Supplemental Material to the Main Manuscript *"BAMLSS: Bayesian Additive Models for Location, Scale and Shape (and Beyond)"*

Nikolaus Umlauf
Department of Statistics, Universität Innsbruck
and
Nadja Klein
Chairs of Statistics and Econometrics, Universität Göttingen
and
Achim Zeileis
Department of Statistics, Universität Innsbruck

Abstract

This document provides more detailed information about algorithmic derivations and "Lego bricks" used in the main manuscript. Section 1 provides an overview of commonly used priors for GAM-type model terms. Section 2 presents the "Lego bricks" for estimating censored Gaussian models and Section 3 for the Cox model. The detailed derivations for posterior mode estimation and MCMC are then presented in Section 4 and 5

1 “Lego brick” B3, prior structures in BAMLSS

Table 1 provides an overview of commonly used “Lego bricks” for GAM-type prior structures used within building block B3 in Section 4.

<table>
<tr><th rowspan="2">Covariates</th><th rowspan="2">Effect types $f_{jk}(\mathbf{x}; \boldsymbol{\beta}_{jk})$</th><th colspan="2">Prior densities $p_{jk}(\boldsymbol{\beta}_{jk}; \boldsymbol{\tau}_{jk}, \boldsymbol{\alpha}_{jk})$</th></tr>
<tr><th>$d_{\boldsymbol{\beta}_{jk}}(\boldsymbol{\beta}_{jk} \vert\, \boldsymbol{\tau}_{jk}; \boldsymbol{\alpha}_{\boldsymbol{\beta}_{jk}})$</th><th>$d_{\tau_{ljk}}(\tau_{ljk} \vert\, \boldsymbol{\alpha}_{\tau_{ljk}})$</th></tr>
<tr><td rowspan="6">Scalar covariates</td><td>Intercept β</td><td rowspan="3">$\propto$ constant
$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m})^\top \mathbf{P}(\boldsymbol{\tau})(\boldsymbol{\beta} - \mathbf{m})\right)$</td><td rowspan="3">$\emptyset$</td></tr>
<tr><td>Linear effect $x \cdot \beta$</td></tr>
<tr><td>Linear interaction $x_1 \cdot x_2 \cdot \beta$</td></tr>
<tr><td>Smooth effect $f(x)$</td><td rowspan="8">$\propto \vert\mathbf{P}(\boldsymbol{\tau})\vert^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^\top \mathbf{P}(\boldsymbol{\tau})\boldsymbol{\beta}\right)$</td><td rowspan="2">IG $\propto \tau^{-(a+1)} \exp(-b/\tau)$</td></tr>
<tr><td>Varying coefficient $f(x_2) \cdot x_1$</td></tr>
<tr><td>Smooth effect $f(x_1, \ldots, x_L)$</td><td rowspan="2">HC $\propto (1 + \tau/a^2)^{-1}(\tau/a^2)^{-1/2}$</td></tr>
<tr><td rowspan="2">Grouping variable s</td><td>Random intercept β_s</td></tr>
<tr><td>Spatial effect $f(s)$</td><td rowspan="2">SD $\propto (\tau/\sqrt{\tau})^{-1/2} \exp(-(\tau/a)^{1/2})$</td></tr>
<tr><td rowspan="3">Grouping and scalar, time variable t</td><td>Random slope $x \cdot \beta_s$</td></tr>
<tr><td>Space-time effect $f(s, t)$</td><td rowspan="2">HN $\propto \tau^{1/2-1} \exp(-\tau/(2a^2))$</td></tr>
<tr><td>Functional random intercept $f_s(t)$</td></tr>
</table>

Table 1: Commonly used “Lego bricks”, building block B3, for model terms in BAMLSS. Priors for linear effects assume that the precision matrix $\mathbf{P}(\boldsymbol{\tau})$ is fixed. For smooth effects, prior densities are: inverse gamma (IG), half-Cauchy (HC), scale-dependent (SD) and half-normal (HN).

