Supplements to "A Class of Hierarchical Multivariate Wiener Processes for Modeling Dependent Degradation Data"

Guanqi Fang[1,2] and Rong Pan[*3]

[1] Collaborative Innovation Center of Statistical Data Engineering Technology & Application, Zhejiang Gongshang University

[2] School of Statistics and Mathematics, Zhejiang Gongshang University

[3] School of Computing and Augmented Intelligence, Arizona State University

A Derivation of Equation (7)

To derive the conditional (posterior) distribution of $\boldsymbol{b}_i$ given $\tilde{\mathbf{y}}_i(t,0)$, we make use of the Bayes' rule. That is

$$
\begin{aligned}
& f\left(\boldsymbol{b}_i|\tilde{\mathbf{y}}_i(t,0)\right) \\
& \propto f\left(\tilde{\mathbf{y}}_i(t,0)|\boldsymbol{b}_i\right) f\left(\boldsymbol{b}_i\right) \\
& \propto \exp\left[-\frac{1}{2}\left(\tilde{\mathbf{y}}_i(t,0)-t\boldsymbol{b}_i\right)'\left(t\boldsymbol{D}\right)^{-1}\left(\tilde{\mathbf{y}}_i(t,0)-t\boldsymbol{b}_i\right)\right] \times \exp\left[-\frac{1}{2}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_b\right)'\boldsymbol{\Sigma}_b^{-1}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_b\right)\right] \\
& \propto \exp\left[-\frac{1}{2}(t\boldsymbol{b}_i)'(t\boldsymbol{D})^{-1}(t\boldsymbol{b}_i)+(t\boldsymbol{b}_i)'(t\boldsymbol{D})^{-1}\tilde{\mathbf{y}}_i(t,0)\right] \times \exp\left[-\frac{1}{2}\boldsymbol{b}_i'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i+\boldsymbol{b}_i'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b\right] \\
& = \exp\left[-\frac{1}{2}\boldsymbol{b}_i'\left(\boldsymbol{\Sigma}_b^{-1}+t\boldsymbol{D}^{-1}\right)\boldsymbol{b}_i+\boldsymbol{b}_i'\left(\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b+\boldsymbol{D}^{-1}\tilde{\mathbf{y}}_i(t,0)\right)\right].
\end{aligned}
$$

The derivation is analogous to finding the posterior mean of a MVN distribution when a semi-conjugate prior for the mean is given. For details, please refer to pages 107–108 of Hoff's book (2009).

B Derivation of Equation (8)

Similar to the above, the conditional (posterior) distribution $\boldsymbol{b}_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)$ is derived as

$$
\begin{aligned}
& f\left(\boldsymbol{b}_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)\right) \\
\propto & f\left(\boldsymbol{y}_i(0)|\boldsymbol{b}_i\right) f\left(\tilde{\mathbf{y}}_i(t,0)|\boldsymbol{b}_i\right) f\left(\boldsymbol{b}_i\right)
\end{aligned}
$$

$$\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_i(0)-\boldsymbol{\mu}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\right)'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\left(\boldsymbol{y}_i(0)-\boldsymbol{\mu}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\right)\right]\times$$

$$\exp\left[-\frac{1}{2}\left(\tilde{\mathbf{y}}_i(t,0)-t\boldsymbol{b}_i\right)'(t\boldsymbol{D})^{-1}\left(\tilde{\mathbf{y}}_i(t,0)-t\boldsymbol{b}_i\right)\right]\times$$

$$\exp\left[-\frac{1}{2}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_b\right)'\boldsymbol{\Sigma}_b^{-1}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_b\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\boldsymbol{b}_i'\left(\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_b^{-1}\right)\boldsymbol{b}_i+\right.$$

$$\left.\boldsymbol{b}_i'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\left(\boldsymbol{y}_i(0)-\boldsymbol{\mu}_a+\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b\right)\right]\times$$

$$\exp\left[-\frac{1}{2}(t\boldsymbol{b}_i)'(t\boldsymbol{D})^{-1}(t\boldsymbol{b}_i)+(t\boldsymbol{b}_i)'(t\boldsymbol{D})^{-1}\tilde{\mathbf{y}}_i(t,0)\right]\times$$

$$\exp\left[-\frac{1}{2}\boldsymbol{b}_i'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b}_i+\boldsymbol{b}_i'\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b\right]$$

$$=\exp\left\{-\frac{1}{2}\boldsymbol{b}_i'\left(\boldsymbol{\Sigma}_b^{-1}+t\boldsymbol{D}^{-1}+\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_b^{-1}\right)\boldsymbol{b}_i+\right.$$

$$\left.\boldsymbol{b}_i'\left[\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b+\boldsymbol{D}^{-1}\tilde{\mathbf{y}}_i(t,0)+\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}\left(\boldsymbol{y}_i(0)-\boldsymbol{\mu}_a+\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b\right)\right]\right\}.$$

C Derivation of the Conditional (Posterior) Distribution for Models M_3 and M_4

❶ For model M_3, similar to the derivation of Equation (8), if both $y_{ij}(0)$ and $\tilde{\mathrm{y}}_{ij}(t,0)$ are known, the conditional (posterior) distribution of b_{ij} is

$$f\left(b_{ij}|y_{ij}(0),\tilde{\mathrm{y}}_{ij}(t,0)\right)$$

$$\propto f\left(y_{ij}(0)|b_{ij}\right)f\left(\tilde{\mathrm{y}}_{ij}(t,0)|b_{ij}\right)f\left(b_{ij}\right)$$

$$\propto \exp\left[-\frac{1}{2\sigma^2_{Y_{ij}(0)|b_{ij}}}\left(y_{ij}(0)-\mu_{Y_{ij}(0)|b_{ij}}\right)^2\right]\times$$

$$\exp\left[-\frac{1}{2t\sigma_j^2}\left(\tilde{\mathrm{y}}_{ij}(t,0)-tb_{ij}\right)^2\right]\times\exp\left[-\frac{1}{2\sigma^2_{bj}}\left(b_{ij}-\mu_{bj}\right)^2\right]$$

$$\propto \exp\left\{-\frac{(\rho_j^{ab})^2}{2\sigma^2_{bj}\left[1-(\rho_j^{ab})^2\right]}b_{ij}^2+\frac{\rho_j^{ab}\left(y_{ij}(0)-\mu_{aj}+\frac{\sigma_{aj}}{\sigma_{bj}}\rho_j^{ab}\mu_{bj}\right)}{\left[1-(\rho_j^{ab})^2\right]\sigma_{aj}\sigma_{bj}}b_{ij}\right\}\times$$

$$\exp\left[-\frac{t}{2\sigma_j^2}b_{ij}^2+\frac{\tilde{\mathsf{y}}_{ij}(t,0)}{\sigma_j^2}b_{ij}\right]\times\exp\left[-\frac{1}{2\sigma_{bj}^2}b_{ij}^2+\frac{\mu_{bj}}{\sigma_{bj}^2}b_{ij}\right]$$

$$\propto\exp\Bigg\{-\frac{1}{2}\left[\frac{1}{\sigma_{bj}^2}+\frac{t}{\sigma_j^2}+\frac{(\rho_j^{ab})^2}{\sigma_{bj}^2\left[1-(\rho_j^{ab})^2\right]}\right]b_{ij}^2+$$

$$\left[\frac{\mu_{bj}}{\sigma_{bj}^2}+\frac{\tilde{\mathsf{y}}_{ij}(t,0)}{\sigma_j^2}+\frac{\rho_j^{ab}\left(y_{ij}(0)-\mu_{aj}+\frac{\sigma_{aj}}{\sigma_{bj}}\rho_j^{ab}\mu_{bj}\right)}{\left[1-(\rho_j^{ab})^2\right]\sigma_{aj}\sigma_{bj}}\right]b_{ij}\Bigg\}.$$

