An MCMHNR Algorithm

In this section, we use a MCMHNR method to approximate the conditional expectations given the observed data and current estimates taking into account the censoring information.

A random sample from the conditional joint density \( f(y^*, b_i | D_i, \theta) \), let \( f(.) \) be a generic notation denoting a density function, can be generated by using the Gibbs sampler method by iteratively sampling from the full conditionals \( f(y^* | b_i, D_i, \theta) \) and \( f(b_i | y^*, D_i, \theta) \) in turn until the resulting Markov chain converges. To sample these full conditionals, note that generation from \( f(y^* | b_i, D_i, \theta) \) is quite simple, basically consisting of an \( n_i \) dimensional vector with \( j \)-th component having mean \( \eta_{1i} \) and variance unity. To sample from \( f(b_i | y^*, D_i, \theta) \) multivariate rejection sampling method ([14]) may be used. Thus sampling from each of the full conditionals after a burn-in period, will ultimately provide a sample from \( f(y^*, b_i | D_i, \theta) \).

Alternatively, given a sample observation from \( f(y^*, b_i | D_i, \theta) \), we can generate a new value \( y^*_{i,j0} \) for the \( j \)-th element of \( y^*_{i0} \). We take

\[
(y^*_{0}, b) = (y^*_{11}, \ldots, y^*_{ij0} y^*_{i,j+1} \ldots y^*_{inn}, b_{11}, \ldots, b_{1q} \ldots, b_{ni1}, \ldots, b_{nq})
\]

by using the candidate distribution \( f(y^*_{0}, b, \theta) = f(y^*_{0}, b | \theta) f(b | \theta) \). In view of McCulloch [15], with probability

\[
\alpha_{ij}(y^*, b; y^*_{0}, b) = \min \left\{ 1, \frac{f(y^*_{0}, b | D_i, \theta) f(y^* | b_i | D_i, \theta)}{f(y^*_{0}, b | D_i, \theta) f(y^*_{i,j0} | b_i | D_i, \theta)} \right\}
\]

\( i = 1 \ldots n, j = 1 \ldots m \)  \hspace{1cm} (a.1)

we select the new sample \( y^*_{i,j0} \).

Again for generating new value \( b_{0} \) for the \( k \)-th element of \( b_i \)

we take \((y^*_{0}, b) = (y^*_{11}, \ldots, y^*_{nm}, b_{11}, \ldots, b_{1q} \ldots, b_{k0} \ldots, b_{nq})\) by using the candidate distribution
\[ f(y^*, b, \theta) \]

and the acceptance probability

\[ \delta_{ik}(y^*, b; y^*, b_0) = \min \left\{ 1, \frac{f_{y^*, b \mid D}(y^*, b_0 \mid D, \theta) f(y^*, b \mid \theta)}{f_{y^*, b \mid D}(y^*, b \mid D, \theta) f(y^*, b_0 \mid \theta)} \right\} \]

\[ i = 1 \ldots n, k = 1 \ldots q \]

(a.2)

Note that the equation in (a.1) can be simplified to

\[ f_{y_i \mid b \mid D}(y_i^*, b \mid D, \theta) f(y_i^*, b \mid \theta) = f_{D \mid y_i^* \mid b}(D \mid y_i^*, b, \beta, \alpha) \]

(a.3)

The ratio in (a.2) can also be simplified into a similar manner. This calculation of the acceptance functions \( \alpha_{ij} \) and \( \delta_{ij} \) essentially involve only the conditional distribution of data given the shared random effect and the latent random variables (i.e. the likelihood).

APPENDIX 1

**Detailed expressions require to compute score function**

The constant \( E(\psi_{1ci} \mid y_i^*, b_i) \) can be expressed as

\[ E(\psi_{1ci} \mid y_i^*, b_i) = -cP(r_i < -c) + r_i P(-c < r_i < c) + cP(r_i > c) \]

\[ = r_i (\Phi(c) - \Phi(-c)) \]

Also \( \frac{\delta}{\delta\beta_1} E(\psi_{1ci} \mid y_i^*, b_i) = \frac{\delta}{\delta\beta_1} \int \psi_{1ci} f_{y_i \mid b \mid D}(y_i^*, b_i, \theta) dy_i \]

\[ = E(\psi_{1ci}(r_i) \mid y_i^*, b_i) - E(\psi_{1ci}(r_i) \mid y_i^*, b_i) \]

where \( E(\psi_{1ci}'(r_i) \mid y_i^*, b_i) = (\Phi(c) - \Phi(-c)) \)

which implies

\[ \frac{\delta\mu_{1i}}{\delta\beta_1} = x_{1i} w_{1i} \left[ \psi_{1ci}(r_i) - (\Phi(c) - \Phi(-c)) + E(\psi_{1ci}(r_i)) \right] \]
Similarly for the parameters $\alpha_k$, we can compute

$$
\frac{\delta \mu_{1i}}{\delta \alpha_k} = \pi_i w_{1i}[\psi'(r_i) - (\Phi(c) - \Phi(-c)) + E(r_i \Psi_{xci}(r_i))]
$$