2 “Lego bricks” for the censored Gaussian distribution

The following presents the “Lego bricks” of example Section 6.1 in the main manuscript that are used for the construction of iteratively weighted least squares (IWLS) updating functions $U_{jk}(\cdot)$ of the Gaussian model left censored at zero.

B1. The density function of a Gaussian distribution left censored at zero is given by

$$f(y;\, \mu = \eta_\mu,\, \log(\sigma) = \eta_\sigma) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) & y > 0 \\ \Phi\left(\frac{-\mu}{\sigma}\right) & \text{else,} \end{cases} \tag{1}$$

where ϕ is the probability density and Φ the cumulative distribution function of the standard normal distribution.

B6b. Score vectors $\mathbf{u}_k = \partial \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X})/\partial \boldsymbol{\eta}_k$ are computed with

$$\frac{\partial \ell(\boldsymbol{\beta}; y, \mathbf{x})}{\partial \eta_\mu} = \begin{cases} \frac{y-\mu}{\sigma^2} & y > 0 \\ -\frac{1}{\sigma}\frac{\phi\left(\frac{-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)} & \text{else,} \end{cases}$$

and

$$\frac{\partial \ell(\boldsymbol{\beta}; y, \mathbf{x})}{\partial \eta_\sigma} = \begin{cases} -1 + \frac{(y-\mu)^2}{\sigma^2} & y > 0 \\ -\frac{-\mu}{\sigma}\frac{\phi\left(\frac{-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)} & \text{else.} \end{cases}$$

B7b. The diagonal elements of the weight matrix $\mathbf{W}_{kk} = -\text{diag}(\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X})/\partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_k^\top)$ are derived using

$$\frac{\partial^2 \ell(\boldsymbol{\beta}; y, \mathbf{x})}{\partial \eta_\mu^2} = \begin{cases} -\frac{1}{\sigma^2} & y > 0 \\ -\frac{-\mu}{\sigma^3}\frac{\phi\left(\frac{-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)} - \frac{1}{\sigma^2}\frac{\phi\left(\frac{-\mu}{\sigma}\right)^2}{\Phi\left(\frac{-\mu}{\sigma}\right)^2} & \text{else,} \end{cases}$$

and

$$\frac{\partial^2 \ell(\boldsymbol{\beta}; y, \mathbf{x})}{\partial \eta_\sigma^2} = \begin{cases} -2\frac{(y-\mu)^2}{\sigma^2} & y > 0 \\ -\frac{-\mu}{\sigma}\frac{\phi\left(\frac{-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)} - \frac{(-\mu)^3}{\sigma^3}\frac{\phi\left(\frac{-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)} - \frac{(-\mu)^2}{\sigma^2}\frac{\phi\left(\frac{-\mu}{\sigma}\right)^2}{\Phi\left(\frac{-\mu}{\sigma}\right)^2} & \text{else.} \end{cases}$$

3 "Lego bricks" for the Cox model

Based on the survival function

$$S(t) = \text{Prob}(T > t) = \exp\left(-\int_0^t \lambda(u)du\right),$$

for full Bayesian inference the following "Lego bricks" need to be implemented for updating functions $U_{jk}(\cdot)$ using algorithms A1, A2a and A2b (algorithms are presented in Section 3.2 in the main manuscript):

B1. The log-likelihood function of the continuous time Cox model is given by

$$\ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \sum_{i=1}^{n}\left(\delta_i \eta_{i\gamma} - \int_0^{t_i} \exp(\eta_{i\lambda}(u)du)\right).$$

where δ_i is the usual censoring indicator, which equals to $\delta_i = 1$ in this example, because we focus on real fire events.

