Supplementary Material: "Generalized Link-Based Additive Survival Models with Informative Censoring"

Supplementary Material A: Software

The models proposed in this article can be employed via the `gamlss()` function in the `R` package `GJRM` (Marra & Radice, 2019). As an example, consider the following call

```
fl <- list(u ~ s(u, bs = "mpi") + z1 + s(z2), u ~ s(u, bs = "mpi") + z1 + s(z2))

M1 <- gamlss(fl, data = data, surv = TRUE, margin = "PH", margin2 = "PH"

   cens = delta, informative = "yes", inform.cov = c("z1"))
```

where `fl` is a list containing the two additive predictors of the informative model, and `s(u, bs = "mpi")` represents the monotonic P-spline function which models a transformation of the baseline survival function. As for `s(z2)`, the default is `bs = "tp"` (penalized low rank thin plate spline) with `k = 10` (number of basis functions) and `m = 2` (order of derivatives). However, argument `bs` can also be set to, for example, `cr` (penalized cubic regression spline), `ps` (P-spline) and `mrf` (Markov random field), to name but a few. In the `gamlss` function, `surv = TRUE` indicates that a survival model is fitted. The arguments `margin ="PH"` and `margin2 ="PH"` specify the link functions for the survival and censoring times, respectively. Table 1 shows the possible choices for the links that have been implemented for this article. In this example, we specify the proportional hazard link (`"PH"`) for the two equations. Argument `cens = delta` is a binary censoring indicator; this variable has to be equal to 1 if the event occurred and 0 otherwise. Finally, `informative = "yes"` indicates that we are fitting a survival model with informative censoring, and `inform.cov = c("z1")` specifies the set of informative covariates.

Model	Link $g(S)$	Inverse link $g^{-1}(\xi) = G(\xi)$	$G'(\xi)$
Prop.hazards("PH")	$\log\{-\log(S)\}$	$\exp\{-\exp(\xi)\}$	$-G(\xi)\exp(\xi)$
Prop.odds("PO")	$-\log\left(\frac{S}{1-S}\right)$	$\frac{\exp(-\xi)}{1+\exp(-\xi)}$	$-G^2(\xi)\exp(-\xi)$
Probit("probit")	$-\Phi^{-1}(S)$	$\Phi(-\xi)$	$-\phi(-\xi)$

Table 1: Link functions implemented in GJRM. Φ and ϕ are the cumulative distribution and density functions of a univariate standard normal distribution. Alternative links can be implemented. The first two functions can be called log-log and -logit links, respectively.

Supplementary Material B: Scores and Hessians

In this section, the detailed derivations of the informative and non-informative Scores and Hessians are presented.

B.1. Informative and Non-informative Scores

If censoring is informative then $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ would have some components in common. Because the first Q components of $\boldsymbol{\gamma}_1$ are the same as the first Q components of $\boldsymbol{\gamma}_2$, we have

$$\boldsymbol{\mathcal{Q}}_{\nu i}^{\top}\boldsymbol{\gamma}_{\nu} = \boldsymbol{\mathcal{Q}}_{i}^{0\top}\boldsymbol{\alpha}_0 + \boldsymbol{\mathcal{Q}}_{\nu i}^{1\top}\boldsymbol{\alpha}_{\nu}.$$

Therefore, defining $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^{\top}, \boldsymbol{\alpha}_1^{\top}, \boldsymbol{\alpha}_2^{\top})^{\top}$, the informative penalized log-likelihood function can be written as

$$\ell_p(\boldsymbol{\alpha}) = \ell(\boldsymbol{\alpha}) - \frac{1}{2}\boldsymbol{\alpha}^{\top}\boldsymbol{\mathcal{S}}\boldsymbol{\alpha}, \tag{1}$$

where $\ell(\boldsymbol{\alpha})$ is defined as

$$\begin{aligned}\ell(\boldsymbol{\alpha}) = &\sum_{i=1}^{n}\left\{\log\mathcal{G}_1\left[\xi_{1i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)\right] + \delta_{1i}\log\left\{-\frac{\mathcal{G}_1'\left[\xi_{1i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)\right]}{\mathcal{G}_1\left[\xi_{1i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)\right]}\frac{\partial\xi_{1i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)}{\partial y_i}\right\}\right\}\\ &+\sum_{i=1}^{n}\left\{\log\mathcal{G}_2\left[\xi_{2i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_2)\right] + \delta_{2i}\log\left\{-\frac{\mathcal{G}_2'\left[\xi_{2i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_2)\right]}{\mathcal{G}_2\left[\xi_{2i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_2)\right]}\frac{\partial\xi_{2i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_2)}{\partial y_i}\right\}\right\}.\end{aligned}$$

The gradient of equation (1) can be calculated as

$$\boldsymbol{\nabla}_{\boldsymbol{\alpha}}\ell_p(\boldsymbol{\alpha}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}}\ell(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\mathcal{S}},$$

where $\boldsymbol{\nabla}_{\boldsymbol{\alpha}}\ell(\boldsymbol{\alpha}) = \left(\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})^\top, \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha})^\top, \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha})^\top\right)^\top$. where $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha})$ and $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha})$ can be obtained as $\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}_0} = \left[\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{011}}\cdots\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{0QJ_Q}}\right]^\top$, $\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}_1} = \left[\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{111}}\cdots\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{1Q_1J_{1Q_1}}}\right]^\top$ and $\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}_2} = \left[\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{21}}\cdots\dfrac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{2Q_2J_{2Q_2}}}\right]^\top$. In particular, the scalar derivatives of $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha})$ and $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha})$ can be calculated as

$$
\begin{aligned}
\frac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{0j}} &= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\right\} + \sum_{i=1}^{n}\delta_{1i}\left\{\left[-\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial y_i}\right]^{-1}\left[-\frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial y_i} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial y_i} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{0j}}\right]\right\} \\
&\quad + \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\right\} + \sum_{i=1}^{n}\delta_{2i}\left\{\left[-\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial y_i}\right]^{-1}\left[-\frac{\mathcal{G}_2''}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial y_i} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial y_i} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial^2\xi_{2i}}{\partial y_i\partial\alpha_{0j}}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}} + \delta_{1i}\left[\frac{\mathcal{G}_1''}{\mathcal{G}_1'}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}} + \frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{0j}}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-1}\right]\right\} \\
&\quad + \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}} + \delta_{2i}\left[\frac{\mathcal{G}_2''}{\mathcal{G}_2'}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}} + \frac{\partial^2\xi_{2i}}{\partial y_i\partial\alpha_{0j}}\left(\frac{\partial\xi_{2i}}{\partial y_i}\right)^{-1}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\left[\frac{\mathcal{G}_1'}{\mathcal{G}_1} + \delta_{1i}\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1'} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\right)\right] + \frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\left[\frac{\mathcal{G}_2'}{\mathcal{G}_2} + \delta_{2i}\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2'} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\right)\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\Delta_1 + \frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\Delta_2\right\},
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\frac{\partial\ell(\boldsymbol{\alpha})}{\partial\alpha_{1j}} &= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}}\right\} + \sum_{i=1}^{n}\delta_{1i}\left\{\left[-\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial y_i}\right]^{-1}\left[-\frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}}\frac{\partial\xi_{1i}}{\partial y_i} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}}\frac{\partial\xi_{1i}}{\partial y_i} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{1j}}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}} + \delta_{1i}\left[\frac{\mathcal{G}_1''}{\mathcal{G}_1'}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{1j}} + \frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{1j}}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-1}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial\xi_{1i}}{\partial\alpha_{1j}}\left[\frac{\mathcal{G}_1'}{\mathcal{G}_1} + \delta_{1i}\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1'} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\right)\right] + \frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{1j}}\delta_{1i}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-1}\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial\xi_{1i}}{\partial\alpha_{1j}}\Delta_1 + \frac{\partial^2\xi_{1i}}{\partial y_i\partial\alpha_{1j}}\Omega_1\right\},
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\alpha})}{\partial \alpha_{2j}} &= \sum_{i=1}^{n} \left\{ \frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} \right\} + \sum_{i=1}^{n} \delta_{2i} \left\{ \left[-\frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial \xi_{2i}}{\partial y_i} \right]^{-1} \left[-\frac{\mathcal{G}_2''}{\mathcal{G}_2} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} \frac{\partial \xi_{2i}}{\partial y_i} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} \frac{\partial \xi_{2i}}{\partial y_i} - \frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}} \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} + \delta_{2i} \left[\frac{\mathcal{G}_2''}{\mathcal{G}_2'} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}} \left(\frac{\partial \xi_{2i}}{\partial y_i} \right)^{-1} \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} \left[\frac{\mathcal{G}_2'}{\mathcal{G}_2} + \delta_{2i} \left(\frac{\mathcal{G}_2''}{\mathcal{G}_2'} - \frac{\mathcal{G}_2'}{\mathcal{G}_2} \right) \right] + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}} \delta_{2i} \left(\frac{\partial \xi_{2i}}{\partial y_i} \right)^{-1} \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{2i}}{\partial \alpha_{2j}} \Delta_2 + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}} \Omega_2 \right\},
\end{aligned}
\tag{4}
$$

where $\xi_{\nu i} = \xi_{\nu i}(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\nu)$, $\Delta_\nu = \left[\frac{\mathcal{G}_\nu'}{\mathcal{G}_\nu} + \delta_{\nu i} \left(\frac{\mathcal{G}_\nu''}{\mathcal{G}_\nu'} - \frac{\mathcal{G}_\nu'}{\mathcal{G}_\nu} \right) \right]$ and $\Omega_\nu = \delta_{\nu i} \left(\frac{\partial \xi_{\nu i}}{\partial y_i} \right)^{-1}$. The last terms of equations (2), (3) and (4) allow to express $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0} \ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1} \ell(\boldsymbol{\alpha})$ and $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_2} \ell(\boldsymbol{\alpha})$ as follow

$$
\begin{aligned}
\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0} \ell(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} \left[\Delta_1 \frac{\partial \xi_{1i}}{\partial \boldsymbol{\alpha}_0} \; + \; \Delta_2 \frac{\partial \xi_{2i}}{\partial \boldsymbol{\alpha}_0} \right], \\
\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1} \ell(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} \left[\Delta_1 \frac{\partial \xi_{1i}}{\partial \boldsymbol{\alpha}_1} \; + \; \Omega_1 \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \boldsymbol{\alpha}_1} \right], \\
\boldsymbol{\nabla}_{\boldsymbol{\alpha}_2} \ell(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} \left[\Delta_2 \frac{\partial \xi_{2i}}{\partial \boldsymbol{\alpha}_2} \; + \; \Omega_2 \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \boldsymbol{\alpha}_2} \right],
\end{aligned}
$$

where, for all $i = 1, ..., n$ and $\nu = 1, 2$, $\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_0} = \left[\frac{\partial \xi_{\nu i}}{\partial \alpha_{011}} \cdots \frac{\partial \xi_{\nu i}}{\partial \alpha_{0Q J_Q}} \right]^\top$, $\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu} = \left[\frac{\partial \xi_{\nu i}}{\partial \alpha_{\nu 11}} \cdots \frac{\partial \xi_{\nu i}}{\partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \right]^\top$ and $\frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu} = \left[\frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \alpha_{\nu 11}} \cdots \frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \right]^\top$. These expressions can be calculated using the design vectors defined in Section 2.2 as

$$
\begin{aligned}
\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_0} &= \left(\boldsymbol{\mathcal{Q}}_1(\mathbf{x}_{1i}^0)^\top, \ldots, \boldsymbol{\mathcal{Q}}_Q(\mathbf{x}_{Qi}^0)^\top \right)^\top = \boldsymbol{\mathcal{Q}}_i^0, \\
\frac{\partial \xi_{\nu i}}{\partial y_i} &= \lim_{\varepsilon \to 0} \left\{ \frac{\boldsymbol{\mathcal{Q}}_{\nu 0}(y_i + \varepsilon) - \boldsymbol{\mathcal{Q}}_{\nu 0}(y_i - \varepsilon)}{2\varepsilon} \right\}^\top \boldsymbol{\Gamma}_{\nu 0} \tilde{\boldsymbol{\alpha}}_{\nu 0} = \boldsymbol{\mathcal{Q}}'_{\nu 0}(y_i)^\top \boldsymbol{\Gamma}_{\nu 0} \tilde{\boldsymbol{\alpha}}_{\nu 0}, \\
\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_{\nu q_\nu}} &= \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota \triangle}(y_i) & \text{if } \boldsymbol{\alpha}_{\nu q_\nu} = \boldsymbol{\alpha}_{\nu 0} \\ \boldsymbol{\mathcal{Q}}_{\nu q_\nu}(\mathbf{x}_{\nu q_\nu i}^1) & \text{otherwise,} \end{cases} \\
\frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_{\nu q_\nu}} &= \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota \triangle \prime}(y_i) & \text{if } \boldsymbol{\alpha}_{\nu q_\nu} = \boldsymbol{\alpha}_{\nu 0} \\ \mathbf{0} & \text{otherwise,} \end{cases}
\end{aligned}
$$

where $\boldsymbol{\mathcal{Q}}'_{\nu 0}(y_i)$ can be conveniently obtained using a finite-difference method. Moreover, we define the design vectors $\boldsymbol{\mathcal{Q}}^{\iota\triangle}_{\nu 0}(y_i)$ and $\boldsymbol{\mathcal{Q}}^{\iota\triangle'}_{\nu 0}(y_i)$ as

$$\boldsymbol{\mathcal{Q}}^{\iota\triangle}_{\nu 0}(y_i) = \begin{bmatrix} \sum_{j_{\nu 0}=1}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i) \\ \left[\sum_{j_{\nu 0}=2}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 02}\right) \\ \left[\sum_{j_{\nu 0}=3}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 03}\right)] \\ \vdots \\ \mathcal{Q}_{\nu 0 J_{\nu 0}}(y_i) \exp\left(\alpha_{\nu 0 J_{\nu 0}}\right) \end{bmatrix} \quad \boldsymbol{\mathcal{Q}}^{\iota\triangle'}_{\nu 0}(y_i) = \begin{bmatrix} \sum_{j_{\nu 0}=1}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i) \\ \left[\sum_{j_{\nu 0}=2}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 02}\right) \\ \left[\sum_{j_{\nu 0}=3}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 03}\right)] \\ \vdots \\ \mathcal{Q}'_{\nu 0 J_{\nu 0}}(y_i) \exp\left(\alpha_{\nu 0 J_{\nu 0}}\right) \end{bmatrix}.$$

On the other hand, when censoring is non-informative the penalized log-likelihood function is

$$\ell_p(\boldsymbol{\gamma}) = \ell(\boldsymbol{\gamma}) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{\mathcal{S}} \boldsymbol{\gamma}, \tag{5}$$

where $\ell(\boldsymbol{\gamma})$ can be written as

$$\begin{aligned} \ell(\boldsymbol{\gamma}) = &\sum_{i=1}^{n} \left\{ \log \mathcal{G}_1\left[\xi_{1i}(\boldsymbol{\gamma}_1)\right] + \delta_{1i} \log \left\{ -\frac{\mathcal{G}'_1\left[\xi_{1i}(\boldsymbol{\gamma}_1)\right]}{\mathcal{G}_1\left[\xi_{1i}(\boldsymbol{\gamma}_1)\right]} \frac{\partial \xi_{1i}(\boldsymbol{\gamma}_1)}{\partial y_i} \right\} \right\} \\ &+ \sum_{i=1}^{n} \left\{ \log \mathcal{G}_2\left[\xi_{2i}(\boldsymbol{\gamma}_2)\right] + \delta_{2i} \log \left\{ -\frac{\mathcal{G}'_2\left[\xi_{2i}(\boldsymbol{\gamma}_2)\right]}{\mathcal{G}_2\left[\xi_{2i}(\boldsymbol{\gamma}_2)\right]} \frac{\partial \xi_{2i}(\boldsymbol{\gamma}_2)}{\partial y_i} \right\} \right\}. \end{aligned}$$

The gradient of (5) can be calculated as

$$\boldsymbol{\nabla}_{\boldsymbol{\gamma}} \ell_p(\boldsymbol{\gamma}) = \boldsymbol{\nabla}_{\boldsymbol{\gamma}} \ell(\boldsymbol{\gamma}) - \boldsymbol{\gamma} \boldsymbol{\mathcal{S}},$$

where $\boldsymbol{\nabla}_{\boldsymbol{\gamma}} \ell(\boldsymbol{\gamma}) = \left(\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1} \ell(\boldsymbol{\gamma})^\top,\ \boldsymbol{\nabla}_{\boldsymbol{\gamma}_2} \ell(\boldsymbol{\gamma})^\top\right)^\top$. In addition, $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1} \ell(\boldsymbol{\gamma})$ and $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_2} \ell(\boldsymbol{\gamma})$ can be calculated as $\dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1} = \left[\dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{111}} \cdots \dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{1K_1 J_{1K_1}}}\right]^\top$ and $\dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2} = \left[\dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{211}} \cdots \dfrac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{2K_2 J_{2K_2}}}\right]^\top$. Furthermore,

the scalar derivatives of $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1}\ell(\boldsymbol{\gamma})$ and $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_2}\ell(\boldsymbol{\gamma})$ can be obtained as