Similarly it can be shown that

$$
\frac{\delta \mu_{2i}(t_i)}{\delta \beta_2} = -\int_0^{t_i} \exp(\eta_{2i}(s))x_{2i}(s)^2d\Lambda(s) + \frac{\delta}{\delta \beta_2} E(\int_0^{t_i} \exp(\eta_{2i}(s))x_{2i}(s)d\Lambda(s))
$$

$$
= -\exp(\eta_{2i}(t_i))x_{2i}(t_i) + \frac{\delta}{\delta \Lambda(t_i)} E[\exp(\eta_{2i}(t_i))x_{2i}(t_i)]
$$

$$
= -\exp(\eta_{2i}(t_i))x_{2i}(t_i) + E[\exp(\eta_{2i}(t_i))] - E[\exp(\eta_{2i}(t_i)) \int_0^{t_i} \exp(\eta_{2i}(s))d\Lambda(s)]
$$

And

$$
\frac{\delta \mu_{3i}(t_i)}{\delta \Lambda(t_i)} = -\exp(\eta_{2i}(t_i))x_{2i}(t_i) + \frac{\delta}{\delta \Lambda(t_i)} E[\exp(\eta_{2i}(t_i))x_{2i}(t_i)]
$$

$$
= -\exp(\eta_{2i}(t_i))x_{2i}(t_i) + E[\exp(\eta_{2i}(t_i))] - E[\exp(\eta_{2i}(t_i)) \int_0^{t_i} \exp(\eta_{2i}(s))d\Lambda(s)]
$$

APPENDIX 2

The basic assumptions are as follows:

A.1: \{m_i\} is a bounded sequence of positive integers and the distinct values of $t_{ij}$ form a quasi-uniform sequence that grows dense on $[0, 1]$. Also $m_i \geq q \forall i$.

A.2: $|f^r(r(.))| < A_0$ for some non random value $A_0$ for $r \geq 2$.

A.3: For every $i$, $Max\{||X_{1i}||, ||X_{2i}||\} \leq B_0$ for some non random constant $B_0$.

A.4: Conditional on data, for every $i$,

$$
\sup_{t_i \geq 1} E[||u_{1ci}||^{2+\delta}] < \infty \text{ and }
$$

$$
\sup_{t_i \geq 1} E[||u_{2i}||^{2+\delta}] < \infty
$$

for some $\delta > 0$, where $u_{1ci} = w_{1i}(\psi_{1ci} - E(\psi_{1ci}))$, $u_{2i} = w_{2i}(v(\eta_{2i}) - E(v(\eta_{2i})))$ and $E$ denotes the expectation over the joint distribution of $y^*$ and $b$ given the data. In fact

$$
E_D E_{y^*, b_i | D}[u_{1ci} \psi_{1ci}'] = B_i \text{ and } E_D E_{y^*, b_i | D}[u_{2i}^2] = a_i
$$

with $\sup_{i \geq 1} \|B_i\| < \infty$ and $\sup_{i \geq 1} |a_i| < \infty$

A.5: The function $\psi_{1c}(\cdot)$ and $v(\cdot)$ are continuously differentiable. Further $E_{y^*, b_i | \psi_{1c}, i, j} | D$ and $E_{y^*, b_i | v_i, i} | D$ are dominated by uniformly $(r_1 + p_1 + p_2)$ integrable functions for $0 < \delta \leq 1$, $D$ stands for all data sets.

A.6: True parameter for $\theta$ denoted by $\theta_0 = (\beta_1', \alpha_0', \beta_2')$ satisfies $\|\beta_0\| \leq M_0$ for a known positive constant $M_0$.

Assumption A.1 essentially indicates that $n_0(= \sum_{i=1}^{n} m_i)$ and $n$ are of same order. This also means that we have only local dependence in the sample. The smoothness condition A.2 deals with the rate of convergence of the spline estimate $\hat{f} = \pi'(t)\hat{\alpha}$. Assumption A.3 is basically on the compact support for the covariates. Assumptions A.4-A.5 are two technical conditions required to justify consistency.