B6a. For derivative-based estimation using Algorithm A2a and for MCMC simulation with Algorithm A2b, the score vectors and Hessian need to be computed. Assuming a basis function approach, the score vector of the regression coefficients for the time-varying part $\eta_\lambda(t)$ is

$$\mathbf{s}\left(\boldsymbol{\beta}_\lambda\right) = \boldsymbol{\delta}^\top \mathbf{X}_\lambda(\mathbf{t}) - \sum_{i=1}^{n} \exp(\eta_{i\gamma}) \left(\int_0^{t_i} \exp(\eta_{i\lambda}(u)) \mathbf{x}_i(u) du \right).$$

B7a. The elements of the Hessian w.r.t. $\boldsymbol{\beta}_\lambda$ are

$$\mathbf{H}\left(\boldsymbol{\beta}_\lambda\right) = -\sum_{i=1}^{n} \exp\left(\eta_{i\gamma}\right) \int_0^{t_i} \exp(\eta_{i\lambda}(u)) \boldsymbol{x}_{i\lambda}(u) \boldsymbol{x}_{i\lambda}^\top(u) du.$$

Note that the Hessian cannot be fragmented further to obtain building block B7b and IWLS updating functions. The reason is that the diagonal weight matrix based on $\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) / \partial \boldsymbol{\eta}_\lambda(\mathbf{t}) \partial \boldsymbol{\eta}_\lambda(\mathbf{t})^\top$ requires a functional derivative like the Hadamard derivative since the predictor depends on time t. However, it turns out that this derivative forms martingale residuals in the IWLS step (see, e.g., Barlow, 1988) which are incapable of estimating time-varying effects, see also Hofner (2008, Section 5.2) for a detailed discussion. Therefore updating functions $U_{jk}(\cdot)$ for the time-varying predictor $\eta_\lambda(t)$ are based on updating Equation (17) within Algorithm A2a and A2b.

B6b & B7b. Constructing updating functions for the time-constant part η_γ again yields an IWLS updating scheme, see Section 4, with working observations given by

$$\mathbf{z} = \boldsymbol{\eta}_\gamma + \mathbf{W}^{-1}\mathbf{u},$$

with the weight matrix

$$\mathbf{W} = \mathrm{diag}(\mathbf{P} \exp(\boldsymbol{\eta}_\gamma)),$$

where $\mathbf{P}$ is a diagonal matrix with elements $p_{ii} = \int_0^{t_i} \exp(\eta_{i\lambda}(u) du)$. The score vector is

$$\mathbf{u} = \boldsymbol{\delta} - \mathbf{P} \exp(\boldsymbol{\eta}_\gamma).$$

(Hennerfeind et al., 2006)

4 Posterior mode updating based on IWLS

The following shows the steps needed to derive the iterative updating scheme based on IWLS in Section 4.2. Focusing on the j-th row of (14) gives

$$
\begin{aligned}
&(\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))\boldsymbol{\beta}_{jk}^{(t+1)} + \ldots + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t+1)} - \\
&\quad (\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))\boldsymbol{\beta}_{jk}^{(t)} - \ldots - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} - \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})(\boldsymbol{\beta}_{jk}^{(t+1)} - \boldsymbol{\beta}_{jk}^{(t)}) + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t+1)} + \ldots + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t+1)} - \\
&\quad \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t)} - \ldots - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} - \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t)} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})(\boldsymbol{\beta}_{jk}^{(t+1)} - \boldsymbol{\beta}_{jk}^{(t)}) + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t+1)} + \ldots + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t+1)} - \\
&\quad \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t)} - \ldots - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t+1)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t+1)} + \ldots + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t+1)} - \\
&\quad \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t)} - \ldots - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{J_k k}\boldsymbol{\beta}_{J_k k}^{(t)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)}
\end{aligned}
$$

$$
\mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t+1)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk}\boldsymbol{\beta}_{jk}^{(t+1)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_{k,-j}^{(t+1)} - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_k^{(t)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)}
$$

$$
(\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))\boldsymbol{\beta}_{jk}^{(t+1)} = \mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_k^{(t)} - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_{k,-j}^{(t+1)}
$$

$$
\boldsymbol{\beta}_{jk}^{(t+1)} = (\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))^{-1}(\mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_k^{(t)} - \mathbf{X}_{jk}^\top \mathbf{W}_{kk}\boldsymbol{\eta}_{k,-j}^{(t+1)})
$$

