Parametric proportional hazards and accelerated failure time models

Göran Broström

February 16, 2009

Abstract

A unified implementation of parametric proportional hazards (PH) and accelerated failure time (AFT) models for right-censored or interval-censored and left-truncated data is described. The description here is valid for time-constant covariates, but the necessary modifications for handling time-varying covariates are implemented in `eha`. Note that only piecewise constant time variation is handled.

1 Introduction

There is a need for software for analyzing parametric proportional hazards (PH) and accelerated failure time (AFT) data, that are right or interval censored and left truncated.

2 The proportional hazards model

We define proportional hazards models in terms of an expansion of a given survivor function S_0,

$$s_{\boldsymbol{\theta}}(t; \mathbf{z}) = \{S_0(g(t, \boldsymbol{\theta}))\}^{\exp(\mathbf{z}\boldsymbol{\beta})}, \tag{1}$$

where $\boldsymbol{\theta}$ is a parameter vector used in modeling the baseline distribution, $\boldsymbol{\beta}$ is the vector of regression parameters, and g is a positive function, which helps defining a parametric family of baseline survivor functions through

$$S(t; \boldsymbol{\theta}) = S_0\big(g(t, \boldsymbol{\theta})\big), \quad t > 0, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{2}$$

With f_0 and h_0 defined as the density and hazard functions corresponding to S_0, respectively, the density function corresponding to S is

$$\begin{aligned} f(t;\boldsymbol{\theta}) &= -\frac{\partial}{\partial t} S(t,\boldsymbol{\theta}) \\ &= -\frac{\partial}{\partial t} S_0(g(t,\boldsymbol{\theta})) \\ &= g_t(t,\boldsymbol{\theta}) f_0(g(t,\boldsymbol{\theta})), \end{aligned}$$

where

$$g_t(t,\boldsymbol{\theta}) = \frac{\partial}{\partial t} g(t,\boldsymbol{\theta}).$$

Correspondingly, the hazard function is

$$\begin{aligned} h(t;\boldsymbol{\theta}) &= \frac{f(t;\boldsymbol{\theta})}{S(t;\boldsymbol{\theta})} \\ &= g_t(t,\boldsymbol{\theta}) h_0(g(t,\boldsymbol{\theta})). \end{aligned} \tag{3}$$

So, the proportional hazards model is

$$\begin{aligned} \lambda_{\boldsymbol{\theta}}(t;\mathbf{z}) &= h(t;\boldsymbol{\theta}) \exp(\mathbf{z}\boldsymbol{\beta}) \\ &= g_t(t,\boldsymbol{\theta}) h_0(g(t,\boldsymbol{\theta})) \exp(\mathbf{z}\boldsymbol{\beta}), \end{aligned} \tag{4}$$

corresponding to (1).

2.1 Data and the likelihood function

Given left truncated and right or interval censored data $(s_i, t_i, u_i, d_i, \mathbf{z}_i)$, $i = 1, \ldots, n$ and the model (4), the likelihood function becomes

$$\begin{aligned} L\big((\boldsymbol{\theta},\boldsymbol{\beta}); (\mathbf{s},\mathbf{t},\mathbf{u},\mathbf{d}), \mathbf{Z}\big) = \prod_{i=1}^{n} \Big\{ & \big(h(t_i;\boldsymbol{\theta}) \exp(\mathbf{z}_i\boldsymbol{\beta})\big)^{I_{\{d_i=1\}}} \\ & \times \big(S(t_i;\boldsymbol{\theta})^{\exp(\mathbf{z}_i\boldsymbol{\beta})}\big)^{I_{\{d_i \neq 2\}}} \\ & \times \big(S(t_i;\boldsymbol{\theta})^{\exp(\mathbf{z}_i\boldsymbol{\beta})} - S(u_i;\boldsymbol{\theta})^{\exp(\mathbf{z}_i\boldsymbol{\beta})}\big)^{I_{\{d_i=2\}}} \\ & \times S(s_i;\boldsymbol{\theta})^{-\exp(\mathbf{z}_i\boldsymbol{\beta})} \Big\} \end{aligned} \tag{5}$$

Here, for $i = 1, \ldots, n$, $s_i < t_i \leq u_i$ are the left truncation and exit intervals, respectively, $d_i = 0$ means that $t_i = u_i$ and right censoring at u_i, $d_i = 1$ means that $t_i = u_i$ and an event at u_i, and $d_i = 2$ means that $t_i < u_i$ and an event occurs in the interval (t_i, u_i) (interval censoring) and $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})$ is a vector of explanatory variables with corresponding parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, $i = 1, \ldots, n$.

