Appendix A: The Error Variance

Multivariate Proportional Hazards Models

To apply the method proposed in Section 3.2 we need estimated values for the error variance. To obtain estimates we can rely on an asymptotic representation, proposed by Lo and Singh (1986), decomposing $\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)$ as an average of i.i.d. terms and a lower order remainder term $r_{ik}(t)$, where $F_i^{(k)}$ is the continuous failure time distribution function for subjects in cluster $i$ with $x_{ij} = k$.

Let $G$ be the censoring distribution, $1 - H_{ik}(s) = \{1 - F_i^{(k)}(s)\}\{1 - G(s)\}$ and

$$H_{ik}^u(s) = P(T_{ij} \leq s, \delta_{ij} = 1|x_{ij} = k) = \int_0^s 1 - G(y^-)dF_i^{(k)}(y).$$

It follows from Lo and Singh (1986) that

$$\hat{F}_i^{(k)}(t) - F_i^{(k)}(t) = \frac{1}{n_{ik}} \sum_{j : x_{ij} = k} \xi_{ik}(T_{ij}, \delta_{ij}, t) + r_{ik}(t),$$
\[
\xi_{ik}(T_{ij}, \delta_{ij}, t) = \frac{I(T_{ij} \leq t, \delta_{ij} = 1)}{1 - H_{ik}(T_{ij})} - \int_0^t \frac{I(T_{ij} > s)}{\{1 - H_{ik}(s)\}^2} dH_{ik}(s),
\]
for a subject in cluster \(i\) with observed information \((T_{ij}, \delta_{ij})\) and \(x_{ij} = k\).

By using the relationship \(\Lambda_i^{(k)}(t) = -\ln \left(1 - F_i^{(k)}(t)\right)\) and first order Taylor expansions, we obtain
\[
\ln \hat{\Lambda}_i^{(k)}(t) - \ln \Lambda_i^{(k)}(t) \approx \frac{1}{\Lambda_i^{(k)}(t)} \left\{ \frac{1}{1 - F_i^{(k)}(t)} \right\} \left\{ \hat{F}_i^{(k)}(t) - F_i^{(k)}(t) \right\}
\]
\[
\approx \frac{1}{\Lambda_i^{(k)}(t)} \frac{1}{S_i^{(k)}(t)} \sum_{j: x_{ij} = k} \xi_{ik}(T_{ij}, \delta_{ij}, t).
\]

Let \(w(.)\) be a weight function, as defined in Section 3.2. Integrating both sides with respect to \(w(.)\) gives
\[
\hat{\Omega}_{ik} - \Omega_{ik} = \int_0^\infty \ln \hat{\Lambda}_i^{(k)}(t) dW(t) - \int_0^\infty \ln \Lambda_i^{(k)}(t) dW(t)
\]
\[
= \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \int_0^\infty \frac{\xi_{ik}(T_{ij}, \delta_{ij}, t)}{\Lambda_i^{(k)}(t) S_i^{(k)}(t)} dW(t)
\]
\[
= \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \eta_{ik}(T_{ij}, \delta_{ij}),
\]
with \(\eta_{ik}(T_{ij}, \delta_{ij}) = \int_0^\infty \frac{\xi_{ik}(T_{ij}, \delta_{ij}, t)}{\Lambda_i^{(k)}(t) S_i^{(k)}(t)} dW(t)\).

Noting that the function \(\xi_{ik}(T_{ij}, \delta_{ij}, t)\) is a conditional version (conditioned on the cluster \(i\) and the subgroup with \(x_{ij} = k\)) of the function \(\xi\) in Lo and Singh (1986), it follows that for a subject with observed information \((T_{ij}, \delta_{ij})\) in the subgroup with \(x_{ij} = k\):
\[
E \{\xi_{ik}(T_{ij}, \delta_{ij}, t)\} = 0,
\]
\[
\text{Cov} \{\xi_{ik}(T_{ij}, \delta_{ij}, t), \xi_{ik}(T_{ij}, \delta_{ij}, s)\} = \{1 - F_i^{(k)}(t)\} \{1 - F_i^{(k)}(s)\}
\]
\[
\times \int_{t \wedge s}^\infty \frac{dH_{ik}(y)}{\{1 - H_{ik}(y)\}^2}.
\]
Assume that subject \( l \) in cluster \( i \) is in the subgroup with \( x_{il} = k \). The asymptotic variance of the error terms \( \hat{\Omega}_{ik} - \Omega_{ik} \) is given by