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{1j}} &= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}}\right\} + \sum_{i=1}^{n}\delta_{1i}\left\{\left[-\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial y_i}\right]^{-1}\left[-\frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}}\frac{\partial \xi_{1i}}{\partial y_i} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}}\frac{\partial \xi_{1i}}{\partial y_i} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2 \xi_{1i}}{\partial y_i \partial \gamma_{1j}}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}} + \delta_{1i}\left[\frac{\mathcal{G}_1''}{\mathcal{G}_1'}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial \gamma_{1j}} + \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \gamma_{1j}}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-1}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial \xi_{1i}}{\partial \gamma_{1j}}\left[\frac{\mathcal{G}_1'}{\mathcal{G}_1} + \delta_{1i}\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1'} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\right)\right] + \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \gamma_{1j}}\delta_{1i}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-1}\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial \xi_{1i}}{\partial \gamma_{1j}}\Delta_1 + \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \gamma_{1j}}\Omega_1\right\},
\end{aligned} \tag{6}$$

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_{2j}} &= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}}\right\} + \sum_{i=1}^{n}\delta_{2i}\left\{\left[-\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial y_i}\right]^{-1}\left[-\frac{\mathcal{G}_2''}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}}\frac{\partial \xi_{2i}}{\partial y_i} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}}\frac{\partial \xi_{2i}}{\partial y_i} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial^2 \xi_{2i}}{\partial y_i \partial \gamma_{2j}}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}} + \delta_{2i}\left[\frac{\mathcal{G}_2''}{\mathcal{G}_2'}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial \gamma_{2j}} + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \gamma_{2j}}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-1}\right]\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial \xi_{2i}}{\partial \gamma_{2j}}\left[\frac{\mathcal{G}_2'}{\mathcal{G}_2} + \delta_{2i}\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2'} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\right)\right] + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \gamma_{2j}}\delta_{2i}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-1}\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial \xi_{2i}}{\partial \gamma_{2j}}\Delta_2 + \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \gamma_{2j}}\Omega_2\right\},
\end{aligned} \tag{7}$$

where $\xi_{\nu i} = \xi_{\nu i}(\boldsymbol{\gamma}_\nu)$. The last terms of equations (6) and (7) allow $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1}\ell(\boldsymbol{\gamma})$ and $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_2}\ell(\boldsymbol{\gamma})$ to be expressed as

$$\begin{aligned}
\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1}\ell(\boldsymbol{\gamma}) &= \sum_{i=1}^{n}\left[\Delta_1\frac{\partial \xi_{1i}}{\partial \boldsymbol{\gamma}_1} + \Omega_1\frac{\partial^2 \xi_{1i}}{\partial y_i \partial \boldsymbol{\gamma}_1}\right] \\
\boldsymbol{\nabla}_{\boldsymbol{\gamma}_2}\ell(\boldsymbol{\gamma}) &= \sum_{i=1}^{n}\left[\Delta_2\frac{\partial \xi_{2i}}{\partial \boldsymbol{\gamma}_2} + \Omega_2\frac{\partial^2 \xi_{2i}}{\partial y_i \partial \boldsymbol{\gamma}_2}\right],
\end{aligned}$$

where $\dfrac{\partial \xi_{\nu i}}{\partial \boldsymbol{\gamma}_\upsilon} = \left[\dfrac{\partial \xi_{\nu i}}{\partial \gamma_{\nu 11}} \cdots \dfrac{\partial \xi_{\nu i}}{\partial \gamma_{\nu K_\nu J_{\nu K_\nu}}}\right]^\top$ and $\dfrac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\gamma}_\nu} = \left[\dfrac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \gamma_{\nu 11}} \cdots \dfrac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \gamma_{\nu K_\nu J_{\nu K_\nu}}}\right]^\top$ for all $i = 1, ..., n$ and $\nu = 1, 2$. Furthermore, $\dfrac{\partial \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i}$, can be generically calculated using

$$\frac{\partial \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i} = \lim_{\varepsilon \to 0}\left\{\frac{\boldsymbol{\mathcal{Q}}_{\nu 0}(y_i + \varepsilon) - \boldsymbol{\mathcal{Q}}_{\nu 0}(y_i - \varepsilon)}{2\varepsilon}\right\}^\top \boldsymbol{\Gamma}_{\nu 0}\hat{\boldsymbol{\gamma}}_{\nu 0} = \boldsymbol{\mathcal{Q}}_{\nu 0}'(y_i)^\top \boldsymbol{\Gamma}_{\nu 0}\hat{\boldsymbol{\gamma}}_{\nu 0},$$

where $\boldsymbol{\mathcal{Q}}'_{\nu 0}(y_i)$ can also be calculated using a finite-difference method. The design vectors for $\dfrac{\partial \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial \boldsymbol{\gamma}_\nu}$ and $\dfrac{\partial^2 \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i \partial \boldsymbol{\gamma}_\nu}$ can be obtained using

$$\frac{\partial \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial \boldsymbol{\gamma}_{\nu k_\nu}} = \begin{cases} \boldsymbol{\mathcal{Q}}^{\triangle}_{\nu 0}(y_i) & \text{if } \boldsymbol{\gamma}_{\nu k_\nu} = \boldsymbol{\gamma}_{\nu 0} \\ \boldsymbol{\mathcal{Q}}_{\nu k_\nu}(\mathbf{x}_{\nu k_\nu i}) & \text{otherwise,} \end{cases}$$

$$\frac{\partial^2 \xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i \partial \boldsymbol{\gamma}_{\nu k_\nu}} = \begin{cases} \boldsymbol{\mathcal{Q}}^{\triangle'}_{\nu 0}(y_i) & \text{if } \boldsymbol{\gamma}_{\nu k_\nu} = \boldsymbol{\gamma}_{\nu 0} \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Finally, we have that

$$\boldsymbol{\mathcal{Q}}^{\triangle}_{\nu 0}(y_i) = \begin{bmatrix} \sum_{j_{\nu 0}=1}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i) \\ \left[\sum_{j_{\nu 0}=2}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 02}\right) \\ \left[\sum_{j_{\nu 0}=3}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 03}\right)] \\ \vdots \\ \mathcal{Q}_{\nu 0 J_{\nu 0}}(y_i) \exp\left(\gamma_{\nu 0 J_{\nu 0}}\right) \end{bmatrix} \quad \boldsymbol{\mathcal{Q}}^{\triangle'}_{\nu 0}(y_i) = \begin{bmatrix} \sum_{j_{\nu 0}=1}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i) \\ \left[\sum_{j_{\nu 0}=2}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 02}\right) \\ \left[\sum_{j_{\nu 0}=3}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 03}\right)] \\ \vdots \\ \mathcal{Q}'_{\nu 0 J_{\nu 0}}(y_i) \exp\left(\gamma_{\nu 0 J_{\nu 0}}\right) \end{bmatrix}.$$

B.2. Informative and Non-informative Hessians

The informative penalized Hessian can be obtained as

$$\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell_p(\boldsymbol{\alpha}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}) - \boldsymbol{\mathcal{S}},$$

where $\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha})$ is

$$\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}) = \begin{bmatrix} \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_1} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_2} \ell(\boldsymbol{\alpha}) \\ \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_0} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2} \ell(\boldsymbol{\alpha}) \\ \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_0} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_1} \ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_2} \ell(\boldsymbol{\alpha}) \end{bmatrix}. \tag{8}$$

In addition, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_\upsilon\boldsymbol{\alpha}_\kappa}\ell(\boldsymbol{\alpha}) = \dfrac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}_\upsilon\partial\boldsymbol{\alpha}_\kappa^\top}$, for all $\upsilon = 0,1,2$ and $\kappa = 0,1,2$. This expression is calculated using

$$\boldsymbol{\nabla}_{\boldsymbol{\alpha}_\upsilon\boldsymbol{\alpha}_\kappa}\ell(\boldsymbol{\alpha}) = \begin{bmatrix} \dfrac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\alpha_{\upsilon 11}\partial\alpha_{\kappa 11}} & \cdots & \dfrac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\alpha_{\upsilon 11}\partial\alpha_{\kappa Q_\kappa J_{\kappa Q_\kappa}}} \\ \cdots & \ddots & \cdots \\ \dfrac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\alpha_{\upsilon Q_\upsilon J_{\upsilon Q_\upsilon}}\partial\alpha_{\kappa 11}} & \cdots & \dfrac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\alpha_{\upsilon Q_\upsilon J_{\upsilon Q_\upsilon}}\partial\alpha_{\kappa Q_\kappa J_{\kappa Q_\kappa}}} \end{bmatrix}.$$

Since $\boldsymbol{\alpha}_1$ appears only in $\xi_{1i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_1)$ and $\boldsymbol{\alpha}_2$ only in $\xi_{2i}(\boldsymbol{\alpha}_0,\boldsymbol{\alpha}_2)$, then $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha}) = \mathbf{0}$. Hence, (8) can be written as

$$\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\ell(\boldsymbol{\alpha}) = \begin{bmatrix} \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha}) \\ \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha}) & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha}) & \mathbf{0} \\ \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha}) & \mathbf{0} & \boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha}) \end{bmatrix}. \tag{9}$$

In equation (9), the scalar derivatives of $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha})$, $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1}\ell(\boldsymbol{\alpha})$ and $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_2\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha})$, can be calculated as

$$\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\alpha})}{\partial\alpha_{0j}\partial\alpha_{0k}} &= \sum_{i=1}^{n}\Bigg\{\frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_1'''}{\mathcal{G}_1}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} \\
&\quad - \frac{\mathcal{G}_1''}{\mathcal{G}_1}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2\xi_{1i}}{\partial\alpha_{0j}\partial\alpha_{0k}} + \frac{\mathcal{G}_1''}{\mathcal{G}_1'}\delta_{1i}\frac{\partial^2\xi_{1i}}{\partial\alpha_{0j}\partial\alpha_{0k}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\delta_{1i}\frac{\partial^2\xi_{1i}}{\partial\alpha_{0j}\partial\alpha_{0k}}\Bigg\} \\
&\quad + \sum_{i=1}^{n}\Bigg\{\frac{\mathcal{G}_2''}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_2'''}{\mathcal{G}_2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} \\
&\quad - \frac{\mathcal{G}_2''}{\mathcal{G}_2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}} + \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial^2\xi_{2i}}{\partial\alpha_{0j}\partial\alpha_{0k}} + \frac{\mathcal{G}_2''}{\mathcal{G}_2'}\delta_{2i}\frac{\partial^2\xi_{2i}}{\partial\alpha_{0j}\partial\alpha_{0k}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\delta_{2i}\frac{\partial^2\xi_{2i}}{\partial\alpha_{0j}\partial\alpha_{0k}}\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}}\left[\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right) + \delta_{1i}\left(\frac{\mathcal{G}_1'''}{\mathcal{G}_1} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2} - \frac{\mathcal{G}_1''}{\mathcal{G}_1} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right)\right] \\
&\quad + \frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}}\left[\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right) + \delta_{2i}\left(\frac{\mathcal{G}_2'''}{\mathcal{G}_2} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2} - \frac{\mathcal{G}_2''}{\mathcal{G}_2} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right)\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{1i}}{\partial\alpha_{0j}}\frac{\partial\xi_{1i}}{\partial\alpha_{0k}}\Phi_1 + \frac{\partial\xi_{2i}}{\partial\alpha_{0j}}\frac{\partial\xi_{2i}}{\partial\alpha_{0k}}\Phi_2\Bigg\},
\end{aligned} \tag{10}$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \alpha_{0j} \partial \alpha_{1k}} &= \sum_{i=1}^{n} \Bigg\{ \frac{\mathcal{G}_1''}{\mathcal{G}_1} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'''}{\mathcal{G}_1} \delta_{1i} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2} \delta_{1i} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} \\
&\quad - \frac{\mathcal{G}_1''}{\mathcal{G}_1} \delta_{1i} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2} \delta_{1i} \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'}{\mathcal{G}_1} \frac{\partial^2 \xi_{1i}}{\partial \alpha_{0j} \partial \alpha_{1k}} + \frac{\mathcal{G}_1''}{\mathcal{G}_1'} \delta_{1i} \frac{\partial^2 \xi_{1i}}{\partial \alpha_{0j} \partial \alpha_{1k}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1} \delta_{1i} \frac{\partial^2 \xi_{1i}}{\partial \alpha_{0j} \partial \alpha_{1k}} \Bigg\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} \left[\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2} \right) + \delta_{1i} \left(\frac{\mathcal{G}_1'''}{\mathcal{G}_1} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2} - \frac{\mathcal{G}_1''}{\mathcal{G}_1} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2} \right) \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{1i}}{\partial \alpha_{0j}} \frac{\partial \xi_{1i}}{\partial \alpha_{1k}} \Phi_1 \right\},
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \alpha_{0j} \partial \alpha_{2k}} &= \sum_{i=1}^{n} \Bigg\{ \frac{\mathcal{G}_2''}{G_2} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} - \frac{\mathcal{G}_2'^2}{G_2^2} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'''}{\mathcal{G}_2} \delta_{2i} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2} \delta_{2i} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} \\
&\quad - \frac{\mathcal{G}_2''}{\mathcal{G}_2} \delta_{2i} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2} \delta_{2i} \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'}{\mathcal{G}_2} \frac{\partial^2 \xi_{2i}}{\partial \alpha_{0j} \partial \alpha_{2k}} + \frac{\mathcal{G}_2''}{\mathcal{G}_2'} \delta_{2i} \frac{\partial^2 \xi_{2i}}{\partial \alpha_{0j} \partial \alpha_{2k}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2} \delta_{2i} \frac{\partial^2 \xi_{2i}}{\partial \alpha_{0j} \partial \alpha_{2k}} \Bigg\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} \left[\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2} \right) + \delta_{2i} \left(\frac{\mathcal{G}_2'''}{\mathcal{G}_2} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2} - \frac{\mathcal{G}_2''}{\mathcal{G}_2} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2} \right) \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial \xi_{2i}}{\partial \alpha_{0j}} \frac{\partial \xi_{2i}}{\partial \alpha_{2k}} \Phi_2 \right\},
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \alpha_{1j}\partial \alpha_{1k}} &= \sum_{i=1}^{n}\Bigg\{ \frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'''}{\mathcal{G}_1}\delta_{1i}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2}\delta_{1i}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} \\
&\quad - \frac{\mathcal{G}_1''}{\mathcal{G}_1}\delta_{1i}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\delta_{1i}\frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}} + \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2 \xi_{1i}}{\partial \alpha_{1j}\partial \alpha_{1k}} + \frac{\mathcal{G}_1''}{\mathcal{G}_1'}\delta_{1i}\frac{\partial^2 \xi_{1i}}{\partial \alpha_{1j}\partial \alpha_{1k}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\delta_{1i}\frac{\partial^2 \xi_{1i}}{\partial \alpha_{1j}\partial \alpha_{1k}} \\
&\quad + \frac{\partial^3 \xi_{1i}}{\partial y_i \partial \alpha_{1j}\partial \alpha_{1k}}\delta_{1i}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-1} - \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1k}}\frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1j}}\delta_{1i}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-2}\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{ \frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}}\left[\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right) + \delta_{1i}\left(\frac{\mathcal{G}_1'''}{\mathcal{G}_1} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2} - \frac{\mathcal{G}_1''}{\mathcal{G}_1} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right)\right] \\
&\quad + \frac{\partial^2 \xi_{1i}}{\partial \alpha_{1j}\partial \alpha_{1k}}\left[\frac{\mathcal{G}_1'}{\mathcal{G}_1} + \delta_{1i}\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1'} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\right)\right] - \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1k}}\frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1j}}\left[\delta_{1i}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-2}\right] \\
&\quad + \frac{\partial^3 \xi_{1i}}{\partial y_i \partial \alpha_{1j}\partial \alpha_{1k}}\left[\delta_{1i}\left(\frac{\partial \xi_{1i}}{\partial y_i}\right)^{-1}\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{ \frac{\partial \xi_{1i}}{\partial \alpha_{1j}}\frac{\partial \xi_{1i}}{\partial \alpha_{1k}}\Phi_1 + \frac{\partial^2 \xi_{1i}}{\partial \alpha_{1j}\partial \alpha_{1k}}\Delta_1 - \frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1k}}\frac{\partial^2 \xi_{1i}}{\partial y_i \partial \alpha_{1j}}\Psi_1 + \frac{\partial^3 \xi_{1i}}{\partial y_i \partial \alpha_{1j}\partial \alpha_{1k}}\Omega_1\Bigg\},
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \alpha_{2j}\partial \alpha_{2k}} &= \sum_{i=1}^{n}\Bigg\{ \frac{\mathcal{G}_2''}{\mathcal{G}_2}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'''}{\mathcal{G}_2}\delta_{2i}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2}\delta_{2i}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} \\
&\quad - \frac{\mathcal{G}_2''}{\mathcal{G}_2}\delta_{2i}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\delta_{2i}\frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}} + \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial^2 \xi_{2i}}{\partial \alpha_{2j}\partial \alpha_{2k}} + \frac{\mathcal{G}_2''}{\mathcal{G}_2'}\delta_{2i}\frac{\partial^2 \xi_{2i}}{\partial \alpha_{2j}\partial \alpha_{2k}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\delta_{2i}\frac{\partial^2 \xi_{2i}}{\partial \alpha_{2j}\partial \alpha_{2k}} \\
&\quad + \frac{\partial^3 \xi_{2i}}{\partial y_i \partial \alpha_{2j}\partial \alpha_{2k}}\delta_{2i}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-1} - \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2k}}\frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}}\delta_{2i}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-2}\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{ \frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}}\left[\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right) + \delta_{2i}\left(\frac{\mathcal{G}_2'''}{\mathcal{G}_2} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2} - \frac{\mathcal{G}_2''}{\mathcal{G}_2} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right)\right] \\
&\quad + \frac{\partial^2 \xi_{2i}}{\partial \alpha_{2j}\partial \alpha_{2k}}\left[\frac{\mathcal{G}_2'}{\mathcal{G}_2} + \delta_{2i}\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2'} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\right)\right] - \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2k}}\frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}}\left[\delta_{2i}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-2}\right] \\
&\quad + \frac{\partial^3 \xi_{2i}}{\partial y_i \partial \alpha_{2j}\partial \alpha_{2k}}\left[\delta_{2i}\left(\frac{\partial \xi_{2i}}{\partial y_i}\right)^{-1}\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{ \frac{\partial \xi_{2i}}{\partial \alpha_{2j}}\frac{\partial \xi_{2i}}{\partial \alpha_{2k}}\Phi_2 + \frac{\partial^2 \xi_{2i}}{\partial \alpha_{2j}\partial \alpha_{2k}}\Delta_2 - \frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2k}}\frac{\partial^2 \xi_{2i}}{\partial y_i \partial \alpha_{2j}}\Psi_2 + \frac{\partial^3 \xi_{2i}}{\partial y_i \partial \alpha_{2j}\partial \alpha_{2k}}\Omega_2\Bigg\},
\end{aligned}
\tag{14}
$$