From (5) we now get the log likelihood and the score vector in a straightforward manner.

$$\begin{aligned}\ell\big((\boldsymbol{\theta},\boldsymbol{\beta});(\mathbf{s},\mathbf{t},\mathbf{u},\mathbf{d}),\mathbf{Z}\big) = &\sum_{i:d_i=1}\big\{\log h(t_i;\boldsymbol{\theta})+\mathbf{z}_i\boldsymbol{\beta}\big\}\\ &+\sum_{i:d_i\neq 2} e^{\mathbf{z}_i\boldsymbol{\beta}}\log S(u_i;\boldsymbol{\theta})\\ &+\sum_{i:d_i=2}\log\big\{S(t_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}-S(u_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}\big\}\\ &-\sum_{i=1}^{n} e^{\mathbf{z}_i\boldsymbol{\beta}}\log S(s_i;\boldsymbol{\theta})\end{aligned}\tag{6}$$

and (in the following we drop the long argument list to ℓ), for the regression parameters $\boldsymbol{\beta}$,

$$\begin{aligned}\frac{\partial}{\partial\beta_j}\ell = &\sum_{i:d_i=1} z_{ij}\\ &+\sum_{i:d_i\neq 2} z_{ij}e^{\mathbf{z}_i\boldsymbol{\beta}}\log S(t_i;\boldsymbol{\theta})\\ &+\sum_{i:d_i=2} z_{ij}e^{\mathbf{z}_i\boldsymbol{\beta}}\frac{S(t_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}\log S(t_i;\boldsymbol{\theta})-S(u_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}\log S(u_i;\boldsymbol{\theta})}{S(t_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}-S(u_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}}\\ &-\sum_{i=1}^{n} z_{ij}e^{\mathbf{z}_i\boldsymbol{\beta}}\log S(s_i;\boldsymbol{\theta}),\quad j=1,\dots,p,\end{aligned}\tag{7}$$

and for the "baseline" parameters $\boldsymbol{\theta}$, in vector form,

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}\ell = &\sum_{i:d_i=1}\frac{h_{\boldsymbol{\theta}}(t_i,\boldsymbol{\theta})}{h(t_i,\boldsymbol{\theta})}\\ &+\sum_{i:d_i\neq 2} e^{\mathbf{z}_i\boldsymbol{\beta}}\frac{S_{\boldsymbol{\theta}}(t_i;\boldsymbol{\theta})}{S(t_i;\boldsymbol{\theta})}\\ &+\sum_{i:d_i=2} e^{\mathbf{z}_i\boldsymbol{\beta}}\frac{S(t_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}-1}S_{\boldsymbol{\theta}}(t_i;\boldsymbol{\theta})-S(u_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}-1}S_{\boldsymbol{\theta}}(u_i;\boldsymbol{\theta})}{S(t_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}-S(u_i;\boldsymbol{\theta})^{e^{\mathbf{z}_i\boldsymbol{\beta}}}}\\ &-\sum_{i=1}^{n} e^{\mathbf{z}_i\boldsymbol{\beta}}\frac{S_{\boldsymbol{\theta}}(s_i;\boldsymbol{\theta})}{S(s_i;\boldsymbol{\theta})}.\end{aligned}\tag{8}$$

From (3),

$$\begin{aligned}h_{\boldsymbol{\theta}}(t,\boldsymbol{\theta}) &= \frac{\partial}{\partial\boldsymbol{\theta}}h(t,\boldsymbol{\theta})\\ &= g_{t\boldsymbol{\theta}}(t,\boldsymbol{\theta})h_0(g(t,\boldsymbol{\theta}))+g_t(t,\boldsymbol{\theta})g_\theta(t,\boldsymbol{\theta})h_0'(g(t,\boldsymbol{\theta})),\end{aligned}\tag{9}$$

and, from (2),

$$\begin{aligned} S_{\boldsymbol{\theta}}(t;\boldsymbol{\theta}) &= \frac{\partial}{\partial\boldsymbol{\theta}} S(t;\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} S_0\bigl(g(t,\boldsymbol{\theta})\bigr) \\ &= -g_{\boldsymbol{\theta}}(t,\boldsymbol{\theta}) f_0\bigl(g(t,\boldsymbol{\theta})\bigr). \end{aligned} \tag{10}$$

For estimating standard errors, the observed information (the negative of the hessian) is useful. However, instead of the error-prone and tedious work of calculating analytic second-order derivatives, we will rely on approximations by numerical differentiation.

