligned}$$

Finally, letting a^ξ and a^τ go to infinity results in a gamma prior on θ_k and a Gaussian prior on β_k :

$$\theta_k | \kappa_B^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{\kappa_B^2}{4}\right), \quad \beta_k | \lambda_B^2 \sim \mathcal{N}\left(0, \frac{2}{\lambda_B^2}\right).$$

Placing priors on the global parameters a^ξ , c^ξ , a^τ , c^τ , κ_B^2 , and λ_B^2 allows them to be learned from the data and gives the model the ability to adapt to the level of sparsity in any given dataset. In choosing these priors, we rely on results from Cadonna et al. (2020), which indicate that different prior distributions are beneficial depending on whether normal-gamma-gamma or normal-gamma priors are employed. In the case of the triple gamma/normal-gamma-gamma, the scaled global shrinkage parameters κ_B^2 and λ_B^2 conditionally follow F distributions, depending on a^ξ , c^ξ , a^τ and c^τ :

$$\frac{\kappa_B^2}{2} | a^\xi, c^\xi \sim F(2a^\xi, 2c^\xi), \quad \frac{\lambda_B^2}{2} | a^\tau, c^\tau \sim F(2a^\tau, 2c^\tau).$$

This choice of F distribution as priors is quite deliberate, as it leads to a uniform distribution on an appropriately defined model size. a^ξ , c^ξ , a^τ and c^τ , in turn, follow scaled beta distributions:

$$\begin{aligned} 2a^\xi &\sim \mathcal{B}(\alpha_{a^\xi}, \beta_{a^\xi}), \quad 2c^\xi \sim \mathcal{B}(\alpha_{c^\xi}, \beta_{c^\xi}), \\ 2a^\tau &\sim \mathcal{B}(\alpha_{a^\tau}, \beta_{a^\tau}), \quad 2c^\tau \sim \mathcal{B}(\alpha_{c^\tau}, \beta_{c^\tau}). \end{aligned}$$

These distributions force the parameters to stay in the range $(0, 0.5)$, thereby preserving the BMA-esque behavior of the triple gamma prior.

In the double gamma/normal-gamma case, κ_B^2 and λ_B^2 follow independent gamma priors:

$$\kappa_B^2 \sim \mathcal{G}(d_1, d_2), \quad \lambda_B^2 \sim \mathcal{G}(e_1, e_2).$$

For learning the parameters a^ξ and c^ξ we follow Knaus et al. (2021) and also employ gamma priors:

$$a^\xi \sim \mathcal{G}(\alpha_{a^\xi}, \alpha_{a^\xi} \beta_{a^\xi}), \quad a^\tau \sim \mathcal{G}(\alpha_{a^\tau}, \alpha_{a^\tau} \beta_{a^\tau}).$$

In this parametrization of the gamma distribution, the expected value depends only on β_{a^ξ} and β_{a^τ} , respectively, and α_{a^ξ} and α_{a^τ} act as degrees of freedom, bounding the prior away from zero.

2.4 Estimation procedure

Estimation of the unknown parameters is done through MCMC. We extend the approach proposed by Wagner (2011), which uses data augmentation to enable Gibbs sampling and avoid Metropolis-Hastings (MH) steps where possible. While this is feasible for most parameters, the conditional posteriors of a^ξ , c^ξ , a^τ and c^τ are not of well-known form and therefore require MH-within-Gibbs steps. The sampler also leans on ideas from Bitto and Frühwirth-Schnatter (2019) and Cadonna et al. (2020) for sampling steps of parameters associated with the shrinkage priors.

2.4.1 Data augmentation

A key ingredient of the MCMC sampler is the data augmentation scheme introduced by Wagner (2011). She shows that the model can be cast into a conditionally Gaussian state space form through the use of a mixture approximation of the standard Gumbel distribution (also known as the generalized extreme value distribution type-I). Beyond simply avoiding Metropolis-Hastings, this allows the use of one-block samplers for the β_{k_j} 's with tools such as forward-filtering, backward-sampling (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994) and related algorithms (e.g. McCausland et al. (2011)).