Thus, this suggests that the distribution of b_{ij} given $y_{ij}(0)$ and $\tilde{\mathsf{y}}_{ij}(t,0)$ is

$$b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)\sim\mathcal{N}\left(\mu_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)},\sigma^2_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)}\right),$$

where

$$\mu_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)}=\left[\frac{1}{\sigma_{bj}^2}+\frac{t}{\sigma_j^2}+\frac{(\rho_j^{ab})^2}{\sigma_{bj}^2\left[1-(\rho_j^{ab})^2\right]}\right]^{-1}\times$$

$$\left[\frac{\mu_{bj}}{\sigma_{bj}^2}+\frac{\tilde{\mathsf{y}}_{ij}(t,0)}{\sigma_j^2}+\frac{\rho_j^{ab}\left(y_{ij}(0)-\mu_{aj}+\frac{\sigma_{aj}}{\sigma_{bj}}\rho_j^{ab}\mu_{bj}\right)}{\left[1-(\rho_j^{ab})^2\right]\sigma_{aj}\sigma_{bj}}\right]\text{ and}$$

$$\sigma^2_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)}=\left[\frac{1}{\sigma_{bj}^2}+\frac{t}{\sigma_j^2}+\frac{(\rho_j^{ab})^2}{\sigma_{bj}^2\left[1-(\rho_j^{ab})^2\right]}\right]^{-1}.$$

It is noted that if $y_{ij}(0)$ and b_{ij} are independent, $b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)$ degenerates to $b_{ij}|\tilde{\mathsf{y}}_{ij}(t,0)$. Its distribution can be found by setting $\rho_j^{ab}=0$; that is

$$b_{ij}|\tilde{\mathsf{y}}_{ij}(t,0)\sim\mathcal{N}\left(\mu_{b_{ij}|\tilde{\mathsf{y}}_{ij}(t,0)},\sigma^2_{b_{ij}|\tilde{\mathsf{y}}_{ij}(t,0)}\right),$$

where

$$\mu_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)}=\left[\frac{1}{\sigma_{bj}^2}+\frac{t}{\sigma_j^2}\right]^{-1}\left[\frac{\mu_{bj}}{\sigma_{bj}^2}+\frac{\tilde{\mathsf{y}}_{ij}(t,0)}{\sigma_j^2}\right]\text{ and}$$

$$\sigma^2_{b_{ij}|y_{ij}(0),\tilde{\mathsf{y}}_{ij}(t,0)}=\left[\frac{1}{\sigma_{bj}^2}+\frac{t}{\sigma_j^2}\right]^{-1}.$$

This result matches the conclusion regarding the distribution of b_{ij} given $\tilde{\mathsf{y}}_{ij}(t,0)$ as presented by Equation (14) in Li et al. (2015).

❷ For model M_4, if both $\boldsymbol{y}_i(0)$ and $\tilde{\mathbf{y}}_i(t,0)$ are known, the conditional (posterior) distribution of b_i is

$$
\begin{aligned}
& f\left(b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)\right) \\
\propto & f\left(\boldsymbol{y}_i(0)|b_i\right) f\left(\tilde{\mathbf{y}}_i(t,0)|b_i\right) f\left(b_i\right) \\
\propto & \exp\left[-\frac{1}{2}\sum_{j=1}^{p}\frac{\left(y_{ij}(0)-\mu_{aj}\mu_{a_i|b_i}\right)^2}{\mu_{aj}^2\sigma_{a_i|b_i}^2}\right]\times\exp\left[-\frac{1}{2}\sum_{j=1}^{p}\frac{\left(\tilde{\mathrm{y}}_{ij}(t,0)-t\mu_{bj}b_i\right)^2}{t\sigma_j^2}\right]\times\exp\left[-\frac{\left(b_i-1\right)^2}{2\sigma_b^2}\right] \\
\propto & \exp\left\{-\frac{b_i^2}{2}\frac{p\rho_{ab}^2}{\sigma_b^2\left(1-\rho_{ab}^2\right)}+b_i\frac{\rho_{ab}}{\sigma_a\sigma_b\left(1-\rho_{ab}^2\right)}\sum_{j=1}^{p}\left(\frac{y_{ij}(0)}{\mu_{aj}}-1+\rho_{ab}\frac{\sigma_a}{\sigma_b}\right)\right\}\times \\
& \exp\left[-\frac{b_i^2}{2}\sum_{j=1}^{p}\frac{t\mu_{bj}^2}{\sigma_j^2}+b_i\sum_{j=1}^{p}\frac{\mu_{bj}\tilde{\mathrm{y}}_{ij}(t,0)}{\sigma_j^2}\right]\times\exp\left[-\frac{b_i^2}{2\sigma_b^2}+\frac{b_i}{\sigma_b^2}\right] \\
\propto & \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_b^2}+\sum_{j=1}^{p}\frac{t\mu_{bj}^2}{\sigma_j^2}+\frac{p\rho_{ab}^2}{\sigma_b^2\left(1-\rho_{ab}^2\right)}\right]b_i^2+\right. \\
& \left.\left[\frac{1}{\sigma_b^2}+\sum_{j=1}^{p}\frac{\mu_{bj}\tilde{\mathrm{y}}_{ij}(t,0)}{\sigma_j^2}+\frac{\rho_{ab}}{\sigma_a\sigma_b\left(1-\rho_{ab}^2\right)}\sum_{j=1}^{p}\left(\frac{y_{ij}(0)}{\mu_{aj}}-1+\rho_{ab}\frac{\sigma_a}{\sigma_b}\right)\right]b_i\right\}.
\end{aligned}
$$

Thus, this suggests that the distribution of b_i given $\boldsymbol{y}_i(0)$ and $\tilde{\mathbf{y}}_i(t,0)$ is

$$
b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)\sim\mathcal{N}\left(\mu_{b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)},\sigma^2_{b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)}\right),
$$

where

$$
\begin{aligned}
\mu_{b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)} = & \left[\frac{1}{\sigma_b^2}+\sum_{j=1}^{p}\frac{t\mu_{bj}^2}{\sigma_j^2}+\frac{p\rho_{ab}^2}{\sigma_b^2\left(1-\rho_{ab}^2\right)}\right]^{-1}\times \\
& \left[\frac{1}{\sigma_b^2}+\sum_{j=1}^{p}\frac{\mu_{bj}\tilde{\mathrm{y}}_{ij}(t,0)}{\sigma_j^2}+\frac{\rho_{ab}}{\sigma_a\sigma_b\left(1-\rho_{ab}^2\right)}\sum_{j=1}^{p}\left(\frac{y_{ij}(0)}{\mu_{aj}}-1+\rho_{ab}\frac{\sigma_a}{\sigma_b}\right)\right] \text{ and} \\
\sigma^2_{b_i|\boldsymbol{y}_i(0),\tilde{\mathbf{y}}_i(t,0)} = & \left[\frac{1}{\sigma_b^2}+\sum_{j=1}^{p}\frac{t\mu_{bj}^2}{\sigma_j^2}+\frac{p\rho_{ab}^2}{\sigma_b^2\left(1-\rho_{ab}^2\right)}\right]^{-1}.
\end{aligned}
$$

D The EM Algorithm for Statistical Inference

D.1 Main Technical Details

To implement the EM algorithm, a useful lemma (shown as Lemma 3 below) indicated on page 107 of Rencher and Schaalje's book (2008) is first presented.

Lemma 3. If $\boldsymbol{x}$ is a random vector with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ and if $\boldsymbol{A}$ is a symmetric matrix of constants, then $E_{\boldsymbol{x}}\left[\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\right] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$.