3 The shape–scale families

From (1) we get a *shape–scale* family of distributions by choosing $\boldsymbol{\theta} = (p, \lambda)$ and

$$g(t,(p,\lambda)) = \left(\frac{t}{\lambda}\right)^p, \quad t \ge 0; \quad p, \lambda > 0.$$

However, for reasons of efficient numerical optimization and normality of parameter estimates, we use the parametrisation $p = \exp(\gamma)$ and $\lambda = \exp(\alpha)$, thus redefining g to

$$g(t,(\gamma,\alpha)) = \left(\frac{t}{\exp(\alpha)}\right)^{\exp(\gamma)}, \quad t \ge 0; \quad -\infty < \gamma, \alpha < \infty. \tag{11}$$

For the calculation of the score and hessian of the log likelihood function, we need some partial derivatives of g. They are found in an appendix.

3.1 The Weibull family of distributions

The Weibull family of distributions is obtained by

$$S_0(t) = \exp(-t), \quad t \ge 0,$$

leading to

$$f_0(t) = \exp(-t), \quad t \ge 0,$$

and

$$h_0(t) = 1, \quad t \ge 0.$$

We need some first and second order derivatives of f and h, and they are particularly simple in this case, for h they are both zero, and for f we get

$$f_0'(t) = -\exp(-t), \quad t \ge 0.$$

3.2 The EV family of distributions

The EV (Extreme Value) family of distributions is obtained by setting

$$h_0(t) = \exp(t), \quad t \ge 0,$$

leading to

$$S_0(t) = \exp\{-(\exp(t) - 1)\}, \quad t \ge 0,$$

The rest of the necessary functions are easily derived from this.

3.3 The Gompertz family of distributions

The Gompertz family of distributions is given by

$$h(t) = \tau \exp(t/\lambda), \quad t \ge 0; \quad \tau, \lambda > 0.$$

This family is not directly possible to generate from the described shape-scale models, so it is treated separately by direct maximum likelihood.

3.4 Other families of distributions

Included in the *eha* package are the lognormal and the loglogistic distributions as well.

4 The accelerated failure time model

In the description of this family of models, we generate a scape-scale family of distributions as defined by the equations (2) and (11). We get

$$\begin{aligned} S(t; (\gamma, \alpha)) &= S_0\bigl(g(t, (\gamma, \alpha))\bigr) \\ &= S_0\left(\left\{\frac{t}{\exp(\alpha)}\right\}^{\exp(\gamma)}\right), \quad t > 0, \quad -\infty < \gamma, \alpha < \infty. \end{aligned} \tag{12}$$

Given a vector $\mathbf{z} = (z_1, \ldots, z_p)$ of explanatory variables and a vector of corresponding regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, the AFT regression model is defined by

$$\begin{aligned} S(t; (\gamma, \alpha, \boldsymbol{\beta})) &= S_0\bigl(g(t\exp(\mathbf{z}\boldsymbol{\beta}), (\gamma, \alpha))\bigr) \\ &= S_0\left(\left\{\frac{t\exp(\mathbf{z}\boldsymbol{\beta})}{\exp(\alpha)}\right\}^{\exp(\gamma)}\right) \\ &= S_0\left(\left\{\frac{t}{\exp(\alpha - \mathbf{z}\boldsymbol{\beta})}\right\}^{\exp(\gamma)}\right) \\ &= S_0\bigl(g(t, (\gamma, \alpha - \mathbf{z}\boldsymbol{\beta}))\bigr), \quad t > 0. \end{aligned} \tag{13}$$

So, by defining $\boldsymbol{\theta} = (\gamma, \alpha - \mathbf{z}\boldsymbol{\beta})$, we are back in the framework of Section 2. We get

$$f(t; \boldsymbol{\theta}) = g_t(t, \boldsymbol{\theta}) f_0(g(t, \boldsymbol{\theta}))$$

and

$$h(t; \boldsymbol{\theta}) = g_t(t, \boldsymbol{\theta}) h_0(g(t, \boldsymbol{\theta})), \tag{14}$$

defining the AFT model generated by the survivor function S_0 and corresponding density f_0 and hazard h_0.