In a first step, each interval that does not end in failure is assumed to be right censored. To capture this, a residual survival time ξ_{ij} is added to each censored observation u_{ij} . This gives rise to the total survival time $\tau_{ij} = u_{ij} + \xi_{ij}$. Due to the memorylessness of the exponential distribution, the residual survival time also follows an exponential $\text{Ex}(\lambda_{ij})$ distribution, conditional on $\tau_{ij} > u_{ij}$. Thus one can write these auxiliary survival times as

$$\tau_{ij} = u_{ij} + \xi_{ij}, \quad \xi_{ij} \sim \text{Ex}(\lambda_{ij}) \text{ for } j = 1, \dots, l-1,$$

with the form of the final auxiliary survival time depending again on the indicator d_i :

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

Represented this way, the normal dynamic survival model can be viewed as a generalized dynamic linear model, where the observations are conditionally exponentially distributed:

$$\begin{aligned} \tau_{ij} | \boldsymbol{\beta}_j &\sim \text{Ex}(\exp(\mathbf{z}'_i \boldsymbol{\beta}_j)), \\ \boldsymbol{\beta}_j &= \boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j, \quad \boldsymbol{\omega}_j \sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})). \end{aligned}$$

In this notation, the baseline hazard is absorbed into $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{Kj})'$, implying that an intercept term is added to $\mathbf{z}_i = (1, z_{i1}, \dots, z_{iK})$. The variances of the innovations are collected in the diagonal matrix $\mathbf{Q}(\boldsymbol{\theta}) = \text{diag}(\theta_0, \theta_1, \dots, \theta_K)$, which also makes clearer the assumption that the random walks evolve independently of one another. Taking the log of the observation equation transforms it to be linear

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

where ϵ_{ij} now follows a standard Gumbel distribution. Frühwirth-Schnatter and Frühwirth (2007b) show that this distribution can be approximated very closely through a mixture of 10 normal distributions

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

The parameters of the mixture distribution, i.e. the weights w_r , the means m_r and the variances v_r , are determined by minimizing the Kullback-Leibler divergence from the approximating mixture to the Gumbel distribution. The exact values can be found in Frühwirth-Schnatter and Frühwirth (2007b, Table 1).

After introducing the component indicators $r_{ij} \in \{1, \dots, 10\}$, the model can be written in conditionally Gaussian state space form:

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

This can be made even more evident by gathering all observations that are still at risk in an interval together in one equation. To this end, let n_j be all observations that are still at risk in a given interval, meaning the observations that have not yet exited the sample due to either failure or censoring. Furthermore, assume that the vector of observed survival times \mathbf{y} is sorted from longest to shortest, with y_1 being the longest and y_N being the shortest, and that the rows of the design matrix \mathbf{Z} were ordered with the same key. Then, through defining a multivariate observation vector \mathbf{x}_j as

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

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

$$\begin{aligned} \mathbf{x}_j &= \mathbf{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, & \boldsymbol{\varepsilon}_j &\sim N(\mathbf{0}, \mathbf{V}_j), \\ \boldsymbol{\beta}_j &= \boldsymbol{\beta}_{j-1} + \boldsymbol{\omega}_j, & \boldsymbol{\omega}_j &\sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})), \end{aligned} \tag{4}$$

where $\mathbf{V}_j = \text{diag}(v_{r_{1,j}}, \dots, v_{r_{n_j,j}})$ and \mathbf{Z}_j consists of the first n_j rows of the reordered design matrix, including the intercept column:

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

Thus, using these data augmentation steps, it is possible to represent the normal dynamic survival model defined in (1) and (3) as a conditionally Gaussian state space model (4), opening up the possibility to use sampling procedures from this well understood class of models.

2.4.2 Sampling Scheme

Draws from the posterior distribution are obtained through a MCMC Gibbs sampler with MH steps. The MCMC scheme extends those developed in Wagner (2011), Bitto and Frühwirth-Schnatter (2019), Cadonna et al. (2020) and Knaus et al. (2021), with much of the implementation being imported from the latter. As such, only a brief sketch of the algorithm is given here and the interested reader is referred to the aforementioned papers for further details.

Algorithm 1: MCMC inference for normal dynamic survival models with shrinkage.

- a) Use steps 1 – 4 from Algorithm 1 in Knaus et al. (2021), using representation (4).
 - b) Sample the auxiliary variables τ_{ij} and the component indicators r_{ij} as in Wagner (2011), Section 3.4, step (c) and use these to update \mathbf{x}_j , for $j = 1, \dots, J$.
-

2.5 Adding a factor

To account for unobserved heterogeneity that may be present in the augmented survival times in the form of covariance, we additionally add a grouped factor component to the hazard rates. Let observation i belong to group g , with $g \in \{1, \dots, G\}$. Then, when the factor component is added, the hazard rates look as follows:

$$\lambda_{ij} = \exp \left(\phi_g f_j + \beta_{0j} + \sum_{k=1}^K z_{ik} \beta_{kj} \right),$$

where f_j is allowed to vary over time according to a zero-mean stochastic volatility law of motion:

$$\begin{aligned} f_j &\sim \mathcal{N}(0, e^{h_j}), \\ h_j | h_{j-1}, \phi_f, \sigma_f^2 &\sim \mathcal{N}(\phi_f h_{j-1}, \sigma_f^2), \\ h_0 &\sim \mathcal{N}(0, \sigma_f^2 / (1 - \phi_f^2)). \end{aligned}$$