Recall that $\{\mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\}$ denotes the complete data, and the total log-likelihood function (up to a constant) is given by

$$\ell\left(\boldsymbol{\theta}; \mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\right) = \ell\left(\boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b}\right) + \ell\left(\boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) + \ell\left(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a; \mathbb{Y}_0\right),$$

where

$$\ell\left(\boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b}\right) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathrm{t}}_{ik}\sigma_j^2\right) + \frac{\left(\tilde{\mathrm{y}}_{ijk} - \tilde{\mathrm{t}}_{ik}b_{ij}\right)^2}{\tilde{\mathrm{t}}_{ik}\sigma_j^2}\right],$$

$$\ell\left(\boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left|\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}\right| + \left(\boldsymbol{b}_i - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}^{-1}\left(\boldsymbol{b}_i - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}\right)\right], \text{ and}$$

$$\ell\left(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a; \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left|\boldsymbol{\Sigma}_a\right| + \left(\boldsymbol{y}_{i0} - \boldsymbol{\mu}_a\right)'\boldsymbol{\Sigma}_a^{-1}\left(\boldsymbol{y}_{i0} - \boldsymbol{\mu}_a\right)\right].$$

Note that once $\mathbb{Y}_0$ is known, it is immediate to get the estimates of $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$:

$$\hat{\boldsymbol{\mu}}_a = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_{i0} \text{ and } \hat{\boldsymbol{\Sigma}}_a = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a\right)\left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a\right)'.$$

So the algorithm only needs to deal with $\ell\left(\boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b}\right)$ and $\ell\left(\boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right)$ iteratively. Given the current EM estimate $\boldsymbol{\theta}^{(s)}$ consisting of $\hat{\boldsymbol{\mu}}_a, \boldsymbol{\mu}_b^{(s)}, \hat{\boldsymbol{\Sigma}}_a, \boldsymbol{\Sigma}_b^{(s)}, \boldsymbol{\Sigma}_{ab}^{(s)}$, and $\boldsymbol{D}^{(s)}$, the algorithm is performed according to the following steps:

- E-step: First, based on Equation (8), $\boldsymbol{b}|\mathbb{Y}_0, \mathbb{D}, \boldsymbol{\theta}^{(s)}$ or concretely $\boldsymbol{b}_i|\boldsymbol{y}_{i0}, \tilde{\mathbf{y}}_i, \tilde{\mathbf{t}}_i, \boldsymbol{\theta}^{(s)}$ is subject to a MVN distribution with mean vector

$$\begin{aligned}\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} &\equiv \left(\check{\mu}_{\boldsymbol{b}_i,1}^{(s)}, \check{\mu}_{\boldsymbol{b}_i,2}^{(s)}, \ldots, \check{\mu}_{\boldsymbol{b}_i,p}^{(s)}\right)' \\ &= \left(\boldsymbol{\Sigma}_b^{(s)^{-1}} + t_{im_i}\boldsymbol{D}^{(s)^{-1}} + \boldsymbol{\Sigma}_b^{(s)^{-1}}\boldsymbol{\Sigma}_{ba}^{(s)}\boldsymbol{\Sigma}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}^{(s)^{-1}}\boldsymbol{\Sigma}_{ab}^{(s)}\boldsymbol{\Sigma}_b^{(s)^{-1}}\right)^{-1}\left[\boldsymbol{\Sigma}_b^{(s)^{-1}}\boldsymbol{\mu}_b^{(s)} + \right. \\ &\quad \left. \boldsymbol{D}^{(s)^{-1}}\left(\boldsymbol{y}_{im_i} - \boldsymbol{y}_{i0}\right) + \boldsymbol{\Sigma}_b^{(s)^{-1}}\boldsymbol{\Sigma}_{ba}^{(s)}\boldsymbol{\Sigma}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}^{(s)^{-1}}\left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a + \boldsymbol{\Sigma}_{ab}^{(s)}\boldsymbol{\Sigma}_b^{(s)^{-1}}\boldsymbol{\mu}_b^{(s)}\right)\right]\end{aligned}$$

and variance-covariance matrix

$$\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)} \equiv \begin{pmatrix} \check{\sigma}_{\boldsymbol{b}_i,1}^{2(s)} & \check{\sigma}_{\boldsymbol{b}_i,12}^{(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,1p}^{(s)} \\ \check{\sigma}_{\boldsymbol{b}_i,12}^{(s)} & \check{\sigma}_{\boldsymbol{b}_i,2}^{2(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,2p}^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ \check{\sigma}_{\boldsymbol{b}_i,1p}^{(s)} & \check{\sigma}_{\boldsymbol{b}_i,2p}^{(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,p}^{2(s)} \end{pmatrix}$$

$$= \left(\boldsymbol{\Sigma}_b^{(s)^{-1}} + t_{im_i} \boldsymbol{D}^{(s)^{-1}} + \boldsymbol{\Sigma}_b^{(s)^{-1}} \boldsymbol{\Sigma}_{ba}^{(s)} \boldsymbol{\Sigma}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}^{(s)^{-1}} \boldsymbol{\Sigma}_{ab}^{(s)} \boldsymbol{\Sigma}_b^{(s)^{-1}} \right)^{-1},$$

where $\boldsymbol{y}_{im_i} - \boldsymbol{y}_{i0} = \sum_{k=1}^{m_i} \tilde{\mathbf{y}}_{ik}$, $t_{im_i} = \sum_{k=1}^{m_i} \tilde{\mathrm{t}}_{ik}$, and $\boldsymbol{\Sigma}_{\boldsymbol{Y}_i(0)|\boldsymbol{b}_i}^{(s)} = \hat{\boldsymbol{\Sigma}}_a - \boldsymbol{\Sigma}_{ab}^{(s)} \boldsymbol{\Sigma}_b^{(s)^{-1}} \boldsymbol{\Sigma}_{ba}^{(s)}$. Here, the conclusion implied by Equation (8) is generalized to the case of multiple measurements. Obviously, due to the Markovian property, only the up-to-date information (i.e., $\boldsymbol{y}_{i0}$, $\boldsymbol{y}_{im_i} - \boldsymbol{y}_{i0}$, and t_{im_i}) is needed to calculate $\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}$ and $\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}$. Then, given Lemma 3, we know

$$\mathrm{E}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}} \left[b_{ij} \right] = \check{\mu}_{\boldsymbol{b}_i,j}^{(s)}, \ \mathrm{E}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}} \left[b_{ij}^2 \right] = \left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)} \right)^2 + \check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)}, \text{ and}$$

$$\mathrm{E}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}} \left[\left(\boldsymbol{b}_i - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}} \right)' \boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}^{-1} \left(\boldsymbol{b}_i - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}} \right) \right] = \mathrm{tr} \left(\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}^{-1} \check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)} \right) + \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}} \right)' \boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}^{-1} \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}} \right).$$

Therefore, $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ is shown as

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right) = & -\frac{1}{2} \Bigg\{ \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{m_i} \left[\ln\left(\tilde{\mathrm{t}}_{ik} \sigma_j^2 \right) + \frac{ \left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)} \right)^2 + \check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)} - 2\tilde{\mathrm{y}}_{ijk} \check{\mu}_{\boldsymbol{b}_i,j}^{(s)} / \tilde{\mathrm{t}}_{ik} + \tilde{\mathrm{y}}_{ijk}^2 / \tilde{\mathrm{t}}_{ik}^2 }{ \sigma_j^2 / \tilde{\mathrm{t}}_{ik} } \right] + \\ & \sum_{i=1}^{n} \Bigg[\ln \left| \boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \boldsymbol{\Sigma}_{ab} \right| + \mathrm{tr}\left((\boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \boldsymbol{\Sigma}_{ab})^{-1} \check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)} \right) + \\ & \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a \right) \right)' (\boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \boldsymbol{\Sigma}_{ab})^{-1} \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a \right) \right) \Bigg] + \\ & \sum_{i=1}^{n} \left[\ln \left| \hat{\boldsymbol{\Sigma}}_a \right| + \left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a \right)' \hat{\boldsymbol{\Sigma}}_a^{-1} \left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a \right) \right] \Bigg\}. \end{aligned}$$