where $\Phi_\nu = \delta_{\nu i}\left(\frac{\mathcal{G}_\nu'''}{\mathcal{G}_\nu} - \frac{\mathcal{G}_\nu''^2}{\mathcal{G}_\nu'^2} - \frac{\mathcal{G}_\nu''}{\mathcal{G}_\nu} + \frac{\mathcal{G}_\nu'^2}{\mathcal{G}_\nu^2}\right)$ and $\Psi_\nu = \left[\delta_{\nu i}\left(\frac{\partial \xi_{\nu i}}{\partial y_i}\right)^{-2}\right]$. Collecting the last terms of (10), (11), (12), (13) and (14), we obtain

$$\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_0 \partial \boldsymbol{\alpha}_0^\top} = \sum_{i=1}^n \left\{ \Phi_1 \frac{\partial \xi_{1i}}{\partial \boldsymbol{\alpha}_0}\left[\frac{\partial \xi_{1i}}{\partial \boldsymbol{\alpha}_0}\right]^\top + \Phi_2 \frac{\partial \xi_{2i}}{\partial \boldsymbol{\alpha}_0}\left[\frac{\partial \xi_{2i}}{\partial \boldsymbol{\alpha}_0}\right]^\top \right\},$$
$$\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_0 \partial \boldsymbol{\alpha}_\nu^\top} = \sum_{i=1}^n \left\{ \Phi_\nu \frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_0}\left[\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu}\right]^\top \right\},$$
$$\frac{\partial^2 \ell(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top} = \sum_{i=1}^n \left\{ \Phi_\nu \frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu}\left[\frac{\partial \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu}\right]^\top + \Delta_\nu \frac{\partial^2 \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top} - \Psi_\nu \frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu}\left[\frac{\partial^2 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu}\right]^\top + \Omega_\nu \frac{\partial^3 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top} \right\},$$

where

$$\frac{\partial^2 \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top} = \begin{bmatrix} \frac{\partial^2 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial \alpha_{\nu 11} \partial \alpha_{\nu 11}} & \cdots & \frac{\partial^2 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial \alpha_{\nu 11} \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \\ \cdots & \ddots & \cdots \\ \frac{\partial^2 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}} \partial \alpha_{\nu 11}} & \cdots & \frac{\partial^2 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}} \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \end{bmatrix},$$

$$\frac{\partial^3 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top} = \begin{bmatrix} \frac{\partial^3 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial y_i \partial \alpha_{\nu 11} \partial \alpha_{\nu 11}} & \cdots & \frac{\partial^3 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial y_i \partial \alpha_{\nu 11} \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \\ \cdots & \ddots & \cdots \\ \frac{\partial^3 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial y_i \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}} \partial \alpha_{\nu 11}} & \cdots & \frac{\partial^3 \xi_{\nu i}(\boldsymbol{\alpha}_\nu)}{\partial y_i \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}} \partial \alpha_{\nu Q_\nu J_{\nu Q_\nu}}} \end{bmatrix}.$$

In particular, the design sub-matrices of $\frac{\partial^2 \xi_{\nu i}}{\partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top}$ and $\frac{\partial^3 \xi_{\nu i}}{\partial y_i \partial \boldsymbol{\alpha}_\nu \partial \boldsymbol{\alpha}_\nu^\top}$ are calculated using

$$\frac{\partial^2 \xi_{\nu i}(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\nu)}{\partial \boldsymbol{\alpha}_{\nu q_\nu} \partial \boldsymbol{\alpha}_{\nu s_\nu}^\top} = \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota \triangle \triangle}(y_i) & \text{if } \boldsymbol{\alpha}_{\nu q_\nu} = \boldsymbol{\alpha}_{\nu s_\nu} = \boldsymbol{\alpha}_{\nu 0} \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

$$\frac{\partial^3 \xi_{\nu i}(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\nu)}{\partial y_i \partial \boldsymbol{\alpha}_{\nu q_\nu} \boldsymbol{\alpha}_{\nu s_\nu}^\top} = \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota \triangle \triangle'}(y_i) & \text{if } \boldsymbol{\alpha}_{\nu q_\nu} = \boldsymbol{\alpha}_{\nu s_\nu} = \boldsymbol{\alpha}_{\nu 0} \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota\triangle\triangle}(y_i)$ and $\boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota\triangle\triangle'}(y_i)$ are defined as

$$\boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota\triangle\triangle}(y_i) = \begin{cases} \dfrac{\partial^2 \xi_{\nu i}}{\partial \alpha_{\nu 0 j_{\nu 0}} \partial \alpha_{\nu 0 k_{\nu 0}}} = \left[\sum_{j_{\nu 0}}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 0 j_{\nu 0}}\right) & \text{if } j = k \neq 1 \\ \dfrac{\partial^2 \xi_{\nu i}}{\partial \alpha_{\nu 0 j_{\nu 0}} \partial \alpha_{\nu 0 k_{\nu 0}}} = 0 & \text{otherwise,} \end{cases}$$

$$\boldsymbol{\mathcal{Q}}_{\nu 0}^{\iota\triangle\triangle'}(y_i) = \begin{cases} \dfrac{\partial^3 \xi_{\nu i}}{\partial y_i \alpha_{\nu 0 j_{\nu 0}} \partial \alpha_{\nu 0 k_{\nu 0}}} = \left[\sum_{j_{\nu 0}}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\alpha_{\nu 0 j_{\nu 0}}\right) & \text{if } j = k \neq 1 \\ \dfrac{\partial^3 \xi_{\nu i}}{\partial y_i \alpha_{\nu 0 j_{\nu 0}} \partial \alpha_{\nu 0 k_{\nu 0}}} = 0 & \text{otherwise.} \end{cases}$$

On the other hand, the non-informative penalized Hessian is

$$\boldsymbol{\nabla}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell_p(\boldsymbol{\gamma}) = \boldsymbol{\nabla}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\boldsymbol{\gamma}) - \boldsymbol{\mathcal{S}}.$$

Since $\xi_{1i}(\boldsymbol{\gamma}_1)$ and $\xi_{2i}(\boldsymbol{\gamma}_2)$) do not have parameters in common, $\boldsymbol{\nabla}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\boldsymbol{\gamma})$ can be written as

$$\boldsymbol{\nabla}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \ell(\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{\nabla}_{\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1} \ell(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\nabla}_{\boldsymbol{\gamma}_2\boldsymbol{\gamma}_2} \ell(\boldsymbol{\gamma}) \end{bmatrix},$$

where $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_\nu\boldsymbol{\gamma}_\nu} \ell(\boldsymbol{\gamma}) = \dfrac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_\nu \partial \boldsymbol{\gamma}_\nu^\top}$. This expression can be obtained using

$$\boldsymbol{\nabla}_{\boldsymbol{\gamma}_\nu\boldsymbol{\gamma}_\nu} \ell(\boldsymbol{\gamma}) = \begin{bmatrix} \dfrac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_{\nu 11} \partial \gamma_{\nu 11}} & \cdots & \dfrac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_{\nu 11} \partial \gamma_{\nu K_\nu J_{\nu K_\nu}}} \\ \cdots & \ddots & \cdots \\ \dfrac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_{\nu K_\nu J_{\nu K_\nu}} \partial \gamma_{\nu 11}} & \cdots & \dfrac{\partial^2 \ell(\boldsymbol{\gamma})}{\partial \gamma_{\nu K_\nu J_{\nu K_\nu}} \partial \gamma_{\nu K_\nu J_{\nu K_\nu}}} \end{bmatrix}.$$

Furthermore, the scalar derivatives of $\boldsymbol{\nabla}_{\gamma_1\gamma_1}\ell(\boldsymbol{\gamma})$ and $\boldsymbol{\nabla}_{\gamma_2\gamma_2}\ell(\boldsymbol{\gamma})$ are

$$
\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\gamma_{1j}\partial\gamma_{1k}} &= \sum_{i=1}^{n}\Bigg\{\frac{\mathcal{G}_1''}{\mathcal{G}_1}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} + \frac{\mathcal{G}_1'''}{\mathcal{G}_1}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} \\
&\quad - \frac{\mathcal{G}_1''}{\mathcal{G}_1}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\delta_{1i}\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}} + \frac{\mathcal{G}_1'}{\mathcal{G}_1}\frac{\partial^2\xi_{1i}}{\partial\gamma_{1j}\partial\gamma_{1k}} + \frac{\mathcal{G}_1''}{\mathcal{G}_1'}\delta_{1i}\frac{\partial^2\xi_{1i}}{\partial\gamma_{1j}\partial\gamma_{1k}} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\delta_{1i}\frac{\partial^2\xi_{1i}}{\partial\gamma_{1j}\partial\gamma_{1k}} \\
&\quad + \frac{\partial^3\xi_{1i}}{\partial y_i\partial\gamma_{1j}\partial\gamma_{1k}}\delta_{1i}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-1} - \frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1k}}\frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1j}}\delta_{1i}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-2}\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}}\left[\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1} - \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right) + \delta_{1i}\left(\frac{\mathcal{G}_1'''}{\mathcal{G}_1} - \frac{\mathcal{G}_1''^2}{\mathcal{G}_1'^2} - \frac{\mathcal{G}_1''}{\mathcal{G}_1} + \frac{\mathcal{G}_1'^2}{\mathcal{G}_1^2}\right)\right] \\
&\quad + \frac{\partial^2\xi_{1i}}{\partial\gamma_{1j}\partial\gamma_{1k}}\left[\frac{\mathcal{G}_1'}{\mathcal{G}_1} + \delta_{1i}\left(\frac{\mathcal{G}_1''}{\mathcal{G}_1'} - \frac{\mathcal{G}_1'}{\mathcal{G}_1}\right)\right] - \frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1k}}\frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1j}}\left[\delta_{1i}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-2}\right] \\
&\quad + \frac{\partial^3\xi_{1i}}{\partial y_i\partial\gamma_{1j}\partial\gamma_{1k}}\left[\delta_{1i}\left(\frac{\partial\xi_{1i}}{\partial y_i}\right)^{-1}\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{1i}}{\partial\gamma_{1j}}\frac{\partial\xi_{1i}}{\partial\gamma_{1k}}\Phi_1 + \frac{\partial^2\xi_{1i}}{\partial\gamma_{1j}\partial\gamma_{1k}}\Delta_1 - \frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1k}}\frac{\partial^2\xi_{1i}}{\partial y_i\partial\gamma_{1j}}\Psi_1 + \frac{\partial^3\xi_{1i}}{\partial y_i\partial\gamma_{1j}\partial\gamma_{1k}}\Omega_1\Bigg\},
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\gamma})}{\partial\gamma_{2j}\partial\gamma_{2k}} &= \sum_{i=1}^{n}\Bigg\{\frac{\mathcal{G}_2''}{\mathcal{G}_2}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} + \frac{\mathcal{G}_2'''}{\mathcal{G}_2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} \\
&\quad - \frac{\mathcal{G}_2''}{\mathcal{G}_2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\delta_{2i}\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}} + \frac{\mathcal{G}_2'}{\mathcal{G}_2}\frac{\partial^2\xi_{2i}}{\partial\gamma_{2j}\partial\gamma_{2k}} + \frac{\mathcal{G}_2''}{\mathcal{G}_2'}\delta_{2i}\frac{\partial^2\xi_{2i}}{\partial\gamma_{2j}\partial\gamma_{2k}} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\delta_{2i}\frac{\partial^2\xi_{2i}}{\partial\gamma_{2j}\partial\gamma_{2k}} \\
&\quad + \frac{\partial^3\xi_{2i}}{\partial y_i\partial\gamma_{2j}\partial\gamma_{2k}}\delta_{2i}\left(\frac{\partial\xi_{2i}}{\partial y_i}\right)^{-1} - \frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2k}}\frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2j}}\delta_{2i}\left(\frac{\partial\xi_{2i}}{\partial y_i}\right)^{-2}\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}}\left[\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2} - \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right) + \delta_{2i}\left(\frac{\mathcal{G}_2'''}{\mathcal{G}_2} - \frac{\mathcal{G}_2''^2}{\mathcal{G}_2'^2} - \frac{\mathcal{G}_2''}{\mathcal{G}_2} + \frac{\mathcal{G}_2'^2}{\mathcal{G}_2^2}\right)\right] \\
&\quad + \frac{\partial^2\xi_{2i}}{\partial\gamma_{2j}\partial\gamma_{2k}}\left[\frac{\mathcal{G}_2'}{\mathcal{G}_2} + \delta_{2i}\left(\frac{\mathcal{G}_2''}{\mathcal{G}_2'} - \frac{\mathcal{G}_2'}{\mathcal{G}_2}\right)\right] - \frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2k}}\frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2j}}\left[\delta_{2i}\left(\frac{\partial\xi_{2i}}{\partial y_i}\right)^{-2}\right] \\
&\quad + \frac{\partial^3\xi_{2i}}{\partial y_i\partial\gamma_{2j}\partial\gamma_{2k}}\left[\delta_{2i}\left(\frac{\partial\xi_{2i}}{\partial y_i}\right)^{-1}\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\xi_{2i}}{\partial\gamma_{2j}}\frac{\partial\xi_{2i}}{\partial\gamma_{2k}}\Phi_2 + \frac{\partial^2\xi_{2i}}{\partial\gamma_{2j}\partial\gamma_{2k}}\Delta_2 - \frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2k}}\frac{\partial^2\xi_{2i}}{\partial y_i\partial\gamma_{2j}}\Psi_2 + \frac{\partial^3\xi_{2i}}{\partial y_i\partial\gamma_{2j}\partial\gamma_{2k}}\Omega_2\Bigg\}.
\end{aligned}
\tag{16}
$$

The last terms of equations (15) and (16) allow to express $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1}\ell(\boldsymbol{\gamma})$ and $\boldsymbol{\nabla}_{\boldsymbol{\gamma}_2\boldsymbol{\gamma}_2}\ell(\boldsymbol{\gamma})$ as

$$\boldsymbol{\nabla}_{\boldsymbol{\gamma}_\nu\boldsymbol{\gamma}_\nu}\ell(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\left\{\Phi_{\nu i}\frac{\partial\xi_{\nu i}}{\partial\boldsymbol{\gamma}_\nu}\left[\frac{\partial\xi_{\nu i}}{\partial\boldsymbol{\gamma}_\nu}\right]^\top + \Delta_{\nu i}\frac{\partial^2\xi_{\nu i}}{\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top} - \Psi_{\nu i}\frac{\partial^2\xi_{\nu i}}{\partial y_i\partial\boldsymbol{\gamma}_\nu}\left[\frac{\partial^2\xi_{\nu i}}{\partial y_i\partial\boldsymbol{\gamma}_\nu}\right]^\top + \Omega_{\nu i}\frac{\partial^3\xi_{\nu i}}{\partial y_i\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top}\right\},$$

where

$$\frac{\partial^2\xi_{\nu i}}{\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top} = \begin{bmatrix} \dfrac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\gamma_{\nu 11}\partial\gamma_{\nu 11}} & \cdots & \dfrac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\gamma_{\nu 11}\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}} \\ \cdots & \ddots & \cdots \\ \dfrac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}\partial\gamma_{\nu 11}} & \cdots & \dfrac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}} \end{bmatrix},$$

$$\frac{\partial^3\xi_{\nu i}}{\partial y_i\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top} = \begin{bmatrix} \dfrac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\gamma_{\nu 11}\partial\gamma_{\nu 11}} & \cdots & \dfrac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\gamma_{\nu 11}\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}} \\ \cdots & \ddots & \cdots \\ \dfrac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}\partial\gamma_{\nu 11}} & \cdots & \dfrac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}\partial\gamma_{\nu K_\nu J_{\nu K_\nu}}} \end{bmatrix}.$$