$$
\boldsymbol{\beta}_{jk}^{(t+1)} = (\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))^{-1}\mathbf{X}_{jk}^\top(\mathbf{u}_k^{(t)} + \mathbf{W}_{kk}\boldsymbol{\eta}_k^{(t)} - \mathbf{W}_{kk}\boldsymbol{\eta}_{k,-j}^{(t+1)})
$$

$$
\boldsymbol{\beta}_{jk}^{(t+1)} = (\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))^{-1}\mathbf{X}_{jk}^\top(\mathbf{W}_{kk}\mathbf{W}_{kk}^{-1}\mathbf{u}_k^{(t)} + \mathbf{W}_{kk}\boldsymbol{\eta}_k^{(t)} - \mathbf{W}_{kk}\boldsymbol{\eta}_{k,-j}^{(t+1)})
$$

$$
\boldsymbol{\beta}_{jk}^{(t+1)} = (\mathbf{X}_{jk}^\top \mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))^{-1}\mathbf{X}_{jk}^\top \mathbf{W}_{kk}(\mathbf{W}_{kk}^{-1}\mathbf{u}_k^{(t)} + \boldsymbol{\eta}_k^{(t)} - \boldsymbol{\eta}_{k,-j}^{(t+1)})
$$

This yields the updating function $U_{jk}(\cdot)$ shown in (16) in the main manuscript.

5 Approximate full conditionals for MCMC

The following shows the steps to derive a multivariate normal jumping distribution based on a second order Taylor series expansion of the log-posterior centered at the last state of $\boldsymbol{\beta}_{jk}$.

$$\begin{aligned}
\pi(\boldsymbol{\beta}^{\star}_{jk}|\cdot) &\propto \exp\left[\log \pi\left(\boldsymbol{\beta}^{(t)}_{jk}|\cdot\right) + \left(\boldsymbol{\beta}^{\star}_{jk} - \boldsymbol{\beta}^{(t)}_{jk}\right)^{\top} \mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) + \right. \\
&\qquad \left. \frac{1}{2}\left(\boldsymbol{\beta}^{\star}_{jk} - \boldsymbol{\beta}^{(t)}_{jk}\right)^{\top} \mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\left(\boldsymbol{\beta}^{\star}_{jk} - \boldsymbol{\beta}^{(t)}_{jk}\right)\right] \\
&\propto \exp\left[(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) + \left(\frac{1}{2}(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) - \right.\right. \\
&\qquad \left.\left. \frac{1}{2}(\boldsymbol{\beta}^{(t)}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\right)\left(\boldsymbol{\beta}^{\star}_{jk} - \boldsymbol{\beta}^{(t)}_{jk}\right)\right] \\
&\propto \exp\left[(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) + \frac{1}{2}(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{\star}_{jk} - \right. \\
&\qquad \left. \frac{1}{2}(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{(t)}_{jk} - \frac{1}{2}(\boldsymbol{\beta}^{(t)}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{\star}_{jk}\right] \\
&= \exp\left[\frac{1}{2}(\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{\star}_{jk} + (\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) - (\boldsymbol{\beta}^{\star}_{jk})^{\top}\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{(t)}_{jk}\right] \\
&= \exp\left[-\frac{1}{2}(\boldsymbol{\beta}^{\star}_{jk})^{\top} - \mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{\star}_{jk} + (\boldsymbol{\beta}^{\star}_{jk})^{\top}\left(\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) - \mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{(t)}_{jk}\right)\right]
\end{aligned}$$