This specification has similarities with shared frailty models (Hougaard, 2000), with the special case $G = N$ closely related to regular (i.e. individual level) frailty models (Hougaard, 1995). A key difference lies in the stochastic volatility specification, as it allows for time variation in the frailty term, whereas standard frailty models usually assume that the term remains constant over time. Frailty models are random effect models applied in the context of time-to-event analysis, where the aim is to capture unobserved heterogeneity. Such unobserved heterogeneity, where it impacts the outcome variable, can induce spurious time-dependent effects of covariates through a selection effect (Balan and Putter, 2020). This effect also has ramifications for the identifiability of frailty terms in models with time-varying parameters, as, without strict assumptions, it is not possible to determine whether time-variation comes from individual frailty or through a truly time-varying effect. In the context of a time-varying parameter model, frailty is therefore particularly useful in the shared frailty specification, where multiple observations share a single frailty term, as this alleviates the identifiability issues. A common example of an application for such a shared frailty model is a scenario where repeated measurements are taken, as the shared frailty term can account for unobserved individual heterogeneity.

2.5.1 Identification of signs of factor loadings

In this specification, the signs of the factor loadings ϕ_1, \dots, ϕ_G and the factor f_j are not identified. To demonstrate this, let ϕ be the vector of factor loadings for all individuals. Then

$$\phi f_j = (-\phi)(-f_j).$$

To remedy this, we first run the MCMC sampler in the unrestricted space and then enforce identification post-hoc through the methods described in Geweke and Zhou (1996) and Lopes and West (2004). In the more general model with multiple factors, this requires a rotation matrix that is positive lower triangular, while in the one-dimensional special case (as it appears above) it amounts to ensuring that ϕ_1 is always positive.

2.5.2 Prior specification

In specifying the priors for the parameters governing the stochastic volatility process, we follow Kastner (2016):

$$\begin{aligned} (\phi_f + 1) / 2 | a_0, b_0 &\sim \mathcal{B}(a_0, b_0), \\ \sigma_f^2 | B_{\sigma_f} &\sim \mathcal{G}(1/2, 1/2 B_{\sigma_f}). \end{aligned}$$

To ensure sparsity in the factor loadings, we place triple gamma shrinkage priors on the factor loadings, analogous to those used in Section 2.3.1:

$$\phi_g \mid a_\phi, c_\phi, \lambda_{B\phi}^2 \sim TG(a_\phi, c_\phi, \lambda_{B\phi}^2).$$

Naturally, as in Section 2.3.1, this includes all special and limiting cases of the triple gamma prior, allowing for a large degree in flexibility in terms of the amount of shrinkage applied. Similarly, the hyperpriors are also specified in an analogous fashion to those in Section 2.3.1. This approach has similarities to that proposed in Kastner (2019), albeit with only one factor instead of multiple and a more general choice of shrinkage prior.

2.5.3 Additional MCMC steps for the factor component

Algorithm 2: MCMC inference for normal dynamic survival models with shrinkage and a factor.

- a) Use steps 1 – 4 from Algorithm 1 in Knaus et al. (2021), using representation (4).
b) Define the following stacked counterparts for parameters used in the model:

- $\tilde{\mathbf{x}}_j = \mathbf{x}_j - \mathbf{z}_j \boldsymbol{\beta}_j$, $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_J]'$,
- $\mathbf{V} = [\mathbf{V}_1, \dots, \mathbf{V}_J]'$,
- $\mathbf{f} = \underbrace{[(f_1, \dots, f_1)]}_{n_1 \text{ elements}}, \dots, \underbrace{[(f_J, \dots, f_J)]}_{n_J \text{ elements}}'$,

and the index p , for $p = 1, \dots, \sum_{j=1}^J n_j$ with group specific indices p_g , with $p_g = \{p \mid \text{individual } i \text{ in group } g\}$. These effectively select the rows belonging to one group from the stacked counterparts of the parameters. Then sample ϕ_g for $g = 1, \dots, G$ from the following conditional posterior:

$$p(\phi_g | \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}_{-\phi_g}) \propto p(\tilde{\mathbf{x}}_g | \phi_g, \mathbf{V}, \mathbf{f}) p(\phi_g) \\ \propto \exp \left(\frac{1}{2\bar{\sigma}_{\phi_g}^2} \left(\phi_g - \bar{\sigma}_{\phi_g}^2 \sum_{e: e \in p_g} \tilde{\mathbf{x}}_e \mathbf{f}_e / \mathbf{V}_e \right)^2 \right)$$

with $\bar{\sigma}_{\phi_g}^2 = \left(\frac{1}{\sum_{e: e \in p_g} \mathbf{f}_e^2 / \mathbf{V}_e} + \frac{1}{\sigma_{\phi_g}^2} \right)$.