Note that for the semiparametric model (1), the covariates $x_{1ij}$ and $t_{ij}$ are dependent. One way to take account of this dependence is to proceed through the linear relationship ([16])

$$
x_{1ijv} = g_v(t_{ij}) + \epsilon_{ijv}, 1 \leq i \leq n, 1 \leq j \leq n, 1 \leq v \leq p
$$

where $g_v(.)$ are $p$ functions for each of which $r$-th derivative is bounded and $\epsilon_{ijv}$’s are independent random variables with mean 0. Further $\epsilon_{ijv}$’s are assumed to be independent of $\{e_{ij}\}$.
In view of the fact that $\alpha$’s are nuisance parameters, we modify the equation (8) as

$$
\hat{\beta} = \left( \begin{array}{c} \hat{\beta}_1 \\ \hat{\beta}_2 \end{array} \right) = \left( \begin{array}{c} \hat{\beta}_{10} \\ \hat{\beta}_{20} \end{array} \right) - |E[ \left( \begin{array}{cc} X_1^* & 0 \\ 0 & X_2 \end{array} \right) \left( \begin{array}{cc} W_1 & 0 \\ 0 & W_2 \end{array} \right) \left( \begin{array}{cc} B_{11} & B_{12} \\ B_{21} & B_{22} \end{array} \right) \times \left( \begin{array}{cc} W_1 & 0 \\ 0 & W_2 \end{array} \right) \left( \begin{array}{cc} X_1^* & 0 \\ 0 & X_2' \end{array} \right) |D\right]^{-1} \times E[ \left( \begin{array}{cc} X_1^* & 0 \\ 0 & X_2 \end{array} \right) \left( \begin{array}{cc} W_1 & 0 \\ 0 & W_2 \end{array} \right) \left( \begin{array}{cc} u_{1c|y^*,b} \\ u_{2|y^*,b} \end{array} \right) \left( \begin{array}{cc} X_1^* & 0 \\ 0 & X_2' \end{array} \right) |D\right] \right)
$$

(a.4)

where $X_1^* = (I - H)X$, $H = P(P'P)^{-1}P'$, $P = ((\pi_{ij}))$ are matrices of order $n_0 \times p_1, n_0 \times n_0$ and $n_0 \times N$ respectively. For the existence of Fisher information the following assumptions are required.

A.7:

$$(i) \lim_{n \to \infty} \frac{k_n}{n} (P'P) = Q \text{ and } (ii) \lim_{n \to \infty} R_n = R(> 0)$$

where $k_n$ is the number of knots and $R_n = (X_1^* X_1^*)$ and $Q$ and $R$ are positive definite matrices with eigenvalues all are bounded.

Assumption A.7(i) is a very standard property of B-spline basis functions and this holds true under general design conditions ([17]). A.7(ii) is a prerequisite for the existence of asymptotic dispersion of the proposed estimator.
Sketch of proofs of Theorems in Section 4

Proof of Theorem 1:

Condition (11) can be shown following Lemma 8 and 9 in [18]. In fact if the number of knots is approximately of order $O(n^{-\frac{1}{r+1}}), r \geq 2$, then it can be shown that

$$\frac{1}{n_0} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (\pi'(t_{ij}) \hat{\alpha} - f_0(t_{ij}))^2 = O_p(n^{-\frac{2}{r+1}})$$

which in view of Stone [18] can be expressed as

$$\int \{f(t) - f_0(t)\}^2 dt = O_p(n^{-\frac{2}{r+1}})$$

The proof of conditions (10) and (11) are rather straightforward application of Zeng and Cai [19]. Under Assumptions A.1-A.7, the following steps that justify consistency can be established.

(i) A solution $(\hat{\alpha}, \hat{\beta}, \hat{\Lambda})$ to equation (8) exists.

(ii) With probability 1, $\hat{\Lambda}(\tau)$ is bounded (as $n$ gets large) and hence there exists a subsequence such that $(\hat{\beta}, \hat{\alpha}, \hat{\Lambda})$ converges to some $(\beta^*, \alpha^*, \Lambda^*)$.

(iii) Finally,

$$\lim_{n \to \infty} (\beta^*, \Lambda^*) = (\beta_0, \Lambda_0)$$

Proof of Theorem 2:

The proof of the theorem can be established through the Taylor expansion of score equation for $\hat{\theta}$ and $\hat{\Lambda}$ around their true values $\theta_0$ and $\Lambda_0$. Using the fact that the information operator for $(\theta_0, \Lambda_0)$ is continuously invertible in an appropriate metric space. In fact this result can be derived by appealing to Theorem 3.3.1 in van der Vaart and Wellner [20], Appendix 1 in Parner [21] and verifying the conditions to suit our result subject to the Assumptions A.1-A.7.
(Specifically using the continuity of $\psi$ function and existence of integrals as assumed under A.5).