4.1 Data and the likelihood function

Given left truncated and right or interval censored data $(s_i, t_i, u_i, d_i, \mathbf{z}_i)$, $i = 1, \ldots, n$ and the model (14), the likelihood function becomes

$$\begin{aligned} L\big((\gamma, \alpha, \boldsymbol{\beta}); (\mathbf{s}, \mathbf{t}, \mathbf{d}), \mathbf{Z}\big) = \prod_{i=1}^{n} \Big\{ & h(t_i; \boldsymbol{\theta}_i)^{I_{\{d_i=1\}}} \\ & \times S(t_i; \boldsymbol{\theta}_i)^{I_{\{i \neq 2\}}} \\ & \times \big(S(t_i; \boldsymbol{\theta}_i) - S(u_i; \boldsymbol{\theta}_i)\big)^{I_{\{d_i=2\}}} \\ & \times S(s_i; \boldsymbol{\theta}_i)^{-1} \Big\} \end{aligned} \tag{15}$$

Here, for $i = 1, \ldots, n$, $s_i < t_i \leq u_i$ are the left truncation and exit intervals, respectively, $d_i = 0$ means that $t_i = u_i$ and right censoring at t_i, $d_i = 1$ means that $t_i = u_i$ and an event at t_i, and $d_i = 2$ means that $t_i < u_i$ and an event uccurs in the interval (t_i, u_i) (interval censoring), and $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})$ is a vector of explanatory variables with corresponding parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, $i = 1, \ldots, n$.

From (15) we now get the log likelihood and the score vector in a straightforward manner.

$$\begin{aligned} \ell\big((\gamma, \alpha, \boldsymbol{\beta}); (\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{d}), \mathbf{Z}\big) = & \sum_{i:d_i=1} \log h(t_i; \boldsymbol{\theta}_i) \\ & + \sum_{i:d_i \neq 2} \log S(t_i; \boldsymbol{\theta}_i) \\ & + \sum_{i:d_i=2} \log\big(S(t_i; \boldsymbol{\theta}_i) - S(u_i; \boldsymbol{\theta}_i)\big) \\ & - \sum_{i=1}^{n} \log S(s_i; \boldsymbol{\theta}_i) \end{aligned}$$

and (in the following we drop the long argument list to ℓ), for the regression parameters $\boldsymbol{\beta}$,

$$
\begin{aligned}
\frac{\partial}{\partial \beta_j}\ell &= \sum_{d_i=1} \frac{h_j(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} + \sum_{d_i\neq 2} \frac{S_j(t_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i)} \\
&\quad + \sum_{d_i=2} \frac{S_j(t_i;\boldsymbol{\theta}_i) - S_j(u_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i) - S(u_i;\boldsymbol{\theta}_i)} - \sum_{i=1}^{n} \frac{S_j(s_i;\boldsymbol{\theta}_i)}{S(s_i;\boldsymbol{\theta}_i)} \\
&= -\sum_{d_i=1} z_{ij} \frac{h_\alpha(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} - \sum_{d_i\neq 2} z_{ij} \frac{S_\alpha(t_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i)} \\
&\quad - \sum_{d_i=2} z_{ij} \frac{S_\alpha(t_i;\boldsymbol{\theta}_i) - S_\alpha(u_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i) - S(u_i;\boldsymbol{\theta}_i)} + \sum_{i=1}^{n} z_{ij} \frac{S_\alpha(s_i;\boldsymbol{\theta}_i)}{S(s_i;\boldsymbol{\theta}_i)}
\end{aligned}
$$