- c) Sample f_j for $j = 1, \dots, J$ by defining the following J univariate regression problems:

$$\tilde{\mathbf{x}}_j = \phi_j f_j + \boldsymbol{\varepsilon}_j,$$

where ϕ_j is a vector of the stacked values of ϕ_g for all individuals i still alive at time j . Posterior samples are generated by drawing from the resulting conditional normal posteriors.

- d) Perform a boosting step to improve the mixing of the factor loadings.

- (a) Define ϕ_m as the largest of ϕ_1, \dots, ϕ_G in absolute value, $\phi_g^* = \frac{\phi_g}{\phi_m}$ and $f_j^* = f_j \phi_m$.
- (b) Sample ϕ_m^2 from the full conditional posterior in Section 6.0.1. Define ϕ_m^{new} as $\phi_m^{\text{new}} = \sqrt{\phi_m^2}$, preserving the sign of the original ϕ_m .
- (c) Determine ϕ_g^{new} through the transformation $\phi_g^{\text{new}} = \phi_g^* \phi_m^{\text{new}}$, for $g = 1, \dots, G$. Furthermore, determine f_j^{new} through the transformation $f_j^{\text{new}} = f_j^* / \phi_m^{\text{new}}$, for $j = 1, \dots, J$.

- d) Sample the variances of the factor loadings τ_g^ϕ for $g = 1, \dots, G$ as in step 4 from Algorithm 1 in Knaus et al. (2021).

- e) Sample f_j for $j = 1, \dots, J$ by defining the following J univariate regression problems:

$$\tilde{\mathbf{x}}_j = \phi_j f_j + \boldsymbol{\varepsilon}_j,$$

where ϕ_j is a vector of the stacked values of ϕ_g for all individuals i still alive at time j . Posterior samples are generated by drawing from the resulting conditional normal posteriors.

- f) Sample the persistence ϕ_f , the volatility of the volatility σ_f^2 and the log-volatilities $h = (h_0, \dots, h_J)$ as in Kastner (2016) using `stochvol` Hosszejni and Kastner (2021).
 - g) Sample the auxiliary variables τ_{ij} and the component indicators r_{ij} as in Wagner (2011), Section 3.4, step (c) and use these to update \mathbf{x}_j , for $j = 1, \dots, J$.
-

3 R Package shrinkDSM

The methods described are implemented in the R package `shrinkDSM` (Winkler and Knaus, 2021), offering an efficient yet easy-to-use tool for estimating piecewise exponential dynamic survival models. It aims to have a low barrier to entry, while still being flexible enough for more advanced users. This is achieved by choosing sensible default values for the hyperparameters, while allowing the user to modify the prior setup in fairly deep ways, should they so desire. The resulting software works well for a wide variety of datasets, requires very little tuning and leads to a default model that can be estimated in only a few lines:

```
set.seed(123)
data("gastric")

# Create intervals for piecewise exponential model
intervals <- divisionpoints(gastric$time, gastric$status, 2)

# Estimate default model
mod <- shrinkDSM(time ~ radiation, gastric,
                 delta = gastric$status, S = intervals)
```

More advanced users can modify the prior setup (e.g. change the class of prior used, modify hyperparameters or even deactivate the learning of certain hyperparameters) in a fashion analogous to the methods presented in Knaus et al. (2021), as the underlying Gaussian state space model is very similar. While this similarity leads to many of the methods carrying over, a distinction lies in the factor model, which is only present in the model described in this paper. To add a factor to the model, a grouping variable is supplied via the argument `groups`. Then, the prior setup for ϕ_1, \dots, ϕ_G can be modified via the argument `phi_param`:

```
# Modify hyperparameters for phi
# Note that using group = 1:nrow(gastric) is
# the special case G = N
mod2 <- shrinkDSM(time ~ radiation, gastric,
                  delta = gastric$status, S = intervals,
                  group = 1:nrow(gastric),
                  phi_param = list(mod_type_phi = "triple",
                                   learn_lambda2_B_phi = FALSE))
```

The arguments provided in the named list for `phi_param` behave in the same fashion as those for the variances of the innovations and the means of the initial states, albeit with the suffix `_phi`.