In addition, the design sub-matrices of $\dfrac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top}$ and $\dfrac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\boldsymbol{\gamma}_\nu\partial\boldsymbol{\gamma}_\nu^\top}$ can be obtained using the following equations

$$\frac{\partial^2\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial\boldsymbol{\gamma}_{\nu k_\nu}\partial\boldsymbol{\gamma}_{\nu s_\nu}^\top} = \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle}(y_i) & \text{if } \boldsymbol{\gamma}_{\nu k_\nu} = \boldsymbol{\gamma}_{\nu s_\nu} = \boldsymbol{\gamma}_{\nu 0} \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

$$\frac{\partial^3\xi_{\nu i}(\boldsymbol{\gamma}_\nu)}{\partial y_i\partial\boldsymbol{\gamma}_{\nu k_\nu}\boldsymbol{\gamma}_{\nu s_\nu}^\top} = \begin{cases} \boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle'}(y_i) & \text{if } \boldsymbol{\gamma}_{\nu k_\nu} = \boldsymbol{\gamma}_{\nu s_\nu} = \boldsymbol{\gamma}_{\nu 0} \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle}(y_i)$ and $\boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle'}(y_i)$ can be calculated as

$$\boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle}(y_i) = \begin{cases} \dfrac{\partial^2 \xi_{\nu i}}{\partial \gamma_{\nu 0 j_{\nu 0}} \partial \gamma_{\nu 0 k_{\nu 0}}} = \left[\sum_{j_{\nu 0}}^{J_{\nu 0}} \mathcal{Q}_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 0 j_{\nu 0}}\right) & \text{if } j = k \neq 1 \\ \dfrac{\partial^2 \xi_{\nu i}}{\partial \gamma_{\nu 0 j_{\nu 0}} \partial \gamma_{\nu 0 k_{\nu 0}}} = 0 & \text{otherwise,} \end{cases}$$

$$\boldsymbol{\mathcal{Q}}_{\nu 0}^{\triangle\triangle'}(y_i) = \begin{cases} \dfrac{\partial^3 \xi_{\nu i}}{\partial y_i \gamma_{\nu 0 j_{\nu 0}} \partial \gamma_{\nu 0 k_{\nu 0}}} = \left[\sum_{j_{\nu 0}}^{J_{\nu 0}} \mathcal{Q}'_{\nu 0 j_{\nu 0}}(y_i)\right] \exp\left(\gamma_{\nu 0 j_{\nu 0}}\right) & \text{if } j = k \neq 1 \\ \dfrac{\partial^3 \xi_{\nu i}}{\partial y_i \gamma_{\nu 0 j_{\nu 0}} \partial \gamma_{\nu 0 k_{\nu 0}}} = 0 & \text{otherwise.} \end{cases}$$

Supplementary Material C: Estimation Algorithm

The optimization method used is the trust region algorithm. At iteration a, for a given vector $\boldsymbol{\alpha}$ and maintaining $\boldsymbol{\lambda}$ fixed at a vector of values, equation (13) in the main paper (or generally, any of the models' likelihoods considered in the paper) is maximized using

$$\boldsymbol{\alpha}^{[a+1]} = \underset{\boldsymbol{\varepsilon}:\|\boldsymbol{\epsilon}\|\leq\boldsymbol{\Xi}^{[a]}}{\arg\min}\ \bar{\ell}_p(\boldsymbol{\alpha}^{[a]}),$$

where $\bar{\ell}_p(\boldsymbol{\alpha}^{[a]}) = -\big\{\ell_p(\boldsymbol{\alpha}^{[a]}) + \boldsymbol{\varepsilon}^\top \mathbf{g}_p(\boldsymbol{\alpha}^{[a]}) + \frac{1}{2}\boldsymbol{\varepsilon}^\top \boldsymbol{\mathcal{H}}_p(\boldsymbol{\alpha}^{[a]})\boldsymbol{\varepsilon}\big\}$, $\mathbf{g}_p(\boldsymbol{\alpha}^{[a]}) = \mathbf{g}(\boldsymbol{\alpha}^{[a]}) - \boldsymbol{\mathcal{S}}\boldsymbol{\alpha}^{[a]}$, $\boldsymbol{\mathcal{H}}_p(\boldsymbol{\alpha}^{[a]}) = \boldsymbol{\mathcal{H}}(\boldsymbol{\alpha}^{[a]}) - \boldsymbol{\mathcal{S}}$. Vector $\mathbf{g}(\boldsymbol{\alpha}^{[a]})$ consists of $\mathbf{g}_0(\boldsymbol{\alpha}^{[a]}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}_0}\ell(\boldsymbol{\alpha})\big|_{\boldsymbol{\alpha}_0=\boldsymbol{\alpha}_0^{[a]}}$ and $\mathbf{g}_\nu(\boldsymbol{\alpha}^{[a]}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}_\nu}\ell(\boldsymbol{\alpha})\big|_{\boldsymbol{\alpha}_\nu=\boldsymbol{\alpha}_\nu^{[a]}}$, and $\boldsymbol{\mathcal{H}}(\boldsymbol{\alpha}^{[a]})_{l,j} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}_l\boldsymbol{\alpha}_j}\ell(\boldsymbol{\alpha})\big|_{\boldsymbol{\alpha}_l=\boldsymbol{\alpha}_l^{[a]},\boldsymbol{\alpha}_j=\boldsymbol{\alpha}_j^{[a]}}$, where $l, j = 0, 1, 2$ and $\nu = 1, 2$. The euclidean norm is denoted by $\|\cdot\|$, and the radius of the trust region is represented by $\boldsymbol{\Xi}^{[a]}$ which is adjusted through the iterations. Close to the solution, the trust region algorithms behaves as a classic Newton-Raphson unconstrained method (Nocedal & Wright, 2006).

Estimation of $\boldsymbol{\lambda}$ is achieved by adapting the general and automatic multiple smoothing parameter estimation method of (Marra et al., 2017) to the context of the proposed survival models. The smoothing criterion is based on the knowledge of $\mathbf{g}(\boldsymbol{\alpha})$ and $\boldsymbol{\mathcal{H}}(\boldsymbol{\alpha})$. The main ideas and some useful results are given here.

To simplify the notation, $\mathbf{g}_p(\boldsymbol{\alpha}^{[a]})$, $\mathbf{g}(\boldsymbol{\alpha}^{[a]})$, $\boldsymbol{\mathcal{H}}_p(\boldsymbol{\alpha}^{[a]})$ and $\boldsymbol{\mathcal{H}}(\boldsymbol{\alpha}^{[a]})$ are denoted as $\mathbf{g}_p^{[a]}$, $\mathbf{g}^{[a]}$, $\boldsymbol{\mathcal{H}}_p^{[a]}$ and $\boldsymbol{\mathcal{H}}^{[a]}$. First, it is necessary to express the parameter estimator in terms of $\mathbf{g}_p^{[a]}$ and $\boldsymbol{\mathcal{H}}_p^{[a]}$. To achieve this, a first order Taylor expansion of $\mathbf{g}_p^{[a+1]}$ about $\boldsymbol{\alpha}^{[a]}$ is used, which yields the following expression: $\mathbf{0} = \mathbf{g}_p^{[a+1]} \approx \mathbf{g}_p^{[a]}(\boldsymbol{\alpha}^{[a+1]} - \boldsymbol{\alpha}^{[a]})\boldsymbol{\mathcal{H}}_p^{[a]}$. After some manipulations, $\boldsymbol{\alpha}^{[a+1]} = (-\boldsymbol{\mathcal{H}}^{[a]} + \boldsymbol{\mathcal{S}})^{-1}\sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}\big[\sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}\boldsymbol{\alpha}^{[a]} + \sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}^{-1}\mathbf{g}^{[a]}\big]$ is obtained, which then becomes $\boldsymbol{\alpha}^{[a+1]} = (-\boldsymbol{\mathcal{H}}^{[a]} + \boldsymbol{\mathcal{S}})^{-1}\sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}\boldsymbol{\mathcal{Z}}^{[a]}$, where $\boldsymbol{\mathcal{Z}}^{[a]} = \boldsymbol{\upsilon}_{\boldsymbol{\mathcal{Z}}}^{[a]} + \boldsymbol{\xi}_{\boldsymbol{\mathcal{Z}}}^{[a]}$, $\boldsymbol{\upsilon}_{\boldsymbol{\mathcal{Z}}}^{[a]} = \sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}\boldsymbol{\alpha}^{[a]}$ and $\boldsymbol{\xi}_{\boldsymbol{\mathcal{Z}}}^{[a]} = \sqrt{-\boldsymbol{\mathcal{H}}^{[a]}}^{-1}\mathbf{g}^{[a]}$. Eigenvalue decomposition is used to obtain the square root of $-\boldsymbol{\mathcal{H}}^{[a]}$ ant its inverse. Furthermore, from likelihood theory, $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{\mathcal{Z}} \sim \mathcal{N}(\boldsymbol{\upsilon}_{\mathcal{N}}, \mathbf{I})$, where $\boldsymbol{\upsilon}_{\boldsymbol{\mathcal{Z}}} = \sqrt{-\boldsymbol{\mathcal{H}}}\boldsymbol{\alpha}^0$, $\boldsymbol{\alpha}^0$ is the true parameter vector and $\mathbf{I}$ is the identity matrix. $\hat{\boldsymbol{\upsilon}}_{\boldsymbol{\mathcal{Z}}} = \sqrt{-\boldsymbol{\mathcal{H}}}\hat{\boldsymbol{\alpha}} = \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{Z}}$ is the predicted value vector for $\boldsymbol{\mathcal{Z}}$, where $\boldsymbol{\mathcal{B}} = \sqrt{-\boldsymbol{\mathcal{H}}}(-\boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{S}})^{-1}\sqrt{-\boldsymbol{\mathcal{H}}}$. Since our objective is to estimate $\boldsymbol{\lambda}$ so that the smooth terms' complexity which is not supported by the data is removed, the following criterion

is used

$$\mathbb{E}\big(\|\boldsymbol{v}_{\boldsymbol{\mathcal{Z}}} - \hat{\boldsymbol{v}}_{\boldsymbol{\mathcal{Z}}}\|^2\big) = \mathbb{E}\big(\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{Z}}\|^2\big) - \bar{\mathrm{n}} + 2\mathrm{tr}(\boldsymbol{\mathcal{B}}), \tag{17}$$

where $\bar{\mathrm{n}} = 2\,\mathrm{n}$ and $\mathrm{tr}(\boldsymbol{\mathcal{B}})$ represent the number of effective degrees of freedom of the penalized model. In applications, $\boldsymbol{\lambda}$ is estimated by minimizing an estimate of equation (17), in other words

$$\widehat{\|\boldsymbol{v}_{\boldsymbol{\mathcal{Z}}} - \hat{\boldsymbol{v}}_{\boldsymbol{\mathcal{Z}}}\|^2} = \|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{Z}}\|^2 - \bar{\mathrm{n}} + 2\mathrm{tr}(\boldsymbol{\mathcal{B}}). \tag{18}$$

The RHS of equation (18) depends on $\boldsymbol{\lambda}$ through $\boldsymbol{\mathcal{B}}$ while $\boldsymbol{\mathcal{Z}}$ is associated with the un-penalized part of the model. Equation (17) is approximately equivalent to the AIC (Akaike, 1973). This implies that $\boldsymbol{\lambda}$ is estimated by minimizing what is effectively the AIC with number of parameters given by $\mathrm{tr}(\boldsymbol{\mathcal{B}})$. Holding the model's parameter vector value fixed at $\boldsymbol{\alpha}^{[a+1]}$, the following problem

$$\boldsymbol{\lambda}^{[a+1]} = \arg\min_{\boldsymbol{\lambda}} \|\boldsymbol{\mathcal{Z}}^{[a+1]} - \boldsymbol{\mathcal{B}}^{[a+1]}\boldsymbol{\mathcal{Z}}^{[a+1]}\|^2 - \bar{\mathrm{n}} + 2\mathrm{tr}(\boldsymbol{\mathcal{B}}^{[a+1]}) \tag{19}$$

is solved using the automatic efficient and stable computational method proposed by Wood (2004). This approach uses the performance iteration idea of Gu (1992), which is based on Newton's method and can evaluate in an efficient and stable way the components in (19) along with their first and second derivatives with respect to $\log(\boldsymbol{\lambda})$, because the smoothing parameters can only take positive values.

The methods for estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are iterated until the algorithm satisfies the criterion $\big|\ell(\boldsymbol{\alpha}^{[a+1]}) - \ell(\boldsymbol{\alpha}^{[a]})\big| \,/\, \big(0.1 + \big|\ell(\boldsymbol{\alpha}^{[a+1]})\big|\big) \leq (1e - 0.7)$. Starting values are obtained by fitting two non-informative models for the survival and censoring times.

Supplementary Material D: Proofs of the Theorems

This section provides the proofs of Theorems 1, 2 and 3 stated in Section 2.4. First, we establish the main set of assumptions (regularity conditions and vanishing penalties), then the main results are presented.

D.1. Assumptions

Since the same set of assumptions are used to proof Theorems 1 and 2, we use $\boldsymbol{\theta}$ to represents the generic vector of parameters. In particular, $\boldsymbol{\theta} = \boldsymbol{\alpha}$ in Theorem 1 and $\boldsymbol{\theta} = \boldsymbol{\gamma}$ in Theorem 2. Hence, the generic log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \left[[f_1(y_i|\mathbf{z}_i;\boldsymbol{\theta}_1) S_2(y_i|\mathbf{z}_i;\boldsymbol{\theta}_2)]^{\delta_{1i}} [f_2(y_i|\mathbf{z}_i;\boldsymbol{\theta}_2) S_1(y_i|\mathbf{z}_i;\boldsymbol{\theta}_1)]^{\delta_{2i}} \right]. \tag{20}$$

In (20), it has been assumed that $\mathbf{z}_{1i} = \mathbf{z}_{2i}$. In what follows $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta})$, where $\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta}) = \boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta}) = \left[[f_1(y_i|\mathbf{z}_i;\boldsymbol{\theta}_1) S_2(y_i|\mathbf{z}_i;\boldsymbol{\theta}_2)]^{\delta_{1i}} [f_2(y_i|\mathbf{z}_i;\boldsymbol{\theta}_2) S_1(y_i|\mathbf{z}_i;\boldsymbol{\theta}_1)]^{\delta_{2i}} \right]$ and $\mathbf{w}_i = (y_i, \mathbf{z}_i^\top)^\top \in \mathbb{R}_+ \times \mathbb{R}^p$, and $\mathbb{R}_+ = (0, \infty)$. In addition, $\ell(\mathbf{w}_i;\boldsymbol{\theta}) = \log \boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta})$, $\ell_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \ell(\mathbf{w}_i;\boldsymbol{\theta})$, $\nabla_{\boldsymbol{\theta}} \ell(\mathbf{w}_i;\boldsymbol{\theta}) = \dfrac{\partial \ell(\mathbf{w}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}) = \dfrac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell(\mathbf{w}_i;\boldsymbol{\theta}) = \dfrac{\partial^2 \ell(\mathbf{w}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_n(\boldsymbol{\theta}) = \dfrac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$. The penalised likelihood is $\ell_p(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}) - \frac{1}{2} \boldsymbol{\theta}^\top \boldsymbol{\mathcal{S}} \boldsymbol{\theta}$.

Assumption 1 (Regularity Conditions)**.**

(i) The parameter space $\boldsymbol{\Theta}_{\boldsymbol{\theta}}$ is a compact subset of $\mathbb{R}^p$.

(ii) For all $\mathbf{w}_i$, $\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Furthermore, $\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta})$ is measurable in $\mathbf{w}_i$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{\theta}}$.

(iii) Identification condition. $\mathbb{P}[\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta}) \neq \boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta}^*)] > 0$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \in \boldsymbol{\Theta}_{\boldsymbol{\theta}}$.

(iv) Dominance. $\mathbb{E}\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{\theta}}} |\ell(\mathbf{w}_i;\boldsymbol{\theta})|\} < \infty$

(v) The true vector of parameters $\boldsymbol{\theta}^*$ is in the interior of $\boldsymbol{\Theta}_{\boldsymbol{\theta}}$, and $\boldsymbol{\Theta}_0$ is an open neighbourhood around $\boldsymbol{\theta}^*$.

(vi) For all $\mathbf{w}_i$, $\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta})$ is three times continuously differentiable in $\boldsymbol{\theta}$ in an open neighbourhood around $\boldsymbol{\theta}^*$. That is $\boldsymbol{\omega}(\mathbf{w}_i;\boldsymbol{\theta}) \in \mathcal{C}^3(\boldsymbol{\Theta}_0)$

(vii) $\int \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \|\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta})\|\, d\mathbf{w}_i < \infty$ and $\int \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \|\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta})\|\, d\mathbf{w}_i < \infty$.

(viii) For $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, $\mathcal{I}(\boldsymbol{\theta}^*) = \text{Cov}\{\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta})\} = \mathbb{E}\{\{\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta}^*) - \mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta}^*)]\}\{\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta}^*) - \mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w}_i;\boldsymbol{\theta}^*)]\}^\top\}$ exists and is positive-definite.

(ix) For all $1 \leq e, f, h \leq p+1$, there exist a function $\boldsymbol{\phi} : \mathbb{R}_+ \times \mathbb{R}^p \longrightarrow \mathbb{R}$ such that, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ and $\mathbf{w}_i \in \mathbb{R}_+ \times \mathbb{R}^p$, $\left|\dfrac{\partial^3 \ell(\mathbf{w}_i;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_f\partial\boldsymbol{\theta}_h}\right| \leq \boldsymbol{\phi}(\mathbf{w}_i)$, with $\mathbb{E}[\boldsymbol{\phi}(\mathbf{w}_i)] < \infty$.

Assumption 2. $\boldsymbol{\lambda} = o(n^{-1/2})$.

In addition, the following lemmas are required to prove Theorems 1, 2 and 3.