and for the "baseline" parameters γ and α,

$$
\begin{aligned}
\frac{\partial}{\partial \gamma}\ell &= \sum_{i:d_i=1} \frac{h_\gamma(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} + \sum_{i:d_i\neq 2} \frac{S_\gamma(t_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i)} \\
&\quad + \sum_{i:d_i=2} \frac{S_\gamma(t_i;\boldsymbol{\theta}_i) - S_\gamma(u_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i) - S(u_i;\boldsymbol{\theta}_i)} - \sum_{i=1}^{n} \frac{S_\gamma(s_i;\boldsymbol{\theta}_i)}{S(s_i;\boldsymbol{\theta}_i)},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial \alpha}\ell &= \sum_{i:d_i=1} \frac{h_\alpha(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} + \sum_{i:d_i\neq 2} \frac{S_\alpha(t_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i)} \\
&\quad + \sum_{i:d_i=2} \frac{S_\alpha(t_i;\boldsymbol{\theta}_i) - S_\alpha(u_i;\boldsymbol{\theta}_i)}{S(t_i;\boldsymbol{\theta}_i) - S(u_i;\boldsymbol{\theta}_i)} - \sum_{i=1}^{n} \frac{S_\alpha(s_i;\boldsymbol{\theta}_i)}{S(s_i;\boldsymbol{\theta}_i)}.
\end{aligned}
$$

Here, from (3),

$$
\begin{aligned}
h_\gamma(t,\boldsymbol{\theta}_i) &= \frac{\partial}{\partial \gamma} h(t,\boldsymbol{\theta}_i) \\
&= g_{t\gamma}(t,\boldsymbol{\theta}_i) h_0(g(t,\boldsymbol{\theta}_i)) + g_t(t,\boldsymbol{\theta}_i) g_\gamma(t,\boldsymbol{\theta}_i) h_0'(g(t,\boldsymbol{\theta}_i)),
\end{aligned}
$$

$$
\begin{aligned}
h_\alpha(t,\boldsymbol{\theta}_i) &= \frac{\partial}{\partial \alpha} h(t,\boldsymbol{\theta}_i) \\
&= g_{t\alpha}(t,\boldsymbol{\theta}_i) h_0(g(t,\boldsymbol{\theta}_i)) + g_t(t,\boldsymbol{\theta}_i) g_\alpha(t,\boldsymbol{\theta}_i) h_0'(g(t,\boldsymbol{\theta}_i)),
\end{aligned}
$$

and

$$
\begin{aligned}
h_j(t,\boldsymbol{\theta}_i) &= \frac{\partial}{\partial \beta_j} h(t,\boldsymbol{\theta}_i) = \frac{\partial}{\partial \alpha} h(t,\boldsymbol{\theta}_i) \frac{\partial}{\partial \beta_j}(\alpha - \mathbf{z}_i\boldsymbol{\beta}) \\
&= -z_{ij} h_\alpha(t,\boldsymbol{\theta}_i), \quad j = 1, \ldots, p.
\end{aligned}
$$

Similarly, from (2) we get

$$\begin{aligned}S_\gamma(t;\boldsymbol{\theta}_i) &= \frac{\partial}{\partial\gamma}S(t;\boldsymbol{\theta}_i) = \frac{\partial}{\partial\gamma}S_0\big(g(t,\boldsymbol{\theta}_i)\big)\\ &= -g_\gamma(t,\boldsymbol{\theta}_i)f_0\big(g(t,\boldsymbol{\theta}_i)\big),\end{aligned}$$

$$\begin{aligned}S_\alpha(t;\boldsymbol{\theta}_i) &= \frac{\partial}{\partial\alpha}S(t;\boldsymbol{\theta}_i) = \frac{\partial}{\partial\alpha}S_0\big(g(t,\boldsymbol{\theta}_i)\big)\\ &= -g_\alpha(t,\boldsymbol{\theta}_i)f_0\big(g(t,\boldsymbol{\theta}_i)\big).\end{aligned}$$

and

$$\begin{aligned}S_j(t;\boldsymbol{\theta}_i) &= \frac{\partial}{\partial\beta_j}S(t;\boldsymbol{\theta}_i) = \frac{\partial}{\partial\alpha}S_0\big(g(t,\boldsymbol{\theta}_i)\big)\frac{\partial}{\partial\beta_j}(\alpha - \mathbf{z}_i\boldsymbol{\beta})\\ &= -z_{ij}S_\alpha(t,\boldsymbol{\theta}_i), \quad j = 1,\ldots,p.\end{aligned}$$