A further goal is to combine the benefits of the low-level (and therefore fast) programming language C++ with those of the higher level language R, with its interpreted code and rich suite of data wrangling software. All computationally intensive code is implemented in C++ and then interfaced with R via the `Rcpp` (Eddelbuettel and François, 2011) package. Code dealing with plotting and interpreting results, as well as checking MCMC convergence, is then written in R, allowing for easy interfacing with other R packages (e.g. `coda`; Plummer et al. 2006). Furthermore, forecasting of survival times based on new covariates and the use of time-varying covariates is supported.

4 Application

Esophageal Cancer (EC) is the eighth most common cancer worldwide, with less than 20% of patients surviving more than five years. Whereas in western countries the number of esophageal squamous cell carcinoma (ESCC) is declining, the number of adenocarcinomas of the esophagogastric junction (AEG) diagnosed is increasing (Bray et al., 2018). Despite the development of multimodal approaches, combining surgical resection with perioperative chemo-(radio)therapy, 5-year survival rate for patients diagnosed with AEG remains poor (Behrens et al., 2015).

Inflammation, as one of the hallmarks of cancer, is an acknowledged factor in tumor biology. Inflammation-driven tumorigenesis and tumor progression plays a crucial role in malignant diseases. Systemic inflammatory response (SIR) in the context of tumor-associated inflammation has been demonstrated to diminish outcome and be of major prognostic importance in various cancers. Investigation of tumor-driving inflammation-based components is of major importance and targeting pathways of inflammatory response might become a cornerstone of cancer treatment (Hanahan and Weinberg, 2011). Therefore, identification of easy-available markers might help to determine individual treatment approaches. The utility of inflammation-based scores, such as the systemic immune-inflammation index (SII), are based on routinely obtained markers that are available before surgery. Elevated SII has been reported to be associated with clinicopathological parameters and has been proven to be an independent prognostic factor in a number of malignancies (Aziz et al., 2019; Hong et al., 2015; Hu et al., 2014).

In the present application, we investigate the prognostic value of SII in patients with adenocarcinoma of the gastroesophageal junction who underwent primary resection or were treated with neoadjuvant therapy before surgery. The data set consists of 287 consecutive patients that underwent curative resection of locally advanced adenocarcinomas of the gastroesophageal junction between January 1992 and April 2016 at the Department of Surgery at the Medical University Vienna were identified from a prospectively maintained database. Serum concentrations of platelets, neutrophils, and lymphocytes were measured within 3 days before the start of neoadjuvant treatment or surgery in patients treated with primary surgery. The SII is calculated as follows: $SII = \frac{\text{platelet} * \text{neutrophil}}{\text{lymphocyte}}$. In addition to the SII several common control variables described in Table 1 are used for estimation. The sampler was run for 100.000 iterations with a burn-in of 20.000 iterations and a thinning factor of 20. The default prior-hyperparameters in Winkler and Knaus (2021) were used. The partitioning set S contains every observed death.

Table 1: Description of Variables

Variable	Description
SII	Systemic Immune Inflammation Index
age	age at the beginning of treatment
sex	female = 0, male = 1
cT	Clinical tumor stage
cN	Clinical lymph node stage
ASA	American Society of Anesthesiologists physical status classification
G	Tumor differentiation

The plots in the first row of Figure 3 show the posterior parameter associated with SII for patients who were treated with primary surgery and neoadjuvant treatment, respectively. The shaded blue area represent the pointwise 50% and 95% credible intervals, respectively. For both groups the risk associated with SII increases over the first

10 months after starting treatment (either surgery for primary resected patients of adjuvant treatment for those who received it) and decreases after the initial peak. However, while the parameter becomes statistically indistinguishable from 0 for patients who underwent primary resection after ~ 6 years, it stays constant, with a median around 1 for those who received neoadjuvant therapy. The Cox proportional hazards model (Table 2), unable to capture time variation, simply results in a smaller estimate for primary resected patients. Thus the proposed model results in two important interpretative differences. First, the parameters associated with SII for the two groups are practically indistinguishable for the first two years. Second, while SII remains an important factor for survival in patients receiving neoadjuvant chemotherapy, it is only important for the first 6 years or so for those undergoing primary resection.

To showcase that constant parameters can also be captured with the proposed model, the posterior parameters associated with the clinical lymph node stage (cN) are also included in Figure 3. All insignificant parameters, according to the Cox PH model, are also shrunk to 0 and constant using the Bayesian approach. Furthermore, the parameter associated with cN = 0 is significantly below 0 in the Cox PH model ($z\text{-score} = -4.761$). In the proposed model the vast majority of posterior mass is constant around -1 .