Lemma 1. Let $s(\mathbf{w},\boldsymbol{\theta})$ be a continuously differentiable function, a.s. $d\mathbf{w}$, on $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$.
If $\int \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \left\| \dfrac{\partial s(\mathbf{w},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right\| d\mathbf{w} < \infty$, then for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$,

(i) $\int s(\mathbf{w},\boldsymbol{\theta})d\mathbf{w}$ is continuously differentiable.

(ii) $\int[\partial s(\mathbf{w},\boldsymbol{\theta})/\partial\boldsymbol{\theta}]d\mathbf{w} = \partial[\int s(\mathbf{w},\boldsymbol{\theta})d\mathbf{w}]/\partial\boldsymbol{\theta}$.

Proof. Newey & McFadden (1994, Lemma 3.6). □

Lemma 2. If Assumption **1** hold, then

(i) $\mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)] = \mathbf{0}$

(ii) $\mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)] = \mathcal{I}(\boldsymbol{\theta}^*)$

Proof.

(i) Since $\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})$ is a hypothetical density, its integral is unity:

$$\int \boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy = 1.$$

This is an identity, valid for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{\theta}}$. Differentiating both sides of this identity with respect to $\boldsymbol{\theta}$, we obtain

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int \boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy = \mathbf{0}.$$

Then, by Assumptions **1**(vi) and **1**(vii), and Lemma 1 (the order of differentiation and integration can be interchanged), the following expression is obtained

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int \boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy = \int \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy. \tag{21}$$

By the definition of the score, we have $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})$. Substituting into (21), we obtain

$$\int \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy = \mathbf{0}. \tag{22}$$

This holds for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, in particular, for $\boldsymbol{\theta}^*$. Setting $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, the following equation is obtained

$$\int \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta}^*)dy = \mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)|\mathbf{z}] = \mathbf{0}.$$

Then, applying the Law of Total Expectations, we obtain the required result

$$\mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)] = \mathbb{E}\{\mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)|\mathbf{z}]\} = \mathbf{0}.$$

(ii) Differentiating both sides of identity (22) and by Assumptions **1**(vi) and **1**(vii), and Lemma 1, we obtain

$$\int \frac{\partial}{\partial \boldsymbol{\theta}^\top}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})]dy = \mathbf{0}. \tag{23}$$

The integrand of (23) can be written as $\frac{\partial}{\partial \boldsymbol{\theta}^\top}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})] = \boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta}) + \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})^\top\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})$. Substituting into (23), we obtain

$$-\int \boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy = \int \boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta})^\top\boldsymbol{\omega}(y|\mathbf{z};\boldsymbol{\theta})dy \tag{24}$$

Setting $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, the following equation is obtained

$$\mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)|\mathbf{z}] = \mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)^\top|\mathbf{z}].$$

Then, applying the Law of Total Expectations, we obtain the desired result

$$\begin{aligned}\mathbb{E}\{\mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)|\mathbf{z}]\} &= \mathbb{E}\{\mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)^\top|\mathbf{z}]\}.\\ \mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)] &= \mathbb{E}[\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)^\top].\\ \mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\mathbf{w};\boldsymbol{\theta}^*)] &= \mathcal{I}(\boldsymbol{\theta}^*)\end{aligned}$$

□

Lemma 3. Let $r \in \mathbb{R}_+$, and $\boldsymbol{\Theta}_r$ be the surface of a sphere with radius $rn^{-1/2}$ and center $\boldsymbol{\theta}^*$, that is $\boldsymbol{\Theta}_r = \{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{\theta}} : \boldsymbol{\theta} = \boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{r}, \|\boldsymbol{r}\| = r\}$. For any $\epsilon > 0$, there exist r such that $\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_r} \ell_p(\boldsymbol{\theta}) < \ell_p(\boldsymbol{\theta}^*)\right) \geq 1 - \epsilon$, when n is large enough.

Proof. We define $n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) = n\ell_n(\boldsymbol{\theta}) - n\ell_n(\boldsymbol{\theta}^*) - \frac{n}{2}[\boldsymbol{\theta}^\top\boldsymbol{\mathcal{S}}\boldsymbol{\theta} - \boldsymbol{\theta}^{*\top}\boldsymbol{\mathcal{S}}\boldsymbol{\theta}^*]$. A Third Order Taylor expansion around $\boldsymbol{\theta}^*$ yields

$$\begin{aligned}n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) = {} & n\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*)^\top(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{n}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - n\boldsymbol{\theta}^{*\top}\boldsymbol{\mathcal{S}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\\ & + \frac{n}{6}\sum_e\sum_f\sum_h(\boldsymbol{\theta} - \boldsymbol{\theta}^*)_e(\boldsymbol{\theta} - \boldsymbol{\theta}^*)_f(\boldsymbol{\theta} - \boldsymbol{\theta}^*)_h\frac{\partial^3\ell_n(\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_f\partial_h\boldsymbol{\theta}} - \frac{n}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top\boldsymbol{\mathcal{S}}(\boldsymbol{\theta} - \boldsymbol{\theta}^*).\end{aligned} \tag{25}$$

Let $\boldsymbol{\theta} = \boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{r} \in \boldsymbol{\Theta}_r$. Then (25) becomes in

$$\begin{aligned}n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) = {} & n^{1/2}\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*)^\top\boldsymbol{r} + \frac{1}{2}\boldsymbol{r}^\top\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*)\boldsymbol{r} + \frac{n^{-1/2}}{6}\sum_e\sum_f\sum_h\boldsymbol{r}_e\boldsymbol{r}_f\boldsymbol{r}_h\frac{\partial^3\ell_n(\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_f\partial_h\boldsymbol{\theta}}\\ & - n^{1/2}\boldsymbol{\theta}^{*\top}\boldsymbol{\mathcal{S}}\boldsymbol{r} - \frac{1}{2}\boldsymbol{r}^\top\boldsymbol{\mathcal{S}}\boldsymbol{r}\\ n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) = {} & \sum_{i=1}^{5}\boldsymbol{\mathcal{C}}_{in}(\boldsymbol{r}),\end{aligned}$$

where $\bar{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* + n^{-1/2}\boldsymbol{r}$. For the first term, $|\boldsymbol{\mathcal{C}}_{1n}(\boldsymbol{r})| = \mathcal{O}_p(1)\|\boldsymbol{r}\|$ since by Lemma 2(i), Assumption **1**(vii) and the CLT, $n^{1/2}\boldsymbol{\nabla}_{\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}^*)]$. By Lemma

2(ii) and the LLN, $n^{1/2}\boldsymbol{\nabla}_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^*) \xrightarrow{p} -\mathcal{I}(\boldsymbol{\theta}^*)$, which (by the continuous mapping theorem) yields $\boldsymbol{\mathcal{C}}_{2n}(\boldsymbol{r}) \xrightarrow{p} -\frac{1}{2}\boldsymbol{r}^\top\mathcal{I}(\boldsymbol{\theta}^*)\boldsymbol{r}$. Thus, by Assumption **1**(viii), $\boldsymbol{\mathcal{C}}_{2n}(\boldsymbol{r}) \leq -\frac{1}{2}\zeta_{\min}\|\boldsymbol{r}\|^2$, where $\zeta_{\min} > 0$ is the smallest eigenvalue of $\mathcal{I}(\boldsymbol{\theta}^*)$. By Assumption **1**(ix) and the LLN, $\left|\dfrac{\partial^3\ell_n(\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_f\partial\boldsymbol{\theta}_h}\right| \leq \frac{1}{n}\sum_1^n \boldsymbol{\phi}(\mathbf{w}_i) \xrightarrow{p} \mathbb{E}[\boldsymbol{\phi}(\mathbf{w}_i)] < \infty$. This fact and the Cauchy-Schwarz inequality imply that $|\boldsymbol{\mathcal{C}}_{3n}(\boldsymbol{r})| \xrightarrow{p} 0$. Finally, by Assumption **2** we have that $|\boldsymbol{\mathcal{C}}_{4n}(\boldsymbol{r})| \xrightarrow{p} 0$ and $|\boldsymbol{\mathcal{C}}_{5n}(\boldsymbol{r})| \xrightarrow{p} 0$. Therefore, combining all of these results, we have

$$n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) \leq \mathcal{O}_p(1)\|\boldsymbol{r}\| - \frac{1}{2}\zeta_{\min}\|\boldsymbol{r}\|^2 \tag{26}$$

for large enough n. Since the choice of $\boldsymbol{\theta}$ was arbitrary, (26) becomes in

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_r} n\ell_p(\boldsymbol{\theta}) - n\ell_p(\boldsymbol{\theta}^*) \leq \boldsymbol{\mathcal{C}},$$

where $\boldsymbol{\mathcal{C}} = \mathcal{O}_p(1)\|\boldsymbol{r}\| - \frac{1}{2}\zeta_{\min}\|\boldsymbol{r}\|^2$. This implies that $\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_r} \ell_p(\boldsymbol{\theta}) < \ell_p(\boldsymbol{\theta}^*)\right) \geq \mathbb{P}(\boldsymbol{\mathcal{C}} < 0)$. Therefore, because for all $\epsilon > 0$, there exists a $\|\boldsymbol{r}\| \in \mathbb{R}_+$ such that $\mathbb{P}[\boldsymbol{\mathcal{C}} < 0] \geq 1 - \epsilon$, we obtain $\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_r} \ell_p(\boldsymbol{\theta}) < \ell_p(\boldsymbol{\theta}^*)\right) \geq 1 - \epsilon$, as required. □

Lemma 4. (Delta Method). Suppose that $\boldsymbol{\theta}_n$ is a sequence of k-dimensional random vectors and $\boldsymbol{\theta}^*$ be a constant k-vector such that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ for some $k \times k$ matrix $\boldsymbol{\Omega}$. Let $g : \mathbb{R}^k \to \mathbb{R}^l$ be continuously differentiable at $\boldsymbol{\theta}^*$. Then

$$\sqrt{n}(g(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, G\boldsymbol{\Omega}G^\top)$$

where $G = \left.\dfrac{\partial g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ is the $l \times k$ Jacobian matrix.

Proof. Hayashi (2000, Lemma 2.5). □

D.2. Theorems

Theorem 1 (Asymptotic properties of the IPMLE estimator).

Proof. Under Assumptions **1**(i), **1**(ii) and Gourieroux & Monfort (1995, Property 24.1), there exists a well defined random variable (measurable function) $\hat{\boldsymbol{\alpha}}$ that solves the optimization problem

in equation (13). Due to Lemma 3, the informative penalized log-likelihood function has a local maximum $\hat{\boldsymbol{\alpha}}$ in the interior of a sphere centered on $\boldsymbol{\alpha}^*$. Then, $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| = \mathcal{O}_p(n^{-1/2})$, implying that $\hat{\boldsymbol{\alpha}}$ is a $\sqrt{n}$-consistent estimator. Furthermore, by Assumption **1**(iii) and Newey & McFadden (1994, Lemma 2.2), $\boldsymbol{\alpha}^*$ is the unique maximizer of $Q^*(\boldsymbol{\alpha}) = \mathbb{E}[\ell(\mathbf{w}_i; \boldsymbol{\alpha})]$.

(i) To prove the asymptotic normality of the informative penalized likelihood estimator, we take the derivative of the log-likelihood function in equation (13) to obtain

$$\mathbf{0} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \ell_n(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\mathcal{S}}\hat{\boldsymbol{\alpha}}. \tag{27}$$

Applying a second order Taylor expansion in equation (27) yields

$$\mathbf{0} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \ell_n(\boldsymbol{\alpha}^*) - \boldsymbol{\mathcal{S}}\boldsymbol{\alpha}^* + \boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell_n(\boldsymbol{\alpha}^*)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) - \boldsymbol{\mathcal{S}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) + \boldsymbol{\Delta}, \tag{28}$$

where the last term is defined as

$$\boldsymbol{\Delta} = \begin{bmatrix} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)^\top [\boldsymbol{\nabla}^2 \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \ell_n(\bar{\boldsymbol{\alpha}})]_1 (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \\ \vdots \\ (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)^\top [\boldsymbol{\nabla}^2 \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \ell_n(\bar{\boldsymbol{\alpha}})]_p (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \end{bmatrix}, \tag{29}$$

and $\bar{\boldsymbol{\alpha}}$ lies between $\boldsymbol{\alpha}^*$ and $\hat{\boldsymbol{\alpha}}$, therefore $\|\bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| \leq \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|$. We can rewrite equation (28) to obtain

$$\mathbf{0} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \ell_n(\boldsymbol{\alpha}^*) - \boldsymbol{\mathcal{S}}\boldsymbol{\alpha}^* + \boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell_n(\boldsymbol{\alpha}^*)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) - \boldsymbol{\mathcal{S}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) + \boldsymbol{\Delta}_p(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*), \tag{30}$$

where $\boldsymbol{\Delta}_p$ is defined as

$$\boldsymbol{\Delta}_p = \begin{bmatrix} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)^\top [\boldsymbol{\nabla} \boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell_n(\bar{\boldsymbol{\alpha}})]_1 \\ \vdots \\ (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)^\top [\boldsymbol{\nabla} \boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \ell_n(\bar{\boldsymbol{\alpha}})]_p \end{bmatrix}.$$

Multiplying the right hand side of equation (30) by $\sqrt{n}$, leads

$$[\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\ell_n(\boldsymbol{\alpha}^*) - \boldsymbol{\mathcal{S}} + \boldsymbol{\Delta}_p]\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) = \sqrt{n}[\boldsymbol{\mathcal{S}}\boldsymbol{\alpha}^* - \boldsymbol{\nabla}_{\boldsymbol{\alpha}}\ell_n(\boldsymbol{\alpha}^*)] \tag{31}$$

By assumption **2**, $\boldsymbol{\mathcal{S}} \xrightarrow{p} 0$ and $\boldsymbol{\mathcal{S}}\boldsymbol{\alpha}^* \xrightarrow{p} 0$. Furthermore, by assumption **1**(ix), $\boldsymbol{\Delta}_p \xrightarrow{p} 0$. As earlier mentioned, by Lemma 2(i), Assumption **1**(vii) and the CLT, $n^{1/2}\boldsymbol{\nabla}_{\boldsymbol{\alpha}}\ell_n(\boldsymbol{\alpha}^*) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathcal{I}(\boldsymbol{\alpha}^*)]$, and by Lemma 2(ii) and the LLN, $n^{1/2}\boldsymbol{\nabla}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\ell_n(\boldsymbol{\alpha}^*) \xrightarrow{p} -\mathcal{I}(\boldsymbol{\alpha}^*)$. Finally, by Slutsky's theorem, we obtain

$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \xrightarrow{d} \mathcal{N}\left\{\mathbf{0}, [\mathcal{I}(\boldsymbol{\alpha}^*)]^{-1}\right\},$$

as required.

(ii) Under Theorem 1, $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \xrightarrow{d} \mathcal{N}\{\mathbf{0}, [\mathcal{I}(\boldsymbol{\alpha}^*)]^{-1}\}$. In particular, for $\hat{\boldsymbol{\alpha}_{\nu 0}} \in \hat{\boldsymbol{\alpha}}$ we have $\sqrt{n}(\hat{\boldsymbol{\alpha}_{\nu 0}} - \boldsymbol{\alpha}^*_{\nu 0}) \xrightarrow{d} \mathcal{N}\{\mathbf{0}, [\mathcal{I}(\boldsymbol{\alpha}^*_{\nu 0})]^{-1}\}$. In addition, $S : \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable at $\boldsymbol{\alpha}^*_{\nu 0}$, with gradient defined as $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_{\nu 0}} S(\boldsymbol{\alpha}^*_{\nu 0}) = \mathcal{G}'_{\nu 0}[s(\boldsymbol{\alpha}^*_{\nu 0})]\boldsymbol{\nabla}_{\boldsymbol{\alpha}_{\nu 0}} s(\boldsymbol{\alpha}^*_{\nu 0})$. Then, we can applied Lemma 4 to obtain

$$\sqrt{n}[\hat{S}_{\nu 0}(\hat{\boldsymbol{\alpha}}_{\nu 0}) - S_{\nu 0}(\boldsymbol{\alpha}^*_{\nu 0})] \xrightarrow{d} \mathcal{N}\big\{\mathbf{0}, \mathcal{G}'_{\nu 0}[s(\boldsymbol{\alpha}^*_{\nu 0})]\boldsymbol{\nabla}_{\boldsymbol{\alpha}_{\nu 0}} s(\boldsymbol{\alpha}^*_{\nu 0})[\mathcal{I}(\boldsymbol{\alpha}^*_{\nu 0})]^{-1}\boldsymbol{\nabla}_{\boldsymbol{\alpha}_{\nu 0}} s(\boldsymbol{\alpha}^*_{\nu 0})^\top \mathcal{G}'_{\nu 0}[s(\boldsymbol{\alpha}^*_{\nu 0})]\big\}.$$

Furthermore, we know that $\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha}) = \mathbf{0}$, therefore $\mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2}\ell(\boldsymbol{\alpha}_0)] = \mathbf{0}$. This also implies that $\mathbb{E}[-\boldsymbol{\nabla}_{\boldsymbol{\alpha}_{10}\boldsymbol{\alpha}_{20}}\ell(\boldsymbol{\alpha}_0)] = \mathbf{0}$, which means that $\boldsymbol{\alpha}_{10}$ and $\boldsymbol{\alpha}_{20}$ are independent. Then, $S(\boldsymbol{\alpha}_{10})$ and $S(\boldsymbol{\alpha}_{20})$ are also independent, as required.

□

Theorem 2 (Asymptotic properties of the NPMLE estimator).

Proof. This proof follows similar arguments of Theorem 1. □

Theorem 3 (Efficiency of the IPMLE estimator).