For estimating standard errors, the observed information (the negative of the hessian) is useful, so

$$\begin{aligned}-\frac{\partial^2}{\partial\beta_j\partial\beta_m}\ell = &-\sum_{i:d_i=1} z_{ij}z_{im}\left\{\frac{h_{\alpha\alpha}(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} - \left(\frac{h_\alpha(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)}\right)^2\right\}\\ &-\sum_{i:i\neq 2} z_{ij}z_{im}\left\{\frac{S_{\alpha\alpha}(t_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i)} - \left(\frac{S_\alpha(t_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i)}\right)^2\right\}\\ -\sum_{i:i=2} z_{ij}z_{im}&\left\{\frac{S_{\alpha\alpha}(t_i,\boldsymbol{\theta}_i) - S_{\alpha\alpha}(u_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)} - \left(\frac{S_\alpha(t_i,\boldsymbol{\theta}_i) - S_\alpha(u_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)}\right)^2\right\}\\ &+\sum_{i=1}^{n} z_{ij}z_{im}\left\{\frac{S_{\alpha\alpha}(s_i,\boldsymbol{\theta}_i)}{S(s_i,\boldsymbol{\theta}_i)} - \left(\frac{S_\alpha(s_i,\boldsymbol{\theta}_i)}{S(s_i,\boldsymbol{\theta}_i)}\right)^2\right\}, \quad j,m = 1,\ldots,p,\end{aligned}$$

and

$$
\begin{aligned}
-\frac{\partial^2}{\partial\beta_j\partial\tau}\ell = \sum_{i:d_i=1} z_{ij}\left\{\frac{h_{\alpha\tau}(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} - \frac{h_\alpha(t_i,\boldsymbol{\theta}_i)h_\tau(t_i,\boldsymbol{\theta}_i)}{h^2(t_i,\boldsymbol{\theta}_i)}\right\} \\
+ \sum_{i:i\neq 2} z_{ij}\left\{\frac{S_{\alpha\tau}(t_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i)} - \frac{S_\alpha(t_i,\boldsymbol{\theta}_i)S_\tau(t_i,\boldsymbol{\theta}_i)}{S^2(t_i,\boldsymbol{\theta}_i)}\right\} \\
+ \sum_{i:i=2} z_{ij}\left\{\frac{S_{\alpha\tau}(t_i,\boldsymbol{\theta}_i) - S_{\alpha\tau}(u_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)}\right. \\
\left. - \frac{\big(S_\alpha(t_i,\boldsymbol{\theta}_i) - S_\alpha(u_i,\boldsymbol{\theta}_i)\big)\big(S_\tau(t_i,\boldsymbol{\theta}_i) - S_\tau(u_i,\boldsymbol{\theta}_i)\big)}{\big(S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)\big)^2}\right\} \\
- \sum_{i=1}^{n} z_{ij}\left\{\frac{S_{\alpha\tau}(s_i,\boldsymbol{\theta}_i)}{S(s_i,\boldsymbol{\theta}_i)} - \frac{S_\alpha(s_i,\boldsymbol{\theta}_i)S_\tau(s_i,\boldsymbol{\theta}_i)}{S^2(s_i,\boldsymbol{\theta}_i)}\right\} \\
j = 1,\ldots,p;\ \tau = \gamma,\alpha,
\end{aligned}
$$