This example highlights the importance of the time-varying parameter approach when it comes to understanding the effect of SII on survival time. The proposed model is shown to be flexible enough to capture both constant and time-varying parameters. In addition, the shrinkage prior used gives a clear indication of which parameters should be included in the model without the need for hyperparameter tuning.

5 Concluding remarks

In this paper we extend existing Bayesian dynamic survival models (Wagner, 2011; Gamerman, 1991) in three ways. First, in order to reduce overfitting, even in high dimensional models, we employ state of the art global-local shrinkage priors similar to Bitto and Frühwirth-Schnatter (2019) and Cadonna et al. (2020). Second, to account for unobserved heterogeneity a time-varying grouped factor model is introduced. The factor can be used, for example, in the case of repeated observations similar to a frailty parameter. Third, a Gibbs-sampler for the proposed model is developed and implemented in a convenient R-package which allows users to specify time varying inputs to the model and sample from the posterior predictive distribution of survival times for new observations. The flexibility of the proposed model in conjunction with the effective shrinkage prior allows practitioners to employ it in a large variety of applications. These include, for example, the survival times of machinery where usage and maintenance data can change over time, customer lifetime value analysis with repeated observations, or evaluating prognostic factors in a medical context.

We showcase the effectiveness of the prior specification and the advantages of the dynamic survival model by investigating the prognostic value of SII in patients with adenocarcinoma of the gastroesophageal junction and comparing it to the current standard in the medical literature, the Cox proportional hazard model. We show that the global-local shrinkage prior is effective in selecting both parameters and time-variation. In addition, the plots of the posterior parameter distributions, readily available in the R-package, allow for convenient interpretation of the complex model. The framework is especially appealing in a medical context where often a large number of potential prognostic factors and relatively few observations are available.