Proof. For $\nu = 1, 2$, we define $\boldsymbol{\gamma}_\nu = (\boldsymbol{\gamma}^\iota_\nu, \boldsymbol{\gamma}^{n\iota}_\nu)^\top$ so that $\boldsymbol{\mathcal{Q}}_i^\top\boldsymbol{\gamma}_\nu = \boldsymbol{\mathcal{Q}}_i^{0\top}\boldsymbol{\gamma}^\iota_\nu + \boldsymbol{\mathcal{Q}}_{\nu i}^{1\top}\boldsymbol{\gamma}^{n\iota}_\nu$. Where $\boldsymbol{\gamma}^\iota_\nu = (\boldsymbol{\gamma}^{\iota\top}_{\nu 1}, ..., \boldsymbol{\gamma}^{\iota\top}_{\nu Q})^\top$ and $\boldsymbol{\gamma}^{n\iota}_\nu = (\boldsymbol{\gamma}^{n\iota\top}_{\nu(Q+1)}, ..., \boldsymbol{\gamma}^{n\iota\top}_{\nu Q_\nu})^\top$ are the informative and non-informative

parameters of the non-informative model respectively. Thus, under Assumption **1**(viii) and Lemma 2(ii), $\mathcal{I}(\boldsymbol{\gamma}^*)$ can be written as

$$\mathcal{I}(\boldsymbol{\gamma}^*) = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\gamma}_1^\iota} & \mathcal{I}_{\boldsymbol{\gamma}_1^\iota \boldsymbol{\gamma}_1^{n\iota}} & \mathbf{0} & \mathbf{0} \\ \mathcal{I}_{\boldsymbol{\gamma}_1^{n\iota} \boldsymbol{\gamma}_1^\iota} & \mathcal{I}_{\boldsymbol{\gamma}_1^{n\iota}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{\boldsymbol{\gamma}_2^\iota} & \mathcal{I}_{\boldsymbol{\gamma}_2^\iota \boldsymbol{\gamma}_2^{n\iota}} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{\boldsymbol{\gamma}_2^{n\iota} \boldsymbol{\gamma}_2^\iota} & \mathcal{I}_{\boldsymbol{\gamma}_2^{n\iota}} \end{bmatrix}, \tag{32}$$

where $\mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota} = \mathcal{I}(\boldsymbol{\gamma}_\nu^{*\iota})$, $\mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}} = \mathcal{I}(\boldsymbol{\gamma}_\nu^{*n\iota})$ and $\mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota \boldsymbol{\gamma}_\nu^{n\iota}} = \mathcal{I}(\boldsymbol{\gamma}_\nu^{*n\iota}, \boldsymbol{\gamma}_\nu^{*\iota})$. Taking the inverse of (32), we obtain

$$[\mathcal{I}(\boldsymbol{\gamma}^*)]^{-1} = \begin{bmatrix} \Sigma_{\boldsymbol{\gamma}_1^{*\iota}} & \Sigma_{\boldsymbol{\gamma}_1^{*\iota} \boldsymbol{\gamma}_1^{*n\iota}} & \mathbf{0} & \mathbf{0} \\ \Sigma_{\boldsymbol{\gamma}_1^{*n\iota} \boldsymbol{\gamma}_1^{*\iota}} & \Sigma_{\boldsymbol{\gamma}_1^{*n\iota}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{\boldsymbol{\gamma}_2^{*\iota}} & \Sigma_{\boldsymbol{\gamma}_2^{*\iota} \boldsymbol{\gamma}_2^{*n\iota}} \\ \mathbf{0} & \mathbf{0} & \Sigma_{\boldsymbol{\gamma}_2^{*\iota} \boldsymbol{\gamma}_2^{*n\iota}} & \Sigma_{\boldsymbol{\gamma}_2^{*n\iota}} \end{bmatrix}, \tag{33}$$

where $\Sigma_{\boldsymbol{\gamma}_\nu^{*\iota}} = [\mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota} - \mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota \boldsymbol{\gamma}_\nu^{n\iota}} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota} \boldsymbol{\gamma}_\nu^\iota}]^{-1}$, $\Sigma_{\boldsymbol{\gamma}_\nu^{*\iota} \boldsymbol{\gamma}_\nu^{*n\iota}} = -\Sigma_{\boldsymbol{\gamma}_\nu^{*\iota}} \mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota \boldsymbol{\gamma}_\nu^{n\iota}} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1}$, $\Sigma_{\boldsymbol{\gamma}_\nu^{*n\iota} \boldsymbol{\gamma}_\nu^{*\iota}} = -\mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota} \boldsymbol{\gamma}_\nu^\iota} \Sigma_{\boldsymbol{\gamma}_\nu^{*\iota}}$ and $\Sigma_{\boldsymbol{\gamma}_\nu^{*n\iota}} = \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1} + \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota} \boldsymbol{\gamma}_\nu^\iota} \Sigma_{\boldsymbol{\gamma}_\nu^{*\iota}} \mathcal{I}_{\boldsymbol{\gamma}_\nu^\iota \boldsymbol{\gamma}_\nu^{n\iota}} \mathcal{I}_{\boldsymbol{\gamma}_\nu^{n\iota}}^{-1}$.

On the other hand, also by Assumption **1**(viii) and Lemma 2(ii), $\mathcal{I}(\boldsymbol{\alpha}^*)$ can be written as

$$\mathcal{I}(\boldsymbol{\alpha}^*) = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\alpha}_0} & \mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_1} & \mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_2} \\ \mathcal{I}_{\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_0} & \mathcal{I}_{\boldsymbol{\alpha}_1} & \mathbf{0} \\ \mathcal{I}_{\boldsymbol{\alpha}_2 \boldsymbol{\alpha}_0} & \mathbf{0} & \mathcal{I}_{\boldsymbol{\alpha}_2} \end{bmatrix}, \tag{34}$$

where $\mathcal{I}_{\boldsymbol{\alpha}_0} = \mathcal{I}(\boldsymbol{\alpha}_0^*)$, $\mathcal{I}_{\boldsymbol{\alpha}_\nu} = \mathcal{I}(\boldsymbol{\alpha}_\nu^*)$, $\mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_\nu} = \mathcal{I}(\boldsymbol{\alpha}_0^*, \boldsymbol{\alpha}_\nu^*)$ and $\mathcal{I}_{\boldsymbol{\alpha}_\nu \boldsymbol{\alpha}_0} = \mathcal{I}(\boldsymbol{\alpha}_\nu^*, \boldsymbol{\alpha}_0^*)$. Taking the inverse of (34), yields

$$[\mathcal{I}(\boldsymbol{\alpha}^*)]^{-1} = \begin{bmatrix} \Sigma_{\boldsymbol{\alpha}_0^*} & \Sigma_{\boldsymbol{\alpha}_0^* \boldsymbol{\alpha}_1^*} & \Sigma_{\boldsymbol{\alpha}_0^* \boldsymbol{\alpha}_2^*} \\ \Sigma_{\boldsymbol{\alpha}_1^* \boldsymbol{\alpha}_0^*} & \Sigma_{\boldsymbol{\alpha}_1^*} & \mathbf{0} \\ \Sigma_{\boldsymbol{\alpha}_2^* \boldsymbol{\alpha}_0^*} & \mathbf{0} & \Sigma_{\boldsymbol{\alpha}_2^*} \end{bmatrix}, \tag{35}$$

where $\Sigma_{\boldsymbol{\alpha}_0^*} = [\mathcal{I}_{\boldsymbol{\alpha}_0} - \mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_1} \mathcal{I}_{\boldsymbol{\alpha}_1}^{-1} \mathcal{I}_{\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_0} - \mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_2} \mathcal{I}_{\boldsymbol{\alpha}_2}^{-1} \mathcal{I}_{\boldsymbol{\alpha}_2 \boldsymbol{\alpha}_0}]^{-1}$, $\Sigma_{\boldsymbol{\alpha}_0^* \boldsymbol{\alpha}_\nu^*} = -\Sigma_{\boldsymbol{\alpha}_0^*} \mathcal{I}_{\boldsymbol{\alpha}_0 \boldsymbol{\alpha}_\nu} \mathcal{I}_{\boldsymbol{\alpha}_\nu}^{-1}$, $\Sigma_{\boldsymbol{\alpha}_\nu^* \boldsymbol{\alpha}_0^*} =$

$-\mathcal{I}^{-1}_{\boldsymbol{\alpha}_\nu}\mathcal{I}_{\boldsymbol{\alpha}_\nu\boldsymbol{\alpha}_0}\Sigma_{\boldsymbol{\alpha}_0^*}$ and $\Sigma_{\boldsymbol{\alpha}_\nu^*} = \mathcal{I}^{-1}_{\boldsymbol{\alpha}_\nu} + \mathcal{I}^{-1}_{\boldsymbol{\alpha}_\nu}\mathcal{I}_{\boldsymbol{\alpha}_\nu\boldsymbol{\alpha}_0}\Sigma_{\boldsymbol{\alpha}_0^*}\mathcal{I}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_\nu}\mathcal{I}^{-1}_{\boldsymbol{\alpha}_\nu}$.

Thus, by (14), (15), (16), (17), (18) and using that $\boldsymbol{\gamma}^{n\iota}_{\nu 0} = \boldsymbol{\alpha}_{\nu 0}$, we obtain $\mathcal{I}_{\boldsymbol{\alpha}_0} = \mathcal{I}_{\boldsymbol{\gamma}^\iota_1} + \mathcal{I}_{\boldsymbol{\gamma}^\iota_2}$, $\mathcal{I}_{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_\nu} = \mathcal{I}_{\boldsymbol{\gamma}^\iota_\nu\boldsymbol{\gamma}^{n\iota}_\nu}$, $\mathcal{I}_{\boldsymbol{\alpha}_\nu\boldsymbol{\alpha}_0} = \mathcal{I}_{\boldsymbol{\gamma}^{n\iota}_\nu\boldsymbol{\gamma}^\iota_\nu}$ and $\mathcal{I}_{\boldsymbol{\alpha}_\nu} = \mathcal{I}_{\boldsymbol{\gamma}^{n\iota}_\nu}$. This and the fact that $\Sigma^{-1}_{\boldsymbol{\alpha}_0^*}$ and $\Sigma^{-1}_{\boldsymbol{\gamma}^{*\iota}_\nu}$ are positive definite matrices, imply that $[\Sigma_{\boldsymbol{\gamma}^{*\iota}_\nu} - \Sigma_{\boldsymbol{\alpha}_0^*}]$ is positive definite. Therefore, $\Sigma_{\boldsymbol{\alpha}_0^*} < \Sigma_{\boldsymbol{\gamma}^{*\iota}_\nu}$. Using this reasoning, we conclude that $\Sigma_{\boldsymbol{\alpha}_0^*\boldsymbol{\alpha}_\nu^*} < \Sigma_{\boldsymbol{\gamma}^{*\iota}_\nu\boldsymbol{\gamma}^{*n\iota}_\nu}$, $\Sigma_{\boldsymbol{\alpha}_\nu^*\boldsymbol{\alpha}_0^*} < \Sigma_{\boldsymbol{\gamma}^{*n\iota}_\nu\boldsymbol{\gamma}^{*\iota}_\nu}$ and $\Sigma_{\boldsymbol{\alpha}_\nu^*} < \Sigma_{\boldsymbol{\gamma}^{*n\iota}_\nu}$, as required. □

The proof of Lemma 3 in the context of informative and non-informative censoring models was adapted from Xingwei et al. (2010) and Vatter & Chavez-Demoulin (2015). The proofs of the asymptotic normality (part (i) of Theorems 1 and 2) are based on Vatter & Chavez-Demoulin (2015).

Supplementary Material E: Confidence Intervals

At convergence, point-wise intervals for linear and non-linear functions for both the non-informative and informative models' parameters can be obtained using the following Bayesian large sample approximation

$$\boldsymbol{\theta} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}), \tag{36}$$

where $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = [\boldsymbol{\mathcal{H}}_p(\hat{\boldsymbol{\theta}})]^{-1}$. For generalised additive models, intervals derived using equation (36) have good frequentist properties, since they account for both smoothing bias and sampling variability (Marra & Wood, 2012). For the non-informative and informative models, equation (36) can be verified using the distribution of $\boldsymbol{\mathcal{Z}}$ (described in Supplementary Material C), making the large sample assumption that $\boldsymbol{\mathcal{H}}(\boldsymbol{\theta})$ can be treated as fixed, and making the usual prior Bayesian assumption for smooth models $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}}^{-1})$, where $\boldsymbol{\mathcal{S}}^{-1}$ is the Moore-Penrose pseudoinverse of $\boldsymbol{\mathcal{S}}$ (Silverman, 1985; Wood, 2017). In equation (36), smoothing parameter uncertainty is neglected. Nevertheless, according to Marra & Wood (2012) this is not problematic if heavy over-smoothing is avoided so that the smoothing bias is not a large proportion of the sampling variability. See also Marra et al. (2017) for an application of this approach to a more general smoothing spline context.

Following Pya & Wood (2015), confidence interval estimates for the monotonic smooth terms in the models can be obtained using the distribution of $\tilde{\boldsymbol{\beta}}_{\nu 0}$ (defined in Section 2.3 of the main paper) since all smooth components would then depend linearly on $\tilde{\boldsymbol{\beta}}_{\nu 0}$. Such distribution is

$$\tilde{\boldsymbol{\beta}}_{\nu 0} \sim \mathcal{N}(\hat{\tilde{\boldsymbol{\beta}}}_{\nu 0}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_{\nu 0}}),$$

where $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_{\nu 0}} = \operatorname{diag}(\boldsymbol{\Gamma}_{\nu 0})\, [\boldsymbol{\mathcal{H}}_p(\hat{\boldsymbol{\beta}}_{\nu 0})]^{-1} \operatorname{diag}(\boldsymbol{\Gamma}_{\nu 0})$. The derivation of this result can be found in Pya & Wood (2015).

P-values for the smooth components in the non-informative and informative models are obtained by adapting the results discussed in Wood (2013) to the present context, where $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}_{\nu 0}}$ is used for the calculations. The reader is referred to the above citation for the definition of reference degrees of freedom.

Supplementary Material F: Model Selection

In practical situations, it is important to detect if $\sum_{k_1=1}^{K_1} s_{1k_1}(\mathbf{x}_{1k_1i})$ and $\sum_{k_2=1}^{K_2} s_{2k_2}(\mathbf{x}_{2k_2i})$ have components in common. This is basically a model selection problem and, to this end, we propose using the AIC, BIC and K-Fold Cross validation criterion $\left(\Upsilon^{\mathrm{KCV}}\right)$. The AIC and BIC can be defined as

$$\mathrm{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2\,\mathrm{EDF},$$
$$\mathrm{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + \log(n)\,\mathrm{EDF},$$

where the log-likelihood is evaluated at the penalized parameter estimates and $\mathrm{EDF} = \mathrm{tr}(\hat{\boldsymbol{\mathcal{B}}})$ with $\hat{\boldsymbol{\mathcal{B}}}$ defined in Supplementary Material C.

As for Υ^{KCV} (Stone, 1974), we first randomly divide the set of observations in K groups (folds) of approximately equal size. Each fold is then in turn treated as a validation set, and the IPMLE for a given model is used to estimate the vector of parameters $\boldsymbol{\alpha}$ using the remaining $\mathrm{K}-1$ folds. The so obtained estimates are denoted as $\hat{\boldsymbol{\alpha}}_0^{\setminus k}$ and $\hat{\boldsymbol{\alpha}}_\nu^{\setminus k}$, and the log-likelihood function is calculated as

$$\ell_k(\hat{\boldsymbol{\alpha}}^{\setminus k}) = \left\{ \log \mathcal{G}_1\left[\xi_{1i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_1^{\setminus k})\right] + \delta_{1i} \log\left\{ -\frac{\mathcal{G}_1'\left[\xi_{1i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_1^{\setminus k})\right]}{\mathcal{G}_1\left[\xi_{1i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_1^{\setminus k})\right]} \frac{\partial \xi_{1i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_1^{\setminus k})}{\partial y_i} \right\} \right\}$$
$$+ \left\{ \log \mathcal{G}_2\left[\xi_{2i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_2^{\setminus k})\right] + \delta_{2i} \log\left\{ -\frac{\mathcal{G}_2'\left[\xi_{2i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_2^{\setminus k})\right]}{\mathcal{G}_2\left[\xi_{2i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_2^{\setminus k})\right]} \frac{\partial \xi_{2i}(\hat{\boldsymbol{\alpha}}_0^{\setminus k}, \hat{\boldsymbol{\alpha}}_2^{\setminus k})}{\partial y_i} \right\} \right\},$$

and Υ^{KCV} given by

$$\Upsilon^{\mathrm{KCV}} = \sum_{k=1}^{K} \ell_k(\hat{\boldsymbol{\alpha}}^{\setminus k}). \tag{37}$$

We choose the model which maximizes (37). The same procedure is used when Υ^{KCV} is calculated for the non-informative model. In such a case we have

$$\ell_k(\hat{\boldsymbol{\gamma}}^{\setminus k}) = \left\{ \log \mathcal{G}_1\left[\xi_{1i}(\hat{\boldsymbol{\gamma}}_1^{\setminus k})\right] + \delta_{1i} \log\left\{ -\frac{\mathcal{G}_1'\left[\xi_{1i}(\hat{\boldsymbol{\gamma}}_1^{\setminus k})\right]}{\mathcal{G}_1\left[\xi_{1i}(\hat{\boldsymbol{\gamma}}_1^{\setminus k})\right]} \frac{\partial \xi_{1i}(\hat{\boldsymbol{\gamma}}_1^{\setminus k})}{\partial y_i} \right\} \right\}$$
$$+ \left\{ \log \mathcal{G}_2\left[\xi_{2i}(\hat{\boldsymbol{\gamma}}_2^{\setminus k})\right] + \delta_{2i} \log\left\{ -\frac{\mathcal{G}_2'\left[\xi_{2i}(\hat{\boldsymbol{\gamma}}_2^{\setminus k})\right]}{\mathcal{G}_2\left[\xi_{2i}(\hat{\boldsymbol{\gamma}}_2^{\setminus k})\right]} \frac{\partial \xi_{2i}(\hat{\boldsymbol{\gamma}}_2^{\setminus k})}{\partial y_i} \right\} \right\},$$

and therefore $\Upsilon^{\text{KCV}} = \sum_{k=1}^{K} \ell_k(\hat{\boldsymbol{\gamma}}^{\setminus k})$.