and finally

$$
\begin{aligned}
-\frac{\partial^2}{\partial\tau\partial\tau'}\ell = -\sum_{i:d_i=1}\left\{\frac{h_{\tau'\tau}(t_i,\boldsymbol{\theta}_i)}{h(t_i,\boldsymbol{\theta}_i)} - \frac{h_{\tau'}(t_i,\boldsymbol{\theta}_i)h_\tau(t_i,\boldsymbol{\theta}_i)}{h^2(t_i,\boldsymbol{\theta}_i)}\right\} \\
- \sum_{i:i\neq 2}\left\{\frac{S_{\tau'\tau}(t_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i)} - \frac{S_{\tau'}(t_i,\boldsymbol{\theta}_i)S_\tau(t_i,\boldsymbol{\theta}_i)}{S^2(t_i,\boldsymbol{\theta}_i)}\right\} \\
- \sum_{i:i=2}\left\{\frac{S_{\tau'\tau}(t_i,\boldsymbol{\theta}_i) - S_{\tau'\tau}(u_i,\boldsymbol{\theta}_i)}{S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)}\right. \\
\left. - \frac{\big(S_{\tau'}(t_i,\boldsymbol{\theta}_i) - S_{\tau'}(u_i,\boldsymbol{\theta}_i)\big)\big(S_\tau(t_i,\boldsymbol{\theta}_i) - S_\tau(u_i,\boldsymbol{\theta}_i)\big)}{\big(S(t_i,\boldsymbol{\theta}_i) - S(u_i,\boldsymbol{\theta}_i)\big)^2}\right\} \\
+ \sum_{i=1}^{n}\left\{\frac{S_{\tau'\tau}(s_i,\boldsymbol{\theta}_i)}{S(s_i,\boldsymbol{\theta}_i)} - \frac{S_{\tau'}(s_i,\boldsymbol{\theta}_i)S_\tau(s_i,\boldsymbol{\theta}_i)}{S^2(s_i,\boldsymbol{\theta}_i)}\right\} \\
(\tau,\tau') = (\gamma,\gamma),(\gamma,\alpha),(\alpha,\alpha).
\end{aligned}
$$

The second order partial derivatives $h_{\tau\tau'}$ and $S_{\tau\tau'}$ are

$$\begin{aligned}
h_{\tau\tau'}(t,\boldsymbol{\theta}) &= \frac{\partial}{\partial\tau'}h_\tau(t,\boldsymbol{\theta}) \\
&= g_{t\tau\tau'}(t,\boldsymbol{\theta})h_0(g(t,\boldsymbol{\theta})) + g_{t\tau}(t,\boldsymbol{\theta})g_{\tau'}(t,\boldsymbol{\theta})h_0'(g(t,\boldsymbol{\theta})) \\
&\quad + g_t(t,\boldsymbol{\theta})g_\theta(t,\boldsymbol{\theta})g_{\tau'}(t,\boldsymbol{\theta})h_0''(g(t,\boldsymbol{\theta})) \\
&\quad + g_t(t,\boldsymbol{\theta})g_{\theta\theta'}(t,\boldsymbol{\theta})h_0'(g(t,\boldsymbol{\theta})) \\
&\quad + g_{t\tau'}(t,\boldsymbol{\theta})g_\theta(t,\boldsymbol{\theta})h_0'(g(t,\boldsymbol{\theta})) \\
&= h_0(g(t,\boldsymbol{\theta}))g_{t\tau\tau'}(t,\boldsymbol{\theta}) \\
&\quad + h_0'(g(t,\boldsymbol{\theta}))\big\{g_t(t,\theta)g_{\theta\theta'}(t,\boldsymbol{\theta}) \\
&\qquad\qquad + g_{t\tau}(t,\boldsymbol{\theta})g_{\tau'}(t,\boldsymbol{\theta}) \\
&\qquad\qquad + g_{t\tau'}(t,\boldsymbol{\theta})g_\tau(t,\boldsymbol{\theta})\big\} \\
&\quad + h_0''(g(t,\boldsymbol{\theta}))g_t(t,\boldsymbol{\theta})g_\theta(t,\boldsymbol{\theta})g_{\tau'}(t,\boldsymbol{\theta}),\,, \\
&\quad (\tau,\tau') = (\gamma,\gamma),(\gamma,\lambda),(\lambda,\lambda),
\end{aligned} \tag{16}$$

and from (10),

$$\begin{aligned}
S_{\tau\tau'}(t,\boldsymbol{\theta}) &= \frac{\partial}{\partial\tau'}S_\tau(t;\boldsymbol{\theta}) \\
&= -\big\{g_{\tau\tau'}(t,\boldsymbol{\theta})f_0\big(g(t,\boldsymbol{\theta})\big) + g_\tau(t,\boldsymbol{\theta})g_{\tau'}(t,\boldsymbol{\theta})f_0'\big(g(t,\boldsymbol{\theta})\big)\big\}, \\
&\quad (\tau,\tau') = (\gamma,\gamma),(\gamma,\lambda),(\lambda,\lambda).
\end{aligned} \tag{17}$$