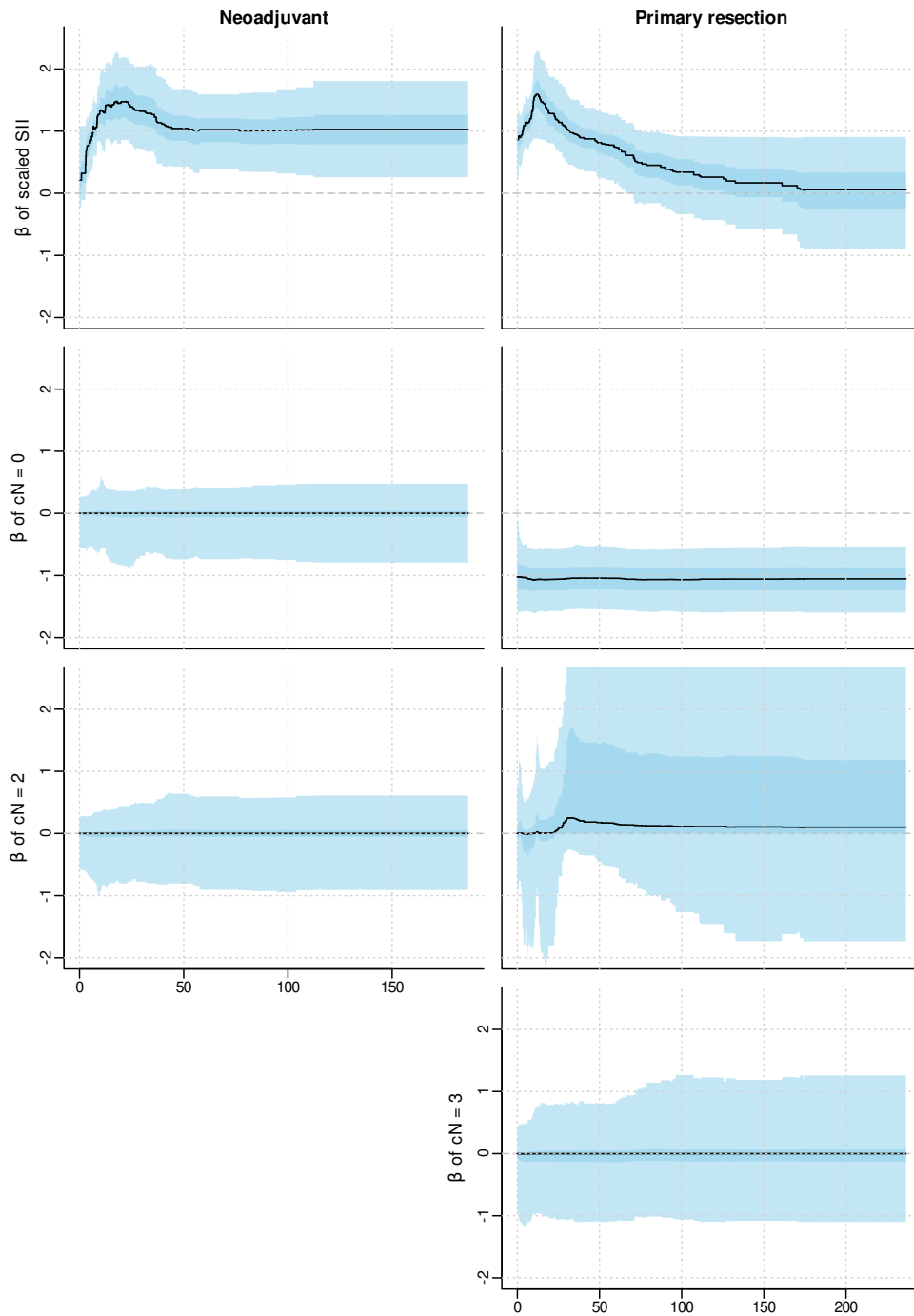


Figure 3: Posterior parameters associated with systemic immune-inflammation index (SII) and clinical lymphnode stage (cN) for primary resected and neoadjuvant treated patients. Black lines represent posterior medians, while shaded areas represent 95% and 50% credible intervals.

Table 2: Parameters estimated using the Cox Proportional Hazards model

	<i>Dependent variable:</i>	
	survival time	
	(Neoadjuvant)	(Primary resection)
scaled SII	1.250*** (0.137)	1.024*** (0.104)
cT2	0.203 (0.248)	
cT3		0.347 (0.236)
cT1		-0.248 (0.271)
cN2	0.032 (0.335)	0.387 (0.369)
cN3		-0.252 (0.730)
cN0	-0.009 (0.322)	-1.237*** (0.260)
age	-0.017 (0.011)	0.012 (0.010)
sex0	-0.225 (0.345)	0.401* (0.238)
G0,1,2	-0.141 (0.231)	
G2		-0.600*** (0.226)
G1		-1.167* (0.652)
ASA1	0.104 (0.254)	0.137 (0.315)
ASA3&4	0.533* (0.316)	-0.339 (0.324)
Observations	150	162
R ²	0.481	0.593
Max. Possible R ²	0.996	0.998
Log Likelihood	-361.742	-450.474
Wald Test	98.000*** (df = 9)	131.470*** (df = 12)
LR Test	98.460*** (df = 9)	145.467*** (df = 12)
Score (Logrank) Test	109.909*** (df = 9)	164.954*** (df = 12)

Note: *p<0.1; **p<0.05; ***p<0.01 SE in parentheses

References

- Aziz, M. H., Sideras, K., Aziz, N. A., Mauff, K., Haen, R., Roos, D., Saida, L., Suker, M., van der Harst, E., Mieog, J. S., et al. (2019). The systemic-immune-inflammation index independently predicts survival and recurrence in resectable pancreatic cancer and its prognostic value depends on bilirubin levels: a retrospective multicenter cohort study. *Annals of surgery*, 270(1):139–146.
- Balan, T. A. and Putter, H. (2020). A tutorial on frailty models. *Statistical methods in medical research*, 29(11):3424–3454.
- Behrens, A., Ell, C., and Lordick, F. (2015). Perioperative and palliative chemotherapy for esophageal cancer. *Visceral Medicine*, 31(5):341–346.
- Belmonte, M. A., Koop, G., and Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94.
- Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424.
- Brown, P. J. and Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian analysis*, 5(1):171–188.
- Cadonna, A., Frühwirth-Schnatter, S., and Knaus, P. (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics*, 8(2):20.
- Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. 81:541–553.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80. PMLR.
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. 15:183–202.
- Frühwirth-Schnatter, S. and Frühwirth, R. (2007a). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, 51(7):3509–3528.
- Frühwirth-Schnatter, S. and Frühwirth, R. (2007b). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, 51(7):3509–3528.
- Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