Which leads to the proposal density $q(\boldsymbol{\beta}^{\star}_{jk}|\,\boldsymbol{\beta}^{(t)}_{jk}) = \mathcal{N}(\boldsymbol{\mu}^{(t)}_{jk}, \boldsymbol{\Sigma}^{(t)}_{jk})$ with precision matrix

$$\left(\boldsymbol{\Sigma}^{(t)}_{jk}\right)^{-1} = -\mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)$$

and mean

$$\begin{aligned}
\boldsymbol{\mu}^{(t)}_{jk} &= \boldsymbol{\Sigma}^{(t)}_{jk}\left[s\left(\boldsymbol{\beta}^{(t)}_{jk}\right) - \mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)\boldsymbol{\beta}^{(t)}_{jk}\right] \\
&= \boldsymbol{\beta}^{(t)}_{jk} - \mathbf{H}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right)^{-1}\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) \\
&= \boldsymbol{\beta}^{(t)}_{jk} - \left[\mathbf{J}_{kk}\left(\boldsymbol{\beta}^{(t)}_{jk}\right) + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))\right]^{-1}\mathbf{s}\left(\boldsymbol{\beta}^{(t)}_{jk}\right).
\end{aligned}$$

Using a basis function representation of functions $f_{jk}(\cdot)$ the precision matrix is

$$\left(\boldsymbol{\Sigma}^{(t)}_{jk}\right)^{-1} = \mathbf{X}^{\top}_{jk}\mathbf{W}_{kk}\mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}),$$

with weights $\mathbf{W}_{kk} = -\mathrm{diag}(\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) / \partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}_k^\top)$ and the mean can be written as

$$\begin{aligned}
\boldsymbol{\mu}_{jk}^{(t)} &= \boldsymbol{\Sigma}_{jk}^{(t)} \left[\mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} - \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk})\boldsymbol{\beta}_{jk}^{(t)} + (\mathbf{X}_{jk}^\top \mathbf{W}_{kk} \mathbf{X}_{jk} + \mathbf{G}_{jk}(\boldsymbol{\tau}_{jk}))\boldsymbol{\beta}_{jk}^{(t)}\right] \\
&= \boldsymbol{\Sigma}_{jk}^{(t)} \left[\mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk} \mathbf{X}_{jk} \boldsymbol{\beta}_{jk}^{(t)}\right] \\
&= \boldsymbol{\Sigma}_{jk}^{(t)} \left[\mathbf{X}_{jk}^\top \mathbf{u}_k^{(t)} + \mathbf{X}_{jk}^\top \mathbf{W}_{kk} \left(\boldsymbol{\eta}_k^{(t)} - \boldsymbol{\eta}_{k,-j}^{(t)}\right)\right] \\
&= \boldsymbol{\Sigma}_{jk}^{(t)} \mathbf{X}_{jk}^\top \left[\mathbf{u}_k^{(t)} + \mathbf{W}_{kk} \left(\boldsymbol{\eta}_k^{(t)} - \boldsymbol{\eta}_{k,-j}^{(t)}\right)\right] \\
&= \boldsymbol{\Sigma}_{jk}^{(t)} \mathbf{X}_{jk}^\top \mathbf{W}_{kk} \left[\boldsymbol{\eta}_k^{(t)} + \mathbf{W}_{kk}^{-1} \mathbf{u}_k^{(t)} - \boldsymbol{\eta}_{k,-j}^{(t)}\right] \\
&= \boldsymbol{\Sigma}_{jk}^{(t)} \mathbf{X}_{jk}^\top \mathbf{W}_{kk} \left[\mathbf{z}_k - \boldsymbol{\eta}_{k,-j}^{(t)}\right]
\end{aligned}$$

with working observations $\mathbf{z}_k = \boldsymbol{\eta}_k^{(t)} + \mathbf{W}_{kk}^{-1} \mathbf{u}_k^{(t)}$.

References

Barlow, R. L. P. W. E. (1988). Residuals for relative risk regression. *Biometrika 75*(1), 65–74.

Hennerfeind, A., A. Brezger, and L. Fahrmeir (2006). Geoadditive survival models. *Journal of the American Statistical Association 101*(475), 1065–1075.

Hofner, B. (2008). *Variable Selection and Model Choice in Survival Models with Time-Varying Effects*. Ph. D. thesis, Institut für Statistik.