- M-step: In this step, we update the parameter estimates by solving the equations that set the first derivative of $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ equal to 0. Particularly, the derivative with respect to $\boldsymbol{\Sigma}_{ab}$ results in a complex cubic matrix-wise equation that is prohibitively hard to solve. To reduce the complexity and also facilitate the computation, $\boldsymbol{\Sigma}_{\boldsymbol{b}_i|\boldsymbol{y}_{i0}}$ (i.e., $\boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{ba} \hat{\boldsymbol{\Sigma}}_a^{-1} \boldsymbol{\Sigma}_{ab}$) is considered as a fixed part when updating $\boldsymbol{\Sigma}_{ab}^{(s+1)}$. Then, $\boldsymbol{\Sigma}_b^{(s+1)}$ can be easily updated given the known $\boldsymbol{\Sigma}_{ab}^{(s+1)}$. After some manipulations based on matrix calculus, the result is given by

$$
\begin{aligned}
\boldsymbol{\mu}_b^{(s+1)} =& \frac{1}{n}\sum_{i=1}^{n}\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)},\\
\boldsymbol{\Sigma}_{ab}^{(s+1)} =& \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{y}_{i0}-\hat{\boldsymbol{\mu}}_a\right)\left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}-\boldsymbol{\mu}_b^{(s+1)}\right)',\\
\boldsymbol{\Sigma}_b^{(s+1)} =& \boldsymbol{\Sigma}_{ba}^{(s+1)}\hat{\boldsymbol{\Sigma}}_a^{-1}\boldsymbol{\Sigma}_{ab}^{(s+1)}+\frac{1}{n}\sum_{i=1}^{n}\left\{\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}+\right.\\
&\left.\left[\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}-\boldsymbol{\mu}_b^{(s+1)}-\boldsymbol{\Sigma}_{ba}^{(s+1)}\hat{\boldsymbol{\Sigma}}_a^{-1}\left(\boldsymbol{y}_{i0}-\hat{\boldsymbol{\mu}}_a\right)\right]\left[\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}-\boldsymbol{\mu}_b^{(s+1)}-\boldsymbol{\Sigma}_{ba}^{(s+1)}\hat{\boldsymbol{\Sigma}}_a^{-1}\left(\boldsymbol{y}_{i0}-\hat{\boldsymbol{\mu}}_a\right)\right]'\right\},\ \text{and}\\
\boldsymbol{D}^{(s+1)} =& \operatorname{diag}\left(\sigma_1^{2^{(s+1)}},\sigma_2^{2^{(s+1)}},\ldots,\sigma_p^{2^{(s+1)}}\right),\\
\text{where } \sigma_j^{2^{(s+1)}} =& \frac{\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\tilde{\mathrm{t}}_{ik}\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2+\tilde{\mathrm{t}}_{ik}\check{\sigma}_{\boldsymbol{b}_i,j}^{2^{(s)}}-2\tilde{\mathrm{y}}_{ijk}\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}+\tilde{\mathrm{y}}_{ijk}^2/\tilde{\mathrm{t}}_{ik}\right]}{\sum_{i=1}^{n}m_i}.
\end{aligned}
$$

D.2 Inference Procedure Adapted to the Special Model Variants

❶ Firstly, recall that models M_1 and M_2 have the structures of random effects as follows:

$$
M_1:\begin{pmatrix}\boldsymbol{Y}_i(0)\\ \boldsymbol{b}_i\end{pmatrix}\sim\mathcal{MVN}\left(\begin{bmatrix}\boldsymbol{\mu}_a\\ \boldsymbol{\mu}_b\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_a & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_b\end{bmatrix}\right),
$$

$$
M_2:\boldsymbol{b}_i\sim\mathcal{MVN}(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b).
$$

One can see that the two models are very alike, except model M_2 excludes the random effect of the initial degradation levels. So we provide the inference procedure for model M_1 as an exemplar, while the one for M_2 can be found analogously. Since model M_1 assumes $\boldsymbol{\Sigma}_{ab}$ is a null matrix, the total log-likelihood function for the complete data reduces to

$$
\ell\left(\boldsymbol{\theta};\mathbb{Y}_0,\mathbb{D},\boldsymbol{b}\right)=\ell\left(\boldsymbol{D};\mathbb{D}\mid\boldsymbol{b}\right)+\ell\left(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b;\boldsymbol{b}\right)+\ell\left(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a;\mathbb{Y}_0\right),
$$

where

$$
\ell\left(\boldsymbol{D};\mathbb{D}\mid\boldsymbol{b}\right)=-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathrm{t}}_{ik}\sigma_j^2\right)+\frac{\left(\tilde{\mathrm{y}}_{ijk}-\tilde{\mathrm{t}}_{ik}b_{ij}\right)^2}{\tilde{\mathrm{t}}_{ik}\sigma_j^2}\right],
$$

$$
\ell\left(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b;\boldsymbol{b}\right)=-\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left|\boldsymbol{\Sigma}_{\boldsymbol{b}}\right|+\left(\boldsymbol{b}_i-\boldsymbol{\mu}_{\boldsymbol{b}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{b}}^{-1}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_{\boldsymbol{b}}\right)\right],\ \text{and}
$$

$$
\ell\left(\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_a;\mathbb{Y}_0\right)=-\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left|\boldsymbol{\Sigma}_a\right|+\left(\boldsymbol{y}_{i0}-\boldsymbol{\mu}_a\right)'\boldsymbol{\Sigma}_a^{-1}\left(\boldsymbol{y}_{i0}-\boldsymbol{\mu}_a\right)\right].
$$

Again, it is immediate to get the estimates of $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$, which are given by

$$\hat{\boldsymbol{\mu}}_a = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_{i0} \text{ and } \hat{\boldsymbol{\Sigma}}_a = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{y}_{i0}-\hat{\boldsymbol{\mu}}_a\right)\left(\boldsymbol{y}_{i0}-\hat{\boldsymbol{\mu}}_a\right)'.$$

So the algorithm only needs to apply the iterative computation with respect to $\ell\left(\boldsymbol{D};\mathbb{D}\mid\boldsymbol{b}\right)$ and $\ell\left(\boldsymbol{\mu}_b,\boldsymbol{\Sigma}_b;\boldsymbol{b}\right)$. Given the current EM estimate $\boldsymbol{\theta}^{(s)}$ consisting of $\hat{\boldsymbol{\mu}}_a,\boldsymbol{\mu}_b^{(s)},\hat{\boldsymbol{\Sigma}}_a,\boldsymbol{\Sigma}_b^{(s)}$, and $\boldsymbol{D}^{(s)}$, the algorithm is performed according to the following steps:

- E-step: First, based on Equation (7), $\boldsymbol{b}|\mathbb{D},\boldsymbol{\theta}^{(s)}$ or concretely $\boldsymbol{b}_i|\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}$ is subject to a MVN distribution with mean vector

$$\begin{aligned}\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} &\equiv \left(\check{\mu}_{\boldsymbol{b}_i,1}^{(s)},\check{\mu}_{\boldsymbol{b}_i,2}^{(s)},\ldots,\check{\mu}_{\boldsymbol{b}_i,p}^{(s)}\right)' \\ &= \left(\boldsymbol{\Sigma}_b^{(s)^{-1}}+t_{im_i}\boldsymbol{D}^{(s)^{-1}}\right)^{-1}\left(\boldsymbol{\Sigma}_b^{(s)^{-1}}\boldsymbol{\mu}_b^{(s)}+\boldsymbol{D}^{(s)^{-1}}(\boldsymbol{y}_{im_i}-\boldsymbol{y}_{i0})\right)\end{aligned}$$

  and variance-covariance matrix

$$\begin{aligned}\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)} &\equiv \begin{pmatrix}\check{\sigma}_{\boldsymbol{b}_i,1}^{2(s)} & \check{\sigma}_{\boldsymbol{b}_i,12}^{(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,1p}^{(s)} \\ \check{\sigma}_{\boldsymbol{b}_i,12}^{(s)} & \check{\sigma}_{\boldsymbol{b}_i,2}^{2(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,2p}^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ \check{\sigma}_{\boldsymbol{b}_i,1p}^{(s)} & \check{\sigma}_{\boldsymbol{b}_i,2p}^{(s)} & \cdots & \check{\sigma}_{\boldsymbol{b}_i,p}^{2(s)}\end{pmatrix} \\ &= \left(\boldsymbol{\Sigma}_b^{(s)^{-1}}+t_{im_i}\boldsymbol{D}^{(s)^{-1}}\right)^{-1},\end{aligned}$$