\[
\sigma^2_{e,ik} = \text{Var}(\hat{\Omega}_{ik} - \Omega_{ik})
\]

\[
= \text{Var} \left\{ \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \eta_{ik}(T_{ij}, \delta_{ij}) \right\}
\]

\[
= \frac{1}{n_{ik}} \text{Var} \{ \eta(T_{il}, \delta_{il}) \}
\]

\[
= \frac{1}{n_{ik}} \mathbb{E} \left\{ \int_0^\infty \frac{\xi_{ik}(T_{il}, \delta_{il}, t)}{\Lambda_i^{(k)}(t)S_i^{(k)}(t)} dW(t) \int_0^\infty \frac{\xi_{ik}(T_{il}, \delta_{il}, s)}{\Lambda_i^{(k)}(s)S_i^{(k)}(s)} dW(s) \right\}
\]

\[
= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t)S_i^{(k)}(t)\Lambda_i^{(k)}(s)S_i^{(k)}(s)} \times \text{Cov} \{ \xi_{ik}(T_{il}, \delta_{il}, t), \xi_{ik}(T_{il}, \delta_{il}, s) \} dW(t)dW(s)
\]

\[
= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t)\Lambda_i^{(k)}(s)} \int_0^{t \wedge s} \frac{dH_i^{u}(y)(1 - H_i(y^-))}{(1 - H_i(y^-))^2} dt ds dW(t)dW(s).
\]

Let \( W \) be a uniform weight function on the interval \((A, B)\), where \( A \) and \( B \) are chosen so that \( 0 < S_i^{(k)}(t) < 1 \) for \( t \in (A, B) \). Then

\[
\sigma^2_{e,ik} = \frac{1}{n_{ik}(B - A)^2} \int_A^B \int_A^s \frac{1}{\Lambda_i^{(k)}(t)\Lambda_i^{(k)}(s)} \int_0^t \frac{dH_i^{u}(y)(1 - H_i(y^-))}{(1 - H_i(y^-))^2} dt ds dW(t)dW(s)
\]

\[
+ \frac{1}{n_{ik}(B - A)^2} \int_A^B \int_s^B \frac{1}{\Lambda_i^{(k)}(t)\Lambda_i^{(k)}(s)} \int_0^s \frac{dH_i^{u}(y)(1 - H_i(y^-))}{(1 - H_i(y^-))^2} dt ds dW(t)dW(s).
\]

To obtain an estimate of the asymptotic error variance, we replace \( H_i(y^-) \) and \( H_i^{u}(y) \) by the following empirical estimators:

\[
\hat{H}_{ik}(y) = \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} \leq y, \delta_{ij} = 1)
\]

\[
\hat{H}_{ik}(y^-) = \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < y).
\]
This gives the following estimated variances of the error terms:

\[
\hat{\sigma}_{e,ik}^2 = \frac{1}{n_{ik}} \frac{1}{(B - A)^2} \int_A^B \int_s^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)} \frac{1}{\Lambda_i^{(k)}(t)} \times \sum_{j: x_{ij} = k} \left\{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})\right\} I(0 \leq t_{ij} \leq t, \delta_{ij} = 1) \right)^2 dt \, ds \\
+ \frac{1}{n_{ik}^2} \frac{1}{(B - A)^2} \int_A^B \int_A^B \frac{1}{\Lambda_i^{(n)}(s)} \frac{1}{\Lambda_i^{(n)}(t)} \times \sum_{j: x_{ij} = k} \left\{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})\right\} I(0 \leq t_{ij} \leq s, \delta_{ij} = 1) \right)^2 dt \, ds.
\]

**Multivariate Proportional Odds Models**

It follows from the first order Taylor expansion and the i.i.d. representation by Lo and Singh (1986) that

\[
\logit\left\{\hat{F}_i^{(k)}(t)\right\} - \logit\left\{F_i^{(k)}(t)\right\} \approx \frac{1}{\hat{F}_i^{(k)}(t) \left\{1 - \hat{F}_i^{(k)}(t)\right\}} \left\{\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)\right\} \\
\approx \frac{1}{1 - S_i^{(k)}(t)} S_i^{(k)}(t) \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \xi_{ik} (T_{ij}, \delta_{ij}, t).
\]