- Gamerman, D. (1991). Dynamic bayesian models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 40(1):63–79.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statistica sinica*, pages 339–373.
- Geweke, J. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *The review of financial studies*, 9(2):557–587.
- Griffin, J., Brown, P., et al. (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis*, 12(1):135–159.
- Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell*, 144(5):646–674.
- Hemming, K. and Shaw, E. (2002). A parametric dynamic survival model applied to breast cancer survival times. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(4):421–435.
- Hemming, K. and Shaw, E. (2005). A class of parametric dynamic survival models. *Lifetime data analysis*, 11(1):81–98.
- Hong, X., Cui, B., Wang, M., Yang, Z., Wang, L., and Xu, Q. (2015). Systemic immune-inflammation index, based on platelet counts and neutrophil-lymphocyte ratio, is useful for predicting prognosis in small cell lung cancer. *The Tohoku journal of experimental medicine*, 236(4):297–304.
- Hosszejni, D. and Kastner, G. (2021). Modeling univariate and multivariate stochastic volatility in r with stochvol and factorstochvol. *Journal of Statistical Software*, 100:1–34.
- Hotz-Behofsits, C., Huber, F., and Zörner, T. O. (2018). Predicting crypto-currencies using sparse non-gaussian state space models. *Journal of Forecasting*, 37(6):627–640.
- Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, 1(3):255–273.
- Hougaard, P. (2000). Shared frailty models. In *Analysis of multivariate survival data*, pages 215–262. Springer.
- Hu, B., Yang, X.-R., Xu, Y., Sun, Y.-F., Sun, C., Guo, W., Zhang, X., Wang, W.-M., Qiu, S.-J., Zhou, J., et al. (2014). Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. *Clinical Cancer Research*, 20(23):6212–6222.
- Huber, F., Koop, G., and Onorante, L. (2020). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, pages 1–15.
- Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, 69(5):1–30.

- Kastner, G. (2019). Sparse bayesian time-varying covariance estimation in many dimensions. *Journal of Econometrics*, 210(1):98–115.
- Knaus, P., Bitto-Nemling, A., Cadonna, A., and Frühwirth-Schnatter, S. (2021). Shrinkage in the time-varying parameter model framework using the R package shrinkTVP. *Journal of Statistical Software*. conditionally accepted (available as arXiv report 1907.07065).
- Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, pages 41–67.
- McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics & Data Analysis*, 55(1):199–212.
- Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.
- Wagner, H. (2011). Bayesian estimation and stochastic model specification search for dynamic survival models. *Statistics and Computing*, 21(2):231–246.
- Winkler, D. and Knaus, P. (2021). *shrinkDSM: Efficient Bayesian Inference for Dynamic Survival Models with Shrinkage*. R package version 0.1.0.

6 Appendix

6.0.1 Full conditional posterior of ϕ_m

Let ϕ_m be the largest of ϕ_1, \dots, ϕ_G in absolute value and define the following quantities:

$$\phi_g^* = \frac{\phi_g}{\phi_m}, \quad f_j^* = f_j \phi_m.$$

This implies the following working priors:

$$\phi_m^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2\sigma_{\phi_m}^2}\right), \quad f_j^* \sim \mathcal{N}(0, e^{h_j} \phi_m^2), \quad \phi_g^* \sim \mathcal{N}\left(0, \frac{\sigma_{\phi_g}^2}{\phi_m^2}\right),$$

which, in turn, leads to the following full conditional posterior for ϕ_m^2 :

$$\begin{aligned} p(\phi_m^2 | \phi_g^*, h_1, \dots, h_J, f_1^*, \dots, f_J^*) &\propto \prod_{g:g \neq m} p(\phi_g^* | \phi_m^2) \prod_{j=1}^J p(f_j^* | \phi_m^2) p(\phi_m^2) \\ &\propto \left[\prod_{g:g \neq m} (\phi_m^2)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{\phi_g^{*2}}{\sigma_{\phi_g}^2} \phi_m^2\right) \right] \left[\prod_{j=1}^J (\phi_m^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{f_j^{*2}}{e^{h_j}} \frac{1}{\phi_m^2}\right) \right] \\ &\times (\phi_m^2)^{\frac{1}{2}-1} \exp\left(-\phi_m^2 \frac{1}{2\sigma_{\phi_m}^2}\right) \\ &= (\phi_m^2)^{\frac{G-J}{2}-1} \exp\left[-\frac{1}{2} \left(\phi_m^2 \left(\frac{1}{\sigma_{\phi_m}^2} + \sum_{g:g \neq m} \frac{\phi_g^{*2}}{\sigma_{\phi_g}^2} \right) + \frac{1}{\phi_m^2} \sum_{j=1}^J \frac{f_j^{*2}}{e^{h_j}} \right)\right], \end{aligned}$$

which is the kernel of a *GIG* $\left(\frac{G-J}{2}, \frac{1}{\sigma_m^2} + \sum_{g:g \neq m} \frac{\phi_g^{*2}}{\sigma_g^2}, \sum_{j=1}^J \frac{f_j^{*2}}{e^{h_j}}\right)$ distribution.