  where $\boldsymbol{y}_{im_i}-\boldsymbol{y}_{i0}=\sum_{k=1}^{m_i}\tilde{\mathbf{y}}_{ik}$ and $t_{im_i}=\sum_{k=1}^{m_i}\tilde{\mathbf{t}}_{ik}$. Here, the conclusion implied by Equation (7) is generalized to the case that multiple measurements are taken. Obviously, due to the Markovian property, only the up-to-date information is needed to calculate $\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}$ and $\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}$. Then, given Lemma 3, we know

$$\mathrm{E}_{\boldsymbol{b}_i|\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}\left[b_{ij}\right]=\check{\mu}_{\boldsymbol{b}_i,j}^{(s)},\ \mathrm{E}_{\boldsymbol{b}_i|\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}\left[b_{ij}^2\right]=\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2+\check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)},\text{ and}$$

$$\mathrm{E}_{\boldsymbol{b}_i|\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}\left[\left(\boldsymbol{b}_i-\boldsymbol{\mu}_{\boldsymbol{b}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{b}}^{-1}\left(\boldsymbol{b}_i-\boldsymbol{\mu}_{\boldsymbol{b}}\right)\right]=\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{b}}^{-1}\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}\right)+\left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}-\boldsymbol{\mu}_{\boldsymbol{b}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{b}}^{-1}\left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}-\boldsymbol{\mu}_{\boldsymbol{b}}\right).$$

Therefore, $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ is shown as

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right) = -\frac{1}{2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathsf{t}}_{ik}\sigma_j^2\right) + \frac{\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2 + \check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)} - 2\tilde{\mathsf{y}}_{ijk}\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}/\tilde{\mathsf{t}}_{ik} + \tilde{\mathsf{y}}_{ijk}^2/\tilde{\mathsf{t}}_{ik}^2}{\sigma_j^2/\tilde{\mathsf{t}}_{ik}}\right] + \right.$$
$$\sum_{i=1}^{n}\left[\ln|\boldsymbol{\Sigma}_b| + \mathrm{tr}\left(\boldsymbol{\Sigma}_b^{-1}\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}\right) + \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b\right)'\boldsymbol{\Sigma}_b^{-1}\left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b\right)\right] +$$
$$\left.\sum_{i=1}^{n}\left[\ln\left|\hat{\boldsymbol{\Sigma}}_a\right| + (\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a)'\hat{\boldsymbol{\Sigma}}_a^{-1}(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a)\right]\right\}.$$

- M-step: In this step, we update the estimated parameters by solving the equations that set the first derivative of $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ equal to 0. And based on some existing properties of matrix algebra, the result is given by

$$\boldsymbol{\mu}_b^{(s+1)} = \frac{1}{n}\sum_{i=1}^{n}\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)},$$
$$\boldsymbol{\Sigma}_b^{(s+1)} = \frac{1}{n}\sum_{i=1}^{n}\left[\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)} + \left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b^{(s+1)}\right)\left(\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)} - \boldsymbol{\mu}_b^{(s+1)}\right)'\right], \text{ and}$$
$$\boldsymbol{D}^{(s+1)} = \mathrm{diag}\left(\sigma_1^{2(s+1)}, \sigma_2^{2(s+1)}, \ldots, \sigma_p^{2(s+1)}\right),$$

where
$$\sigma_j^{2(s+1)} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\tilde{\mathsf{t}}_{ik}\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2 + \tilde{\mathsf{t}}_{ik}\check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)} - 2\tilde{\mathsf{y}}_{ijk}\check{\mu}_{\boldsymbol{b}_i,j}^{(s)} + \tilde{\mathsf{y}}_{ijk}^2/\tilde{\mathsf{t}}_{ik}\right]}{\sum_{i=1}^{n} m_i}.$$

❷ Now, recall that model M_3 is given by

$$M_3 : \begin{cases} Y_{ij}(t) = Y_{ij}(0) + b_{ij}t + \sigma_j\mathcal{B}(t) \\ \begin{pmatrix} Y_{ij}(0) \\ b_{ij} \end{pmatrix} \sim \mathcal{BVN}\left(\begin{bmatrix} \mu_{aj} \\ \mu_{bj} \end{bmatrix}, \begin{bmatrix} \sigma_{aj}^2 & \rho_j^{ab}\sigma_{aj}\sigma_{bj} \\ \rho_j^{ab}\sigma_{aj}\sigma_{bj} & \sigma_{bj}^2 \end{bmatrix}\right) \end{cases}.$$

Since PC-wise dependency is absent, we only need to demonstrate the inference procedure for marginal processes; and it is common for all individuals. For notational convenience, we drop the subscript j. Then, the set of the unknown parameters is $\boldsymbol{\theta} = (\mu_a, \mu_b, \sigma_a^2, \sigma_b^2, \rho_{ab}, \sigma^2)'$. And the complete data becomes $\{\mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\}$, where $\mathbb{Y}_0 = (y_{10}, y_{20}, \ldots, y_{n0})'$, $\mathbb{D} = \{\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \ldots, \tilde{\mathbf{y}}_n, \tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2, \ldots, \tilde{\mathbf{t}}_n\}$, $\tilde{\mathbf{y}}_i = (\tilde{\mathsf{y}}_{i1}, \tilde{\mathsf{y}}_{i2}, \ldots, \tilde{\mathsf{y}}_{im_i})'$, and $\boldsymbol{b} =$

$(b_1, b_2, \ldots, b_n)'$. Under this scenario, the total log-likelihood function is given by

$$\ell\left(\boldsymbol{\theta}; \mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\right) = \ell\left(\sigma^2; \mathbb{D} \mid \boldsymbol{b}\right) + \ell\left(\mu_a, \mu_b, \sigma_a^2, \sigma_b^2, \rho_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) + \ell\left(\mu_a, \sigma_a^2; \mathbb{Y}_0\right),$$

where

$$\ell\left(\sigma^2; \mathbb{D} \mid \boldsymbol{b}\right) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathrm{t}}_{ik}\sigma^2\right) + \frac{(\tilde{\mathrm{y}}_{ik} - \tilde{\mathrm{t}}_{ik}b_i)^2}{\tilde{\mathrm{t}}_{ik}\sigma^2}\right],$$

$$\ell\left(\mu_a, \mu_b, \sigma_a^2, \sigma_b^2, \rho_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left[\left(1-\rho_{ab}^2\right)\sigma_b^2\right] + \frac{\left[b_i - \mu_b - \frac{\sigma_b}{\sigma_a}\rho_{ab}(y_{i0} - \mu_a)\right]^2}{(1-\rho_{ab}^2)\,\sigma_b^2}\right], \text{ and}$$

$$\ell\left(\mu_a, \sigma_a^2; \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\ln\left(\sigma_a^2\right) + \frac{(y_{i0} - \mu_a)^2}{\sigma_a^2}\right].$$

Analogously, it is immediate to get the estimates of μ_a and σ_a^2, which are given by

$$\hat{\mu}_a = \frac{1}{n}\sum_{i=1}^{n} y_{i0} \text{ and } \hat{\sigma}_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{i0} - \hat{\mu}_a\right)^2.$$

So the algorithm only needs to deal with $\ell\left(\sigma^2; \mathbb{D} \mid \boldsymbol{b}\right)$ and $\ell\left(\mu_a, \mu_b, \sigma_a^2, \sigma_b^2, \rho_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right)$ iteratively. Given the current EM estimate $\boldsymbol{\theta}^{(s)}$, the algorithm is performed as follows:

- E-step: First, based on the conclusion provided by Supplementary Section C, $b_i | y_{i0}, \tilde{\mathbf{y}}_i, \tilde{\mathbf{t}}_i, \boldsymbol{\theta}^{(s)}$ is subject to a normal distribution with mean

$$\check{\mu}_{b_i}^{(s)} = \left[\frac{1}{\sigma_b^{2(s)}} + \frac{t_{im_i}}{\sigma^{2(s)}} + \frac{\left(\rho_{ab}^{(s)}\right)^2}{\sigma_b^{2(s)}\left[1 - \left(\rho_{ab}^{(s)}\right)^2\right]}\right]^{-1} \times$$
$$\left[\frac{\mu_b^{(s)}}{\sigma_b^{2(s)}} + \frac{y_{im_i} - y_{i0}}{\sigma^{2(s)}} + \frac{\rho_{ab}^{(s)}\left(y_{i0} - \hat{\mu}_a + \frac{\hat{\sigma}_a}{\sigma_b^{(s)}}\rho_{ab}^{(s)}\mu_b^{(s)}\right)}{\left[1 - \left(\rho_{ab}^{(s)}\right)^2\right]\hat{\sigma}_a\sigma_b^{(s)}}\right]$$

  and variance

$$\check{\sigma}_{b_i}^{2(s)} = \left[\frac{1}{\sigma_b^{2(s)}} + \frac{t_{im_i}}{\sigma^{2(s)}} + \frac{\left(\rho_{ab}^{(s)}\right)^2}{\sigma_b^{2(s)}\left[1 - \left(\rho_{ab}^{(s)}\right)^2\right]}\right]^{-1},$$

  where $y_{im_i} - y_{i0} = \sum_{k=1}^{m_i}\tilde{\mathrm{y}}_{ik}$ and $t_{im_i} = \sum_{k=1}^{m_i}\tilde{\mathrm{t}}_{ik}$. Again, the Markovian property applies here. Then, we know that

$$\mathrm{E}_{b_i|y_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}[b_i] = \check{\mu}_{b_i}^{(s)} \text{ and } \mathrm{E}_{b_i|y_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}[b_i^2] = \left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2(s)}.$$

Therefore, $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ is shown as

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right) = -\frac{1}{2}\Bigg\{ & \sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathsf{t}}_{ik}\sigma^2\right) + \frac{\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2(s)} - 2\tilde{\mathsf{y}}_{ik}\check{\mu}_{b_i}^{(s)}/\tilde{\mathsf{t}}_{ik} + \tilde{\mathsf{y}}_{ik}^2/\tilde{\mathsf{t}}_{ik}^2}{\sigma^2/\tilde{\mathsf{t}}_{ik}}\right] + \\ & \sum_{i=1}^{n}\left[\ln\left[\left(1-\rho_{ab}^2\right)\sigma_b^2\right] + \frac{\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2(s)} - 2\left[\mu_b + \frac{\sigma_b}{\hat{\sigma}_a}\rho_{ab}(y_{i0}-\hat{\mu}_a)\right]\check{\mu}_{b_i}^{(s)} + \left[\mu_b + \frac{\sigma_b}{\hat{\sigma}_a}\rho_{ab}(y_{i0}-\hat{\mu}_a)\right]^2}{\left(1-\rho_{ab}^2\right)\sigma_b^2}\right] + \\ & \sum_{i=1}^{n}\left[\ln\left(\hat{\sigma}_a^2\right) + \frac{(y_{i0}-\hat{\mu}_a)^2}{\hat{\sigma}_a^2}\right]\Bigg\}. \end{aligned}$$

- M-step: In this step, we update the parameter estimates by solving the equations that set the first derivative of $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ equal to 0. After some manipulations based on calculus, the result is given by

$$\begin{aligned} \mu_b^{(s+1)} &= \frac{1}{n}\sum_{i=1}^{n}\check{\mu}_{b_i}^{(s)}, \\ \sigma_{ab}^{(s+1)} &= \frac{1}{n-1}\sum_{i=1}^{n}(y_{i0}-\hat{\mu}_a)\left(\check{\mu}_{b_i}^{(s)} - \mu_b^{(s+1)}\right), \\ \sigma_b^{2(s+1)} &= \left(\frac{\sigma_{ab}^{(s+1)}}{\hat{\sigma}_a}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left\{\check{\sigma}_{b_i}^{2(s)} + \left[\check{\mu}_{b_i}^{(s)} - \mu_b^{(s+1)} - \frac{\sigma_{ab}^{(s+1)}}{\hat{\sigma}_a^2}(y_{i0}-\hat{\mu}_a)\right]^2\right\}, \text{ and} \\ \sigma^{2(s+1)} &= \frac{\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\tilde{\mathsf{t}}_{ik}\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \tilde{\mathsf{t}}_{ik}\check{\sigma}_{b_i}^{2(s)} - 2\tilde{\mathsf{y}}_{ik}\check{\mu}_{b_i}^{(s)} + \tilde{\mathsf{y}}_{ik}^2/\tilde{\mathsf{t}}_{ik}\right]}{\sum_{i=1}^{n}m_i}, \end{aligned}$$

  where $\sigma_{ab}^{(s+1)}$ and $\sigma_b^{2(s+1)}$ are updated after reparameterizing $\rho_{ab}\sigma_b$ as $\sigma_{ab}/\hat{\sigma}_a$ in the Q-function. In such case, $\rho_{ab}^{(s+1)} = \sigma_{ab}^{(s+1)}/\left(\hat{\sigma}_a\sigma_b^{(s+1)}\right)$. The result also mirrors the conclusion regarding model M_0 in a multivariate setting.

❸ Lastly, recall that model M_4 is given by

$$M_4 : \begin{cases} Y_{ij}(t) = \mu_{aj}a_i + \mu_{bj}b_i t + \sigma_j\mathcal{B}(t) \\ \begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \mathcal{BVN}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{bmatrix}\right) \end{cases},$$

for which the set of all the unknown parameters is $\boldsymbol{\theta} = \{\boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \sigma_a, \sigma_b, \rho_{ab}, \boldsymbol{D}\}$. Under this scenario, the total log-likelihood function for the complete data $\{\mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\}$, where $\boldsymbol{b} = (b_1, b_2, \dots, b_n)'$ instead, turns into

$$\ell\left(\boldsymbol{\theta}; \mathbb{Y}_0, \mathbb{D}, \boldsymbol{b}\right) = \ell\left(\boldsymbol{\mu}_b, \boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b}\right) + \ell\left(\boldsymbol{\mu}_a, \sigma_a, \sigma_b, \rho_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) + \ell\left(\boldsymbol{\mu}_a, \sigma_a; \mathbb{Y}_0\right),$$

where

$$\ell\left(\boldsymbol{\mu}_b, \boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b}\right) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathrm{t}}_{ik}\sigma_j^2\right) + \frac{(\tilde{\mathrm{y}}_{ijk} - \tilde{\mathrm{t}}_{ik}\mu_{bj}b_i)^2}{\tilde{\mathrm{t}}_{ik}\sigma_j^2}\right],$$

$$\ell\left(\boldsymbol{\mu}_a, \sigma_a, \sigma_b, \rho_{ab}; \boldsymbol{b} \mid \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\left\{\ln\left[\left(1-\rho_{ab}^2\right)\sigma_b^2\right] + \frac{\left[b_i - 1 - \frac{\sigma_b}{\sigma_a}\rho_{ab}\left(\frac{y_{ij0}}{\mu_{aj}} - 1\right)\right]^2}{\left(1-\rho_{ab}^2\right)\sigma_b^2}\right\}, \text{ and}$$

$$\ell\left(\boldsymbol{\mu}_a, \sigma_a; \mathbb{Y}_0\right) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\ln\left(\mu_{aj}^2\sigma_a^2\right) + \frac{(y_{ij0} - \mu_{aj})^2}{\mu_{aj}^2\sigma_a^2}\right].$$

Firstly, regarding $\ell\left(\boldsymbol{\mu}_a, \sigma_a; \mathbb{Y}_0\right)$, the unbiased estimate of σ_a^2 given $\boldsymbol{\mu}_a$ is given by

$$\hat{\sigma}_a^2(\boldsymbol{\mu}_a) = \frac{1}{np-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{y_{ij0}}{\mu_{aj}} - 1\right)^2.$$

This relation can be substituted into $\ell\left(\boldsymbol{\mu}_a, \sigma_a; \mathbb{Y}_0\right)$ to obtain the profile likelihood of $\boldsymbol{\mu}_a$. Then, $\hat{\boldsymbol{\mu}}_a$ can be easily solved by numerically maximizing the profile likelihood. In turn, $\hat{\sigma}_a$ can be computed.