A Some partial derivatives

The function (see (11))

$$g(t,(\gamma,\alpha)) = \left(\frac{t}{\exp(\alpha)}\right)^{\exp(\gamma)}, \quad t \geq 0; \quad -\infty < \gamma,\alpha < \infty. \tag{18}$$

has the following partial derivatives:

$$\begin{aligned}
g_t\big(t,(\gamma,\alpha)\big) &= \frac{\exp(\gamma)}{t}g\big(t,(\gamma,\alpha)\big), \quad t > 0 \\
g_\gamma\big(t,(\gamma,\alpha)\big) &= g\big(t,(\gamma,\alpha)\big)\log\big\{g\big(t,(\gamma,\alpha)\big)\big\} \\
g_\alpha\big(t,(\gamma,\alpha)\big) &= -\exp(\gamma)g\big(t,(\gamma,\alpha)\big)
\end{aligned}$$

$$
\begin{aligned}
g_{t\gamma}\big(t,(\gamma,\alpha)\big) &= g_t\big(t,(\gamma,\alpha)\big) + \frac{\exp(\gamma)}{t} g_\gamma\big(t,(\gamma,\alpha)\big), \quad t>0 \\
g_{t\alpha}\big(t,(\gamma,\alpha)\big) &= -\exp(\gamma) g_t\big(t,(\gamma,\alpha)\big), \quad t>0 \\
g_{\gamma^2}\big(t,(\gamma,\alpha)\big) &= g_\gamma\big(t,(\gamma,\alpha)\big)\big\{1+\log g\big(t,(\gamma,\alpha)\big)\big\} \\
g_{\gamma\alpha}\big(t,(\gamma,\alpha)\big) &= g_\alpha\big(t,(\gamma,\alpha)\big)\big\{1+\log g\big(t,(\gamma,\alpha)\big)\big\} \\
g_{\alpha^2}\big(t,(\gamma,\alpha)\big) &= -\exp(\gamma) g_\alpha\big(t,(\gamma,\alpha)\big)
\end{aligned}
$$

$$
\begin{aligned}
g_{t\gamma^2}\big(t,(\gamma,\alpha)\big) &= g_{t\gamma}\big(t,(\gamma,\alpha)\big) \\
&\qquad + \frac{\exp(\gamma)}{t} g_\gamma\big(t,(\gamma,\alpha)\big)\big\{2+\log g\big(t,(\gamma,\alpha)\big)\big\} \\
g_{t\gamma\alpha}\big(t,(\gamma,\alpha)\big) &= -\exp(\gamma)\big\{g_t\big(t,(\gamma,\alpha)\big) + g_{t\gamma}\big(t,(\gamma,\alpha)\big)\big\} \\
g_{t\alpha^2}\big(t,(\gamma,\alpha)\big) &= -\exp(\gamma) g_{t\alpha}\big(t,(\gamma,\alpha)\big)
\end{aligned}
$$

The formulas will be easier to read if we remove all function arguments, i.e., $(t,(\gamma,\alpha))$:

$$
\begin{aligned}
g_t &= \frac{\exp(\gamma)}{t} g, \quad t>0 \\
g_\gamma &= g\log g \\
g_\alpha &= -\exp(\gamma) g \\
g_{t\gamma} &= g_t + \frac{\exp(\gamma)}{t} g_\gamma, \quad t>0 \\
g_{t\alpha} &= -\exp(\gamma) g_t, \quad t>0 \\
g_{\gamma^2} &= g_\gamma\big\{1+\log g\big\} \\
g_{\gamma\alpha} &= g_\alpha\big\{1+\log g\big\} \\
g_{\alpha^2} &= -\exp(\gamma) g_\alpha \\
g_{t\gamma^2} &= g_{t\gamma} + \frac{\exp(\gamma)}{t} g_\gamma\big\{2+\log g\big\}, \quad t>0 \\
g_{t\gamma\alpha} &= -\exp(\gamma)\big\{g_t + g_{t\gamma}\big\}, \quad t>0 \\
g_{t\alpha^2} &= -\exp(\gamma) g_{t\alpha}, \quad t>0
\end{aligned}
$$