Integrating with respect to \(w(.)\) gives

\[
\hat{\Omega}_{PO; i}^{(k)} - \Omega_{PO; i}^{(k)} = \int_0^\infty \logit\left\{\hat{F}_i^{(k)}(t)\right\} dW(t) - \int_0^\infty \logit\left\{F_i^{(k)}(t)\right\} dW(t) \\
\approx \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \int_0^\infty \frac{\xi_{ik} (T_{ij}, \delta_{ij}, t)}{1 - S_i^{(k)}(t)} S_i^{(k)}(t) dW(t).
\]

The estimated variance of the error terms \(\hat{\Omega}_{PO; i}^{(k)} - \Omega_{PO; i}^{(k)}\) can be obtained using the same arguments as explained for the multivariate proportional
hazards model (frailty model):

\[
\sigma_{e i}^2 = \frac{1}{n_{ik}} \frac{1}{(B - A)^2} \int_{A}^{B} \int_{A}^{s} \left\{ 1 - S_i^{(k)}(t) \right\} \left\{ 1 - S_i^{(k)}(s) \right\} dt \times ds 
\]

\[
\times \sum_{j: x_{ij} = k} \frac{1}{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})} \frac{1}{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})} dt \times ds 
\]

\[
+ \frac{1}{n_{ik}} \frac{1}{(B - A)^2} \int_{A}^{B} \int_{A}^{s} \left\{ 1 - S_i^{(k)}(t) \right\} \left\{ 1 - S_i^{(k)}(s) \right\} ds 
\]

Multivariate Additive Risks Models

Using \( \hat{\Lambda}_i^{(k)}(t) = -\ln \left\{ 1 - F_i^{(k)}(t) \right\} \) and the first order Taylor expansion gives

\[
\hat{\Lambda}_i^{(k)}(t) - \Lambda_i^{(k)}(t) \approx \frac{1}{1 - F_i^{(k)}(t)} \left\{ \hat{F}_i^{(k)}(t) - F_i^{(k)}(t) \right\} 
\]

\[
\approx \frac{1}{1 - F_i^{(k)}(t)} \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \xi_{ik}(T_{ij}, \delta_{ij}, t). 
\]

The last equation follows by Lo and Singh (1986). By integrating both sides with respect to the weight function \( \tilde{W}(t) = W(t) / \int_{0}^{\infty} sdW(s) \), we obtain

\[
\hat{\Omega}_{AR;i}^{(k)} - \Omega_{AR;i}^{(k)} = \int_{0}^{\infty} \hat{\Lambda}_i^{(k)}(t) d\tilde{W}t - \int_{0}^{\infty} \Lambda_i^{(k)}(t) d\tilde{W}(t) 
\]

\[
\approx \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} \int_{0}^{\infty} \frac{\xi_{ik}(T_{ij}, \delta_{ij}, t)}{S_i^{(k)}(t)} d\tilde{W}(t). 
\]

As in the previous sections, we choose a uniform weight function \( W \) on the interval \( (A, B) \). The estimated variance of the error terms \( \hat{\Omega}_{AR;i}^{(k)} - \Omega_{AR;i}^{(k)} \) can be obtained in a similar way as the discussion given for the frailty model:

\[
\sigma_{e AR;ik}^2 = \frac{1}{n_{ik}} \frac{4}{(B^2 - A^2)^2} \int_{A}^{B} \int_{A}^{s} \sum_{j: x_{ij} = k} \frac{I(0 < t_{ij} \leq t, \delta_{ij} = 1)}{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})} dt \times ds 
\]

\[
+ \frac{1}{n_{ik}} \frac{4}{(B^2 - A^2)^2} \int_{A}^{B} \int_{s}^{B} \sum_{j: x_{ij} = k} \frac{I(0 < t_{ij} \leq s, \delta_{ij} = 1)}{1 - \frac{1}{n_{ik}} \sum_{j: x_{ij} = k} I(T_{ij} < t_{ij})} dt \times ds. 
\]
Appendix B: Fitting the Linear Mixed-Effects Model