Given the current EM estimate $\boldsymbol{\theta}^{(s)}$, the algorithm is performed as follows:

- E-step: First, based on the conclusion provided by Supplementary Section C, $b_i|\boldsymbol{y}_{i0}, \tilde{\mathbf{y}}_i, \tilde{\mathbf{t}}_i, \boldsymbol{\theta}^{(s)}$ is subject to a normal distribution with mean

$$\check{\mu}_{b_i}^{(s)} = \left[\frac{1}{\sigma_b^{2(s)}} + \sum_{j=1}^{p}\frac{t_{im_i}\left(\mu_{bj}^{(s)}\right)^2}{\sigma_j^{2(s)}} + \frac{p\left(\rho_{ab}^{(s)}\right)^2}{\sigma_b^{2(s)}\left[1-\left(\rho_{ab}^{(s)}\right)^2\right]}\right]^{-1} \times$$
$$\left[\frac{1}{\sigma_b^{2(s)}} + \sum_{j=1}^{p}\frac{\mu_{bj}^{(s)}\left(y_{ijm_i} - y_{ij0}\right)}{\sigma_j^{2(s)}} + \frac{\rho_{ab}^{(s)}}{\hat{\sigma}_a\sigma_b^{(s)}\left[1-\left(\rho_{ab}^{(s)}\right)^2\right]}\sum_{j=1}^{p}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}} - 1 + \rho_{ab}^{(s)}\frac{\hat{\sigma}_a}{\sigma_b^{(s)}}\right)\right]$$

and variance

$$\check{\sigma}_{b_i}^{2^{(s)}} = \left[\frac{1}{\sigma_b^{2^{(s)}}} + \sum_{j=1}^{p} \frac{t_{im_i}\left(\mu_{bj}^{(s)}\right)^2}{\sigma_j^{2^{(s)}}} + \frac{p\left(\rho_{ab}^{(s)}\right)^2}{\sigma_b^{2^{(s)}}\left[1-\left(\rho_{ab}^{(s)}\right)^2\right]}\right]^{-1},$$

where $y_{ijm_i} - y_{ij0} = \sum_{k=1}^{m_i} \tilde{\mathrm{y}}_{ijk}$ and $t_{im_i} = \sum_{k=1}^{m_i} \tilde{\mathfrak{t}}_{ik}$. Again, the Markovian property applies here. Then, we know that

$$\mathrm{E}_{b_i|\boldsymbol{y}_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}[b_i] = \check{\mu}_{b_i}^{(s)} \text{ and } \mathrm{E}_{b_i|\boldsymbol{y}_{i0},\tilde{\mathbf{y}}_i,\tilde{\mathbf{t}}_i,\boldsymbol{\theta}^{(s)}}[b_i^2] = \left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2^{(s)}}.$$

Therefore, $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ is shown as

$$\begin{aligned}
Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right) = -\frac{1}{2}\Bigg\{&\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tilde{\mathfrak{t}}_{ik}\sigma_j^2\right) + \frac{\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2^{(s)}} - 2\tilde{\mathrm{y}}_{ijk}\check{\mu}_{b_i}^{(s)}/(\tilde{\mathfrak{t}}_{ik}\mu_{bj}) + \tilde{\mathrm{y}}_{ijk}^2/(\tilde{\mathfrak{t}}_{ik}^2\mu_{bj}^2)}{\sigma_j^2/(\tilde{\mathfrak{t}}_{ik}\mu_{bj}^2)}\right] + \\
&\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\ln\left[\left(1-\rho_{ab}^2\right)\sigma_b^2\right] + \frac{\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2^{(s)}} - 2\left[1+\frac{\sigma_b}{\hat{\sigma}_a}\rho_{ab}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}}-1\right)\right]\check{\mu}_{b_i}^{(s)} + \left[1+\frac{\sigma_b}{\hat{\sigma}_a}\rho_{ab}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}}-1\right)\right]^2}{\left(1-\rho_{ab}^2\right)\sigma_b^2}\right] + \\
&\sum_{i=1}^{n}\sum_{j=1}^{p}\left[\ln\left(\hat{\mu}_{aj}^2\hat{\sigma}_a^2\right) + \frac{(y_{ij0}-\hat{\mu}_{aj})^2}{\hat{\mu}_{aj}^2\hat{\sigma}_a^2}\right]\Bigg\}.
\end{aligned}$$

- M-step: In this step, we update the parameter estimates by solving the equations that set the first derivative of $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ equal to 0. After some manipulations based on calculus, the result is given by

$$\mu_{bj}^{(s+1)} = \frac{\sum_{i=1}^{n}\check{\mu}_{b_i}^{(s)}(y_{ijm_i} - y_{ij0})}{\sum_{i=1}^{n}\left[\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2^{(s)}}\right]t_{im_i}},$$

$$\sigma_j^{2^{(s+1)}} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left\{\tilde{\mathfrak{t}}_{ik}\left(\mu_{bj}^{(s+1)}\right)^2\left[\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2^{(s)}}\right] - 2\tilde{\mathrm{y}}_{ijk}\check{\mu}_{b_i}^{(s)}\mu_{bj}^{(s+1)} + \tilde{\mathrm{y}}_{ijk}^2/\tilde{\mathfrak{t}}_{ik}\right\}}{\sum_{i=1}^{n} m_i},$$

$$\sigma_{ab}^{(s+1)} = \frac{1}{np-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}}-1\right)\left(\check{\mu}_{b_i}^{(s)}-1\right), \text{ and}$$

$$\sigma_b^{2^{(s+1)}} = \left(\frac{\sigma_{ab}^{(s+1)}}{\hat{\sigma}_a}\right)^2 + \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}\left\{\check{\sigma}_{b_i}^{2^{(s)}} + \left[\check{\mu}_{b_i}^{(s)} - 1 - \frac{\sigma_{ab}^{(s+1)}}{\hat{\sigma}_a^2}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}}-1\right)\right]^2\right\},$$

where $\sigma_{ab}^{(s+1)}$ and $\sigma_b^{2^{(s+1)}}$ are updated by making use of the same trick above.