Model	Non-Inf. Covariates	Inf. Covariates	Link T_{1i}	Link T_{2i}	AIC	Υ^{KCV}	BIC
1	s(wmonth) s(mthage) region alcohol nsibs	$\cdots$	PH	PH	13775.68	-6924.20	14015.53
2	s(wmonth) s(mthage) region alcohol nsibs	$\cdots$	PO	PH	13776.87	-8396.57	14016.51
3	s(wmonth) s(mthage) nsibs	alcohol region	PH	PH	13772.60	-6922.63	13981.42
4	s(wmonth) s(mthage) nsibs	alcohol region	PO	PH	13773.80	-8392.31	13982.51

Table 2: Values of three model selection criteria (AIC, BIC and Υ^{KCV}) for the best informative and non-informative models fitted to the real data application of this paper. The models were fitted using `gamlss()` in `GJRM` by employing different combinations of covariates and link functions.

Supplementary Material G: Additional simulation results for DGP1 and DGP2 and findings from a simulation study with mild censoring rate

In the DGP presented in this section (DGP3), z_{1i} is informative, z_{2i} is informative and a mild censoring rate (about 47%) is considered. T_{1i} and T_{2i} were generated using the model defined in equation (19) of the main paper. The baseline survival functions were defined as $S_{10}(t_{1i}) = 0.8\exp\left(-0.4t_{1i}^{2.5}\right) + 0.2\exp\left(-0.1t_{1i}^{1.0}\right)$ and $S_{20}(t_{2i}) = 0.99\exp\left(-0.05t_{2i}^{2.3}\right) + 0.01\exp\left(-0.4t_{2i}^{1.1}\right)$. The informative covariates, z_{1i} and z_{2i}, were generated using a binomial and a uniform distribution respectively. Also, $s_{11}(z_{2i}) = s_{12}(z_{2i}) = \sin(2\pi z_i)$, $\alpha_{01} = -0.10$, $\alpha_{02} = -0.25$ and $\alpha_{11} = \alpha_{12} = -1.5$.

The main findings are:

- Figure 1 and Table 4 show that overall the mean estimates for the two estimators are very close to the respective true values and improve as the sample size increases. However, even though the variability of the estimates (IPMLE and NPMLE) decreases as the sample size grows large, the IPMLE is slightly more efficient than the NPMLE in recovering the true linear effects for all sample sizes examined here. In particular, the RMSE of the IPMLE is slightly smaller than the RMSE of the NPMLE for all sample sizes considered.

- Figures 2 and 3, and Table 4 show that overall the true functions are recovered well by the IPMLE and NPMLE and that the results improve in terms of bias and efficiency as the sample size increases. Furthermore, the IPMLE is slightly more efficient than the NPMLE in recovering the non-linear covariate effects for all sample sizes examined in this section (Table 4). However, this gain in efficiency by the IPMLE is not too significant when a mild censoring rate (47%) is examined.

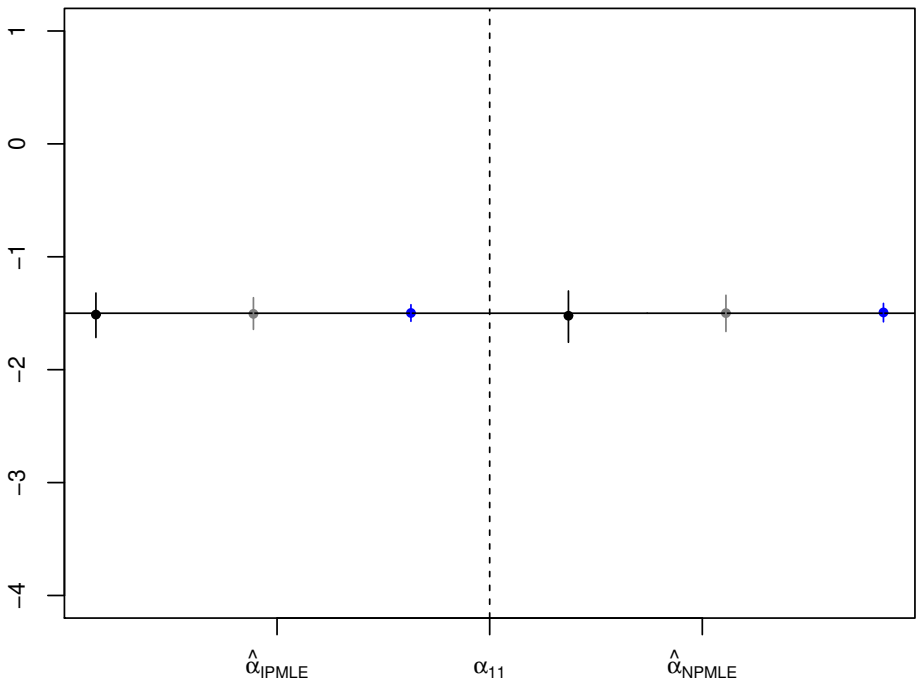

Figure 1: Linear coefficient estimates obtained by applying `gamlss()` to informative survival data simulated according to DGP3 characterised by a censoring rate of about 47%. Circles indicate mean estimates while bars represent the estimates' ranges resulting from 5% and 95% quantiles. True values are indicated by black solid horizontal lines. Black circles and vertical bars refer to the results obtained for $n = 500$, whereas those for $n = 1000$ and $n = 4000$ are given in dark gray and blue, respectively.

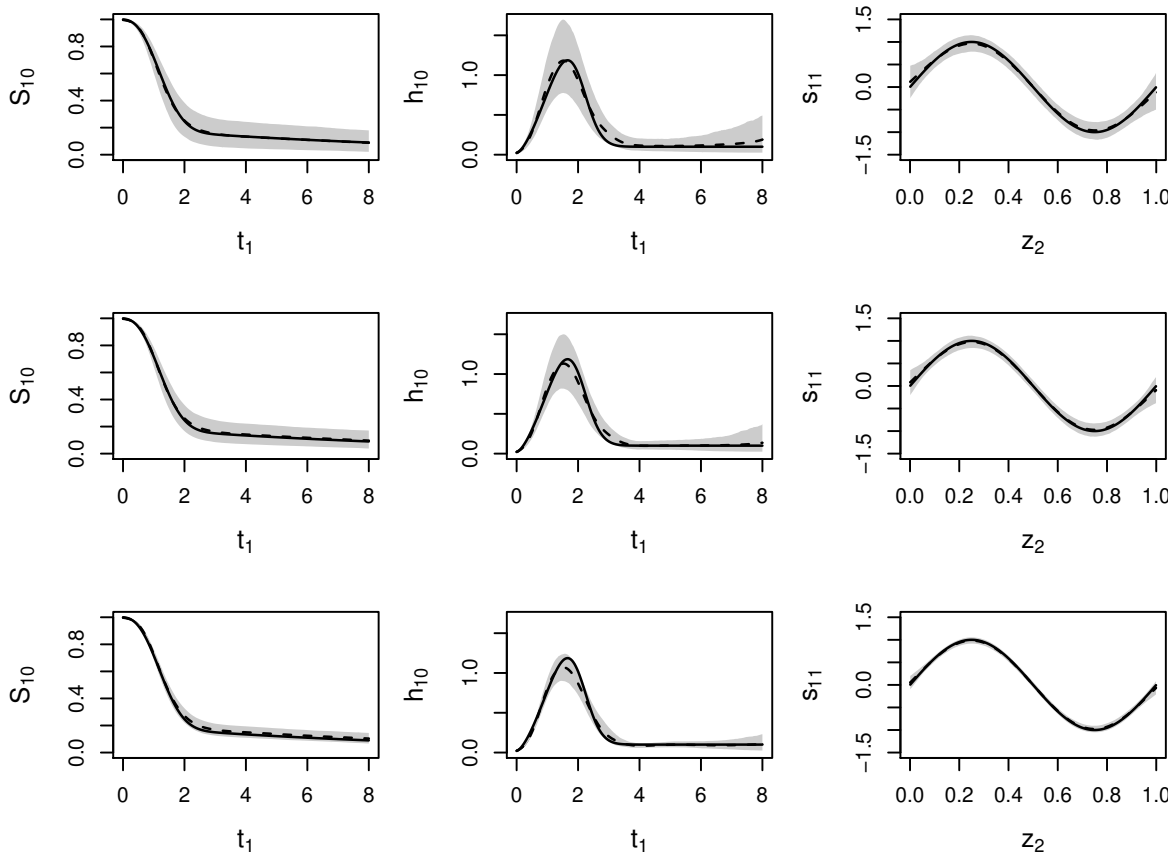

Figure 2: Smooth function estimates for the IPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP3 characterised by a censoring rate of about 47%. True functions are represented by black solid lines, mean estimates by dashed lines and pointwise ranges resulting from 5% and 95% quantiles by shaded areas. The results in the first row refer to $n = 500$, whereas those in the second and third rows to $n = 1000$ and $n = 4000$.

(a) Informative Penalized Maximum Log-likelihood Estimator (IPMLE)						
	Bias			RMSE		
	n = 500	n = 1000	n = 4000	n = 500	n = 1000	n = 4000
α_{11}	-0.024	-0.014	-0.006	0.138	0.100	0.049
s_1	0.039	0.025	0.012	0.154	0.114	0.059
h_{10}	0.084	0.048	0.035	0.262	0.144	0.083
S_{10}	0.028	0.020	0.017	0.063	0.050	0.031
(b) Non-informative Penalized Maximum Log-likelihood Estimator (NPMLE)						
	Bias			RMSE		
	n = 500	n = 1000	n = 4000	n = 500	n = 1000	n = 4000
α_{11}	-0.045	-0.017	-0.007	0.208	0.144	0.071
s_1	0.085	0.068	0.044	0.191	0.206	0.111
h_{10}	0.085	0.057	0.033	0.195	0.292	0.083
S_{10}	0.027	0.021	0.015	0.058	0.068	0.033

Table 3: Bias and root mean squared error (RMSE) for the IPMLE and NPMLE obtained by applying the `gamlss()` to informative survival data simulated according to DGP2 characterised by a censoring rate of about 74%. Further details are given in the caption of Table 1.

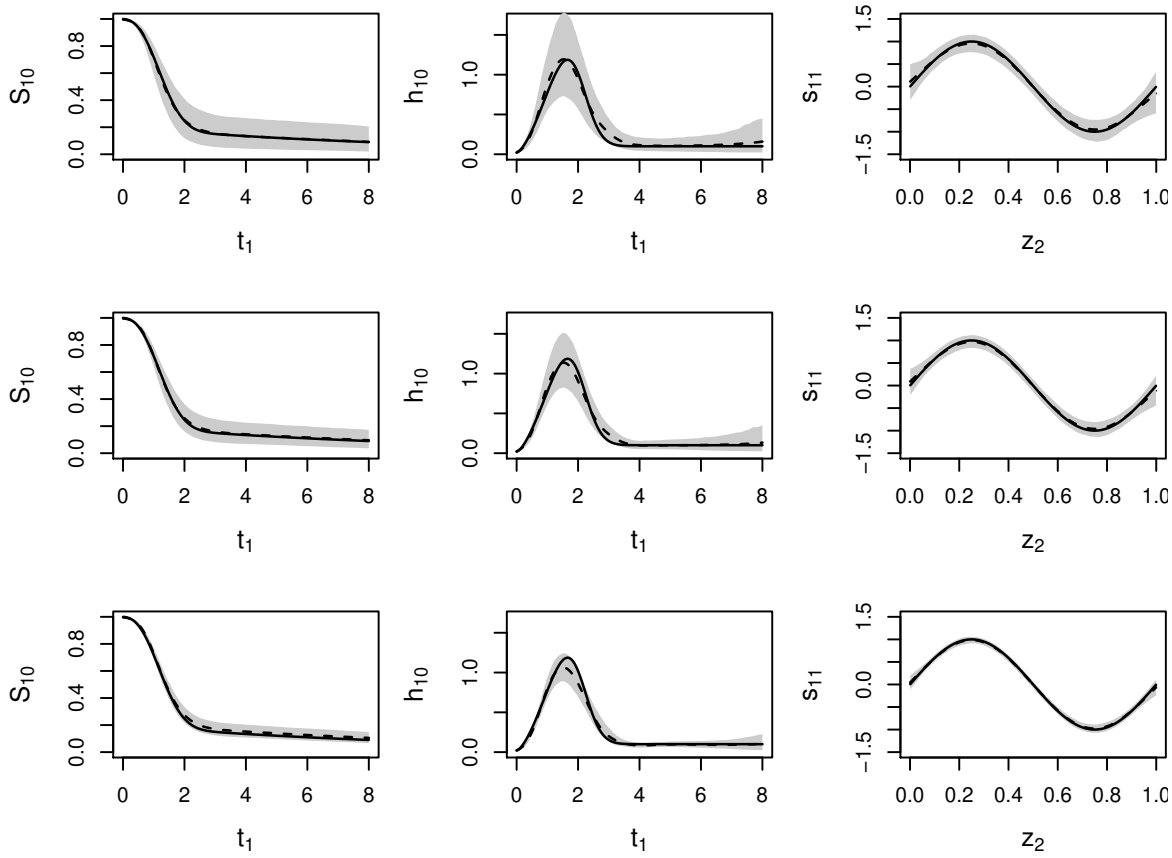

Figure 3: Smooth function estimates for the NPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP3 characterised by a censoring rate of about 47%. Further details are given in the caption of Figure 2.

(a) Informative Penalized Maximum Log-likelihood Estimator (IPMLE)						
	Bias			RMSE		
	$\mathrm{n} = 500$	$\mathrm{n} = 1000$	$\mathrm{n} = 4000$	$\mathrm{n} = 500$	$\mathrm{n} = 1000$	$\mathrm{n} = 4000$
α_{11}	-0.012	-0.006	0.003	0.121	0.058	0.045
s_1	0.031	0.021	0.015	0.124	0.091	0.051
h_{10}	0.040	0.027	0.026	0.135	0.088	0.058
S_{10}	0.003	0.008	0.015	0.057	0.047	0.030
(b) Non-informative Penalized Maximum Log-likelihood Estimator (NPMLE)						
	Bias			RMSE		
	$\mathrm{n} = 500$	$\mathrm{n} = 1000$	$\mathrm{n} = 4000$	$\mathrm{n} = 500$	$\mathrm{n} = 1000$	$\mathrm{n} = 4000$
α_{11}	-0.022	0.001	0.007	0.140	0.100	0.050
s_1	0.036	0.027	0.014	0.142	0.104	0.055
h_{10}	0.037	0.027	0.027	0.131	0.089	0.056
S_{10}	0.004	0.008	0.017	0.065	0.047	0.032

Table 4: Bias and root mean squared error (RMSE) for the IPMLE and NPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP3 characterised by a censoring rate of about 47%. Further details are given in the caption of Table 1.

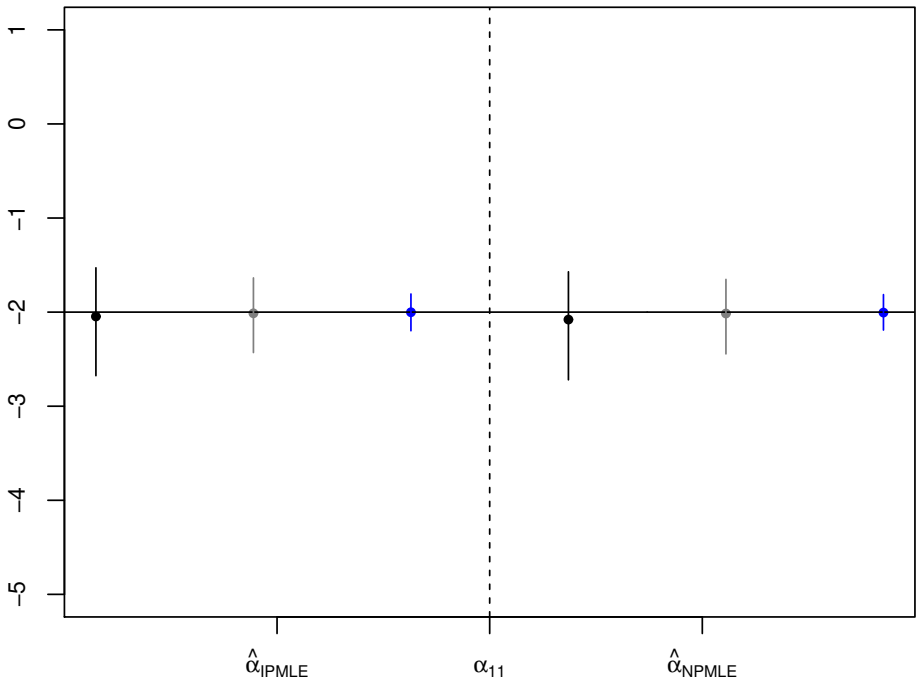

Figure 4: Linear coefficient estimates obtained by applying `gamlss()` to informative survival data simulated according to DGP1 which is characterised by a censoring rate of about 78%. Further details are given in the caption of Figure 1.