To fit the transformed models (6) and (7) in Section 3.2, we use PROC MIXED in SAS. The mixed-effects model is written as

\[ y = X\beta + Z\gamma + e, \]  

(1)

where \( y \) denotes the vector of dependent variable values, \( \beta \) is an unknown vector of fixed effects with known model matrix \( X \), \( \gamma \) is an unknown vector of random effects with known model matrix \( Z \), and \( e \) is the random error vector. A key assumption is that \( \gamma \) and \( e \) are normally distributed with

\[ \mathbb{E}(\gamma | e) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathcal{D}(\gamma | e) = \begin{pmatrix} \text{G} & 0 \\ 0 & \text{R} \end{pmatrix}. \]

The variance-covariance matrix of \( y \) is therefore \( V = ZGZ^T + R \). To estimate the variance-covariance components in model (1), PROC MIXED implements two likelihood-based methods: maximum likelihood (ML) and restricted/residual likelihood (REML). We will consider the REML method. The corresponding log likelihood function is:

\[ l_R(G, R) = -\frac{1}{2} \log|V| - \frac{1}{2} \log|X^TV^{-1}X| - \frac{1}{2} r^TV^{-1}r - \frac{n-p}{2} \log 2\pi, \]  

(2)

where \( r = y - X(X^TV^{-1}X)^{-1}X^TV^{-1}y \) and \( p \) is the rank of \( X \). PROC MIXED minimizes \(-2 l_R(G, R)\) over all unknown parameters using a ridge-stabilized Newton-Raphson algorithm.

For model (7) in Section 3.2, \( y^T = (\hat{\Omega}_1, \hat{\Omega}_2, \ldots, \hat{\Omega}_K) \), \( X = 1_K \), \( Z = I_K \), \( \gamma^T = (b_{01}, b_{02}, \ldots, b_{0K}) \), \( G = \sigma_0^2 I_K \) and \( R = \text{diag}(\sigma^2_{e,1}, \ldots, \sigma^2_{e,K}) \). As already mentioned, we only have one observation per level in model (7).

To be able to estimate \( \sigma_0^2 \), we first estimate \( \sigma^2_{e,1}, \ldots, \sigma^2_{e,K} \) as explained in
Section 3.3. In the PARMS statement of PROC MIXED, initial values for the covariance parameters can be specified. We choose an arbitrary initial value for $\sigma_0^2$. The initial values for the error variances are chosen to be $\hat{\sigma}_{e,1}^2, \ldots, \hat{\sigma}_{e,K}^2$. By using the option EQCONS, the initial residual variances will be held constant during the estimation procedure. Maximization of (2) over $\sigma_0^2$ gives an estimate for the heterogeneity $\sigma_0^2$.

For model (6), $\mathbf{y}^T = \left( \hat{\Omega}_{10}, \hat{\Omega}_{11}, \hat{\Omega}_{20}, \hat{\Omega}_{21}, \ldots, \hat{\Omega}_{K0}, \hat{\Omega}_{K1} \right)$, $\mathbf{X} = \left( \begin{array}{cc} 1 & -1 \\ 1 & 1 \end{array} \right) \otimes \mathbf{1}_K$, $\mathbf{Z} = \left( \begin{array}{cc} 1 & -1 \\ 1 & 1 \end{array} \right) \otimes \mathbf{I}_K$, $\mathbf{y}^T = \mathbf{b}^T$ and $\mathbf{G}$ is as defined in Section 3.1. Further, $\mathbf{R} = \text{diag}(\sigma_{e,10}^2, \sigma_{e,11}^2, \ldots, \sigma_{e,K0}^2, \sigma_{e,K1}^2)$. The error covariance matrix $\mathbf{R}$ can be estimated as explained in Section 3.3. By maximizing (2) over $\mathbf{G}$ in PROC MIXED while fixing the error variances as described above, we obtain estimates for $\sigma_0^2$, $\sigma_1^2$ and $\sigma_{01}$.

To obtain estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the mixed model equations are solved (Henderson, 1984). The solutions can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y}$ and $\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})$.

References