D.3 Consideration of Time Scale Transformation

With the consideration of time scale transformation, we denote the transformed time interval by $\tau_{ijk} = \Lambda_j(t_{ik}) - \Lambda_j(t_{i,k-1}), \forall\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, p, k = 1, 2, \ldots, m_i$, with an unknown parameter γ_j. Note that in the decomposed log-likelihood functions, only $\ell(\boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b})$ involves the information of time intervals directly. Therefore, in the E-step, after updating $\check{\boldsymbol{\mu}}_{\boldsymbol{b}_i}^{(s)}$ and $\check{\boldsymbol{\Sigma}}_{\boldsymbol{b}_i}^{(s)}$ with $\gamma_j^{(s)}$, $\ell(\boldsymbol{D}; \mathbb{D} \mid \boldsymbol{b})$ in $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(s)}\right)$ becomes

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m_i}\left[\ln\left(\tau_{ijk}\sigma_j^2\right) + \frac{\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2 + \check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)} - 2\tilde{\mathsf{y}}_{ijk}\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}/\tau_{ijk} + \tilde{\mathsf{y}}_{ijk}^2/\tau_{ijk}^2}{\sigma_j^2/\tau_{ijk}}\right].$$

Then, in the M-step, $\gamma_j^{(s+1)}$ is equal to the solution of γ_j to the equation below:

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left\{\frac{\tau'_{ijk}}{\tau_{ijk}} + \frac{\tau'_{ijk}\left[\left(\check{\mu}_{\boldsymbol{b}_i,j}^{(s)}\right)^2 + \check{\sigma}_{\boldsymbol{b}_i,j}^{2(s)}\right] - \tilde{\mathsf{y}}_{ijk}^2\tau'_{ijk}/\tau_{ijk}^2}{\sigma_j^{2(s+1)}}\right\} = 0.$$

In this paper, we make use of the power transformation, i.e., $\tau_{ijk} = t_{ik}^{\gamma_j} - t_{i,k-1}^{\gamma_j}$. Thus, τ'_{ijk} should be replaced with $t_{ik}^{\gamma_j}\ln t_{ik} - t_{i,k-1}^{\gamma_j}\ln t_{i,k-1}$ in the equation above. Particularly, when $t_{i,k-1} = 0$, we set $\tau_{ijk} = t_{ik}^{\gamma_j}$. For models M_1 to M_3, the estimate of the time scale transformation parameter is found similarly.

Specially, for model M_4, $\gamma_j^{(s+1)}$ is equal to the solution of γ_j to the equation below:

$$\sum_{i=1}^{n}\sum_{k=1}^{m_i}\left[\frac{\tau'_{ijk}}{\tau_{ijk}} + \frac{\tau'_{ijk}\left(\mu_{bj}^{(s+1)}\right)^2\left[\left(\check{\mu}_{b_i}^{(s)}\right)^2 + \check{\sigma}_{b_i}^{2(s)}\right] - \tilde{\mathsf{y}}_{ijk}^2\tau'_{ijk}/\tau_{ijk}^2}{\sigma_j^{2(s+1)}}\right] = 0.$$

D.4 Helpful Starting Values of Parameter Estimates

Recall that $\hat{\boldsymbol{\mu}}_a$ and $\hat{\boldsymbol{\Sigma}}_a$ are directly obtained by applying MLE on the log-likelihood function in terms of the data $\mathbb{Y}_0$. So there is no need to provide starting values for these two components. For the remaining terms, the following steps are performed:

1. Treat the degradation increments of each degradation path, i.e., $(\tilde{\mathsf{y}}_{ij1}, \tilde{\mathsf{y}}_{ij2}, \ldots, \tilde{\mathsf{y}}_{ijm_i})'$, $\forall\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, p$, as an independent realization from a simple (non)linear Wiener process, $\mathcal{N}\left(b_{ij}\Lambda_j(t), \sigma_{ij}^2\Lambda_j(t)\right)$. Fit the model to the data and obtain estimated parameters $\hat{b}_{ij}$, $\hat{\sigma}_{ij}^2$ and possibly $\hat{\gamma}_{ij}$, $\forall\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, p$.

2. Denote $\hat{\boldsymbol{b}}_i = (\hat{b}_{i1}, \hat{b}_{i2}, \dots, \hat{b}_{ip})'$, $\forall\ i = 1, 2, \dots, n$, and compute

$$\sigma_j^{2(0)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_{ij}^2,\ \boldsymbol{\mu}_b^{(0)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{b}}_i,$$

$$\boldsymbol{\Sigma}_b^{(0)} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\mu}_b^{(0)}\right)\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\mu}_b^{(0)}\right)', \text{ and } \boldsymbol{\Sigma}_{ab}^{(0)} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{y}_{i0} - \hat{\boldsymbol{\mu}}_a\right)\left(\hat{\boldsymbol{b}}_i - \boldsymbol{\mu}_b^{(0)}\right)'.$$

If nonlinearity exists, $\gamma_j^{(0)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_{ij}$ needs also be computed. Then, set these calculated values as the initial parameter estimates and feed them into the E-step.

Specially, for model M_4, $\sigma_b^{2(0)}$ and $\rho_{ab}^{(0)}$ are computed as

$$\sigma_b^{2(0)} = \frac{1}{np-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{\hat{b}_{ij}}{\mu_{bj}^{(0)}} - 1\right)^2 \text{ and } \rho_{ab}^{(0)} = \frac{1}{np-1}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\frac{y_{ij0}}{\hat{\mu}_{aj}} - 1\right)\left(\frac{\hat{b}_{ij}}{\mu_{bj}^{(0)}} - 1\right).$$

E The BCp Bootstrap Method

For illustrative purposes, we use the BCp bootstrap method to demonstrate the computation of the C.I. for $F_{T_{\boldsymbol{\mathcal{D}}}}(t; \boldsymbol{\theta}, \boldsymbol{\mathcal{D}})$ of the proposed model. It is performed according to the followings:

1. Given the observed data $\mathbb{D}$, implement the EM algorithm to obtain the parameter estimate $\hat{\boldsymbol{\theta}}$ and calculate the estimated failure time probability $\hat{F}_{T_{\boldsymbol{\mathcal{D}}}}(t; \hat{\boldsymbol{\theta}}, \boldsymbol{\mathcal{D}})$ (abbreviated $\hat{F}(t)$) at desired values of t.

2. Generate a large number B (say $B = 1,000$) of bootstrap samples that mimic the original sample and compute the corresponding bootstrap estimates $\hat{F}^*_{T_{\boldsymbol{\mathcal{D}}}}(t; \hat{\boldsymbol{\theta}}^*_b, \boldsymbol{\mathcal{D}})_b$ (abbreviated $\hat{F}^*(t)_b$), $b = 1, 2, \dots, B$, according to the following steps:

   (a) Generate n simulated realizations from the random effects, i.e.,

   $$\begin{pmatrix} \boldsymbol{Y}_i^*(0) \\ \boldsymbol{b}_i^* \end{pmatrix} \sim \mathcal{MVN}\left(\begin{bmatrix} \hat{\boldsymbol{\mu}}_a \\ \hat{\boldsymbol{\mu}}_b \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_a & \hat{\boldsymbol{\Sigma}}_{ab} \\ \hat{\boldsymbol{\Sigma}}_{ba} & \hat{\boldsymbol{\Sigma}}_b \end{bmatrix}\right),\ i = 1, 2, \dots, n.$$

   (b) For each pair of $(y_{ij}^*(0), b_{ij}^*)$, generate p simulated degradation paths based on $y_{ijk}^* = y_{ij,k-1}^* + \mathsf{y}_{ijk}^*$, $\forall\ i = 1, 2, \dots, n,\ j = 1, 2, \dots, p,\ k = 1, 2, \dots, m_i$, where $y_{ij0}^* = y_{ij}^*(0)$ and y_{ijk}^* is sampled from $\mathcal{N}(\mathsf{t}_{ik} b_{ij}^*, \mathsf{t}_{ik}\hat{\sigma}_j^2)$.

   (c) Use the simulated degradation paths as inputs to produce the bootstrap estimates $\hat{\boldsymbol{\theta}}^*_b$ and compute $\hat{F}^*(t)_b$ at desired values of t.

3. For each desired value of t, the bootstrap C.I. for $F_{T_{\boldsymbol{\mathcal{D}}}}(t; \boldsymbol{\theta}, \boldsymbol{\mathcal{D}})$ is constructed as below:

(a) Sort the bootstrap estimates $\hat{F}^*(t)_1, \ldots, \hat{F}^*(t)_B$ in increasing order giving $\hat{F}^*(t)_{(b)}$, $b = 1, 2, \ldots, B$.

(b) The lower and upper bounds of point-wise approximate $100(1-\alpha)\%$ C.I. are $[\hat{F}^*(t)_{(L)}, \hat{F}^*(t)_{(U)}]$, where

$$L = B \times \Phi\left[2\Phi^{-1}(q) + \Phi^{-1}(\alpha/2)\