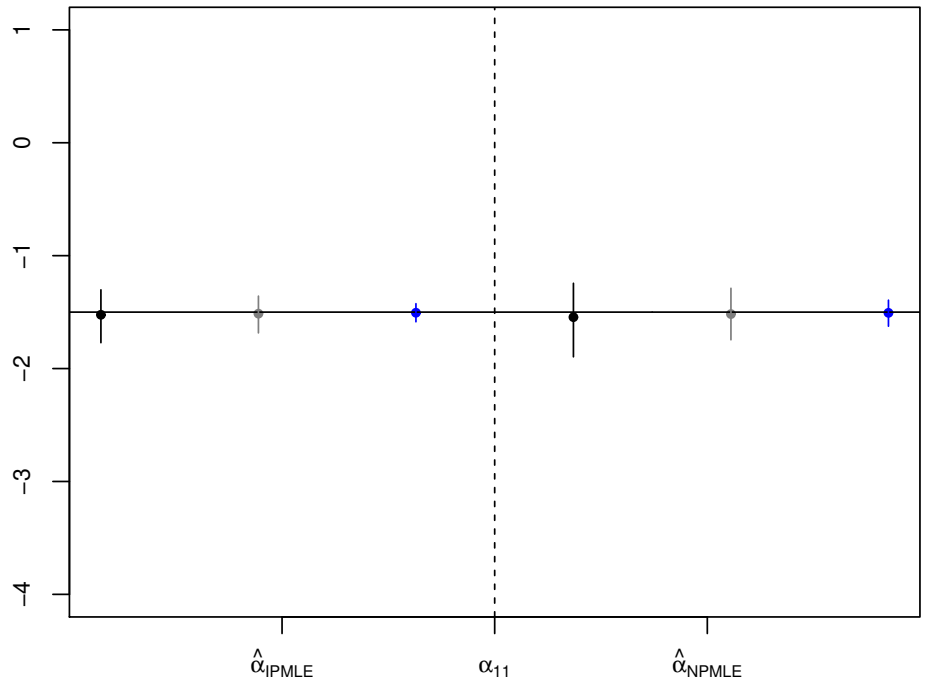

Figure 5: Linear coefficient estimates obtained by applying `gamlss()` to informative survival data simulated according to DGP2 which is characterised by a censoring rate of about 74%. Further details are given in the caption of Figure 1.

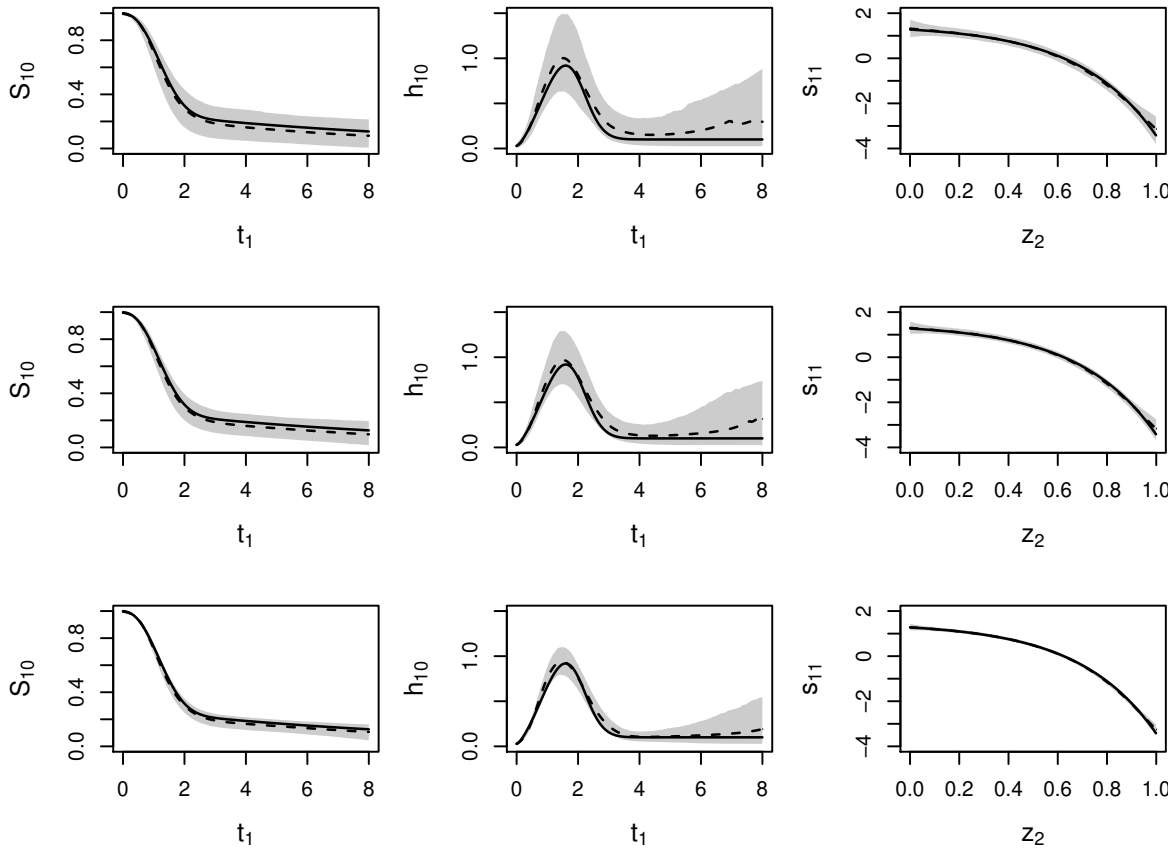

Figure 6: Smooth function estimates for the IPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP1 characterised by a censoring rate of about 78%. Further details are given in the caption of Figure 2.

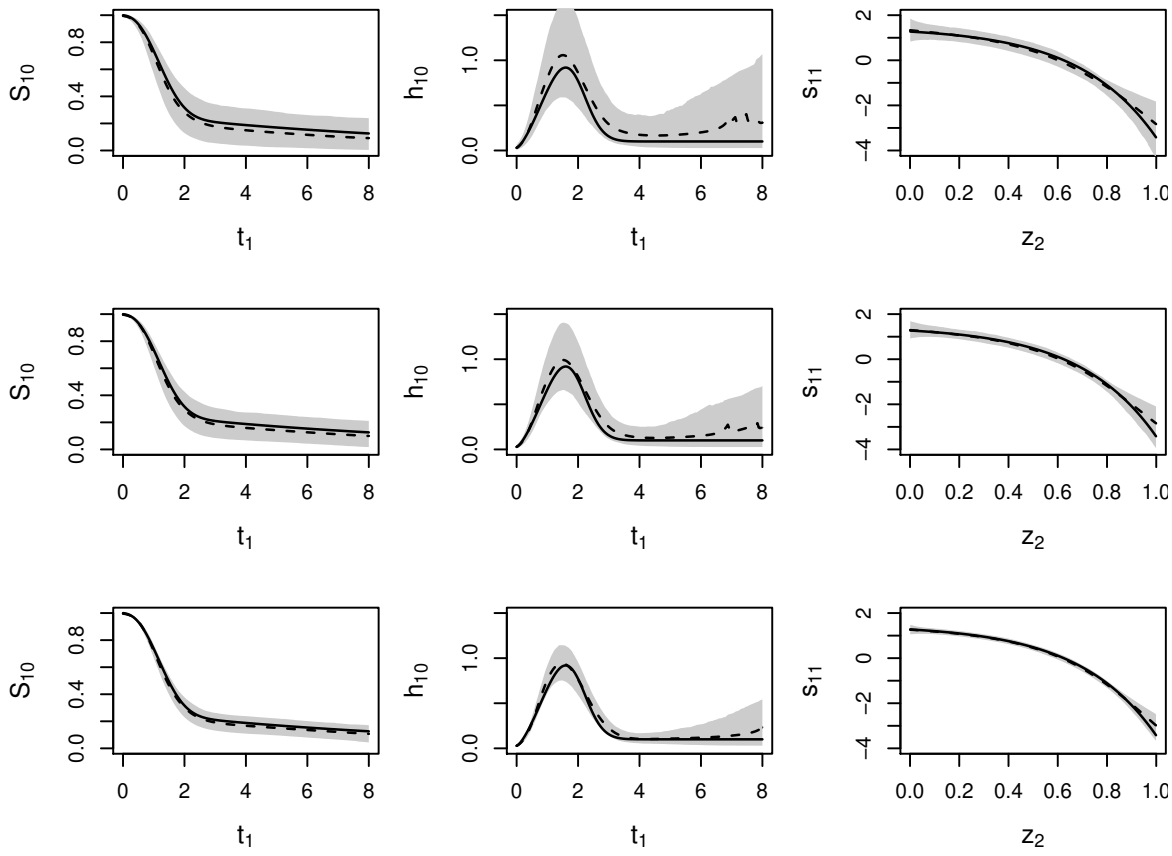

Figure 7: Smooth function estimates for the NPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP1 characterised by a censoring rate of about 78%. Further details are given in the caption of Figure 2.

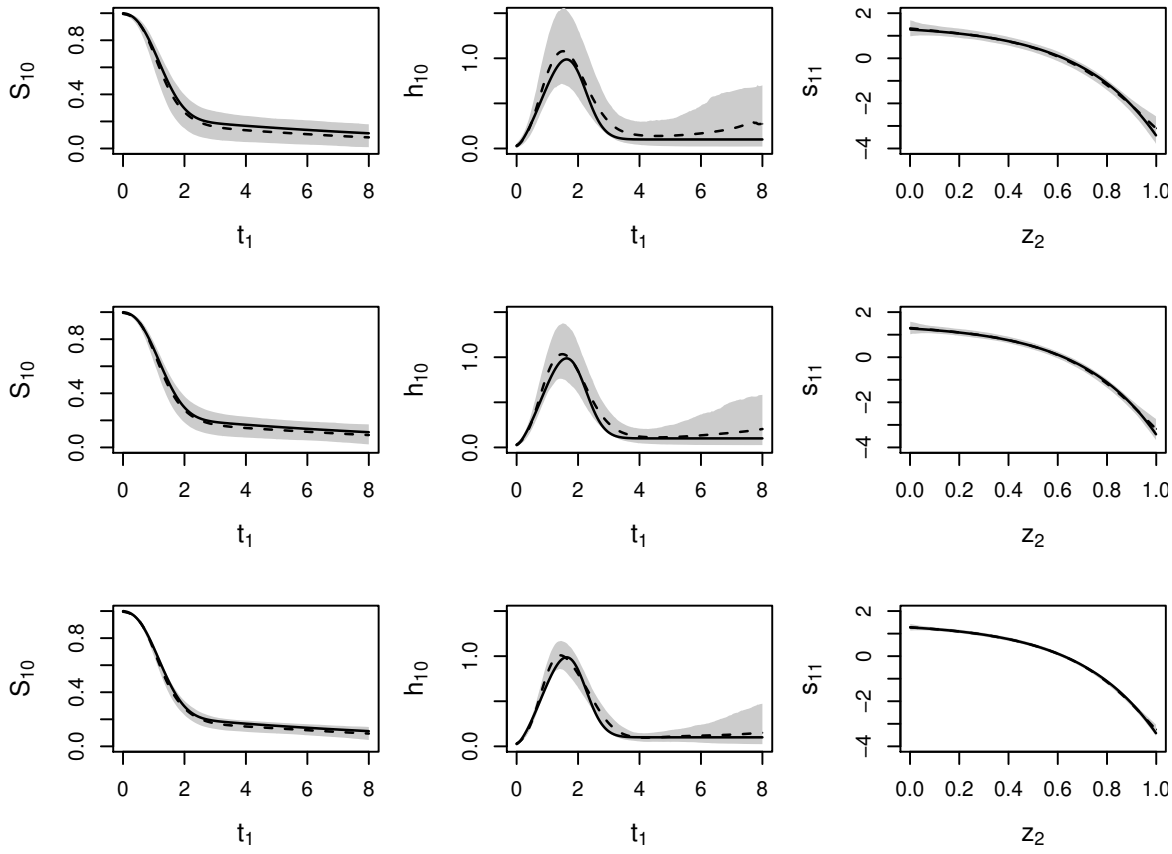

Figure 8: Smooth function estimates for the IPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP2 characterised by a censoring rate of about 74%. Further details are given in the caption of Figure 2

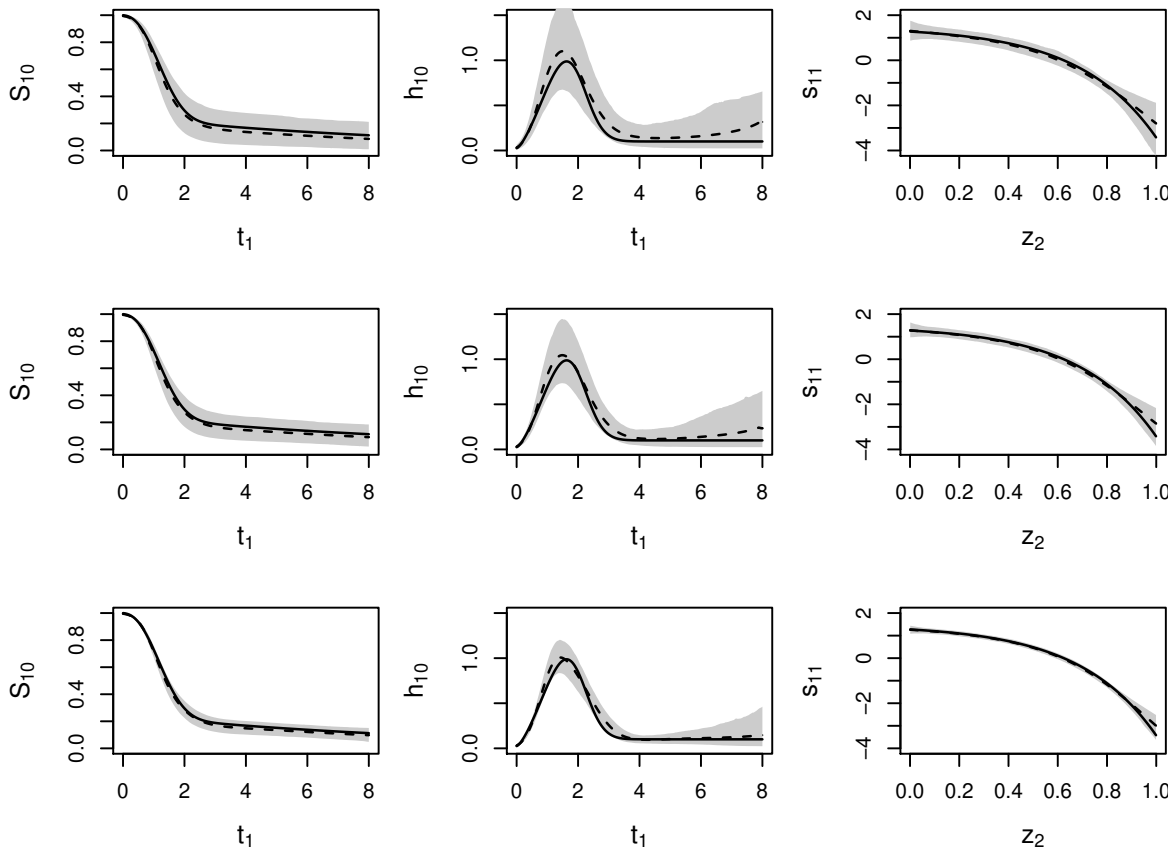

Figure 9: Smooth function estimates for the NPMLE obtained by applying `gamlss()` to informative survival data simulated according to DGP2 characterised by a censoring rate of about 74%. Further details are given in the caption of Figure 2

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In: Petrov, B.N., Csaki, B.F. (eds.) Second International Symposium on Information Theory. Academiai Kiado, Budapest*.

Gourieroux, C. & Monfort, A. (1995). *Statistics and econometric models*, volume 1. Cambridge University Press.

Gu, C. (1992). Cross-validating non-gaussian data. *Journal of Computational and Graphical Statistics*, 1(2), 169–179.

Hayashi, F. (2000). *Econometrics*. Princeton University Press Princeton, NJ.

Marra, G. & Radice, R. (2019). *GJRM: Generalised Joint Regression Modelling*. R package version 0.2-1.

Marra, G., Radice, R., Bärnighausen, T., Wood, S. N., & McGovern, M. E. (2017). A simultaneous equation approach to estimating hiv prevalence with nonignorable missing responses. *Journal of the American Statistical Association*, 112(518), 484–496.

Marra, G. & Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1), 53–74.

Newey, W. K. & McFadden, D. (1994). *Handbook of econometrics*, volume 4. Elsevier.

Nocedal, J. & Wright, S. (2006). *Numerical optimization, series in operations research and financial engineering*. Springer, New York, USA, 2006.

Pya, N. & Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*, 25(3), 543–559.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(1), 1–21.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, 36(2), 111–147.

Vatter, T. & Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141, 147–167.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673–686.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1), 221–228.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Second Edition, Chapman & Hall/CRC, London.

Xingwei, T., Tao, H., & Hengjian, C. (2010). Hazard regression with penalized spline: The smoothing parameter choice and asymptotics. *Acta Mathematica Scientia*, 30(5), 1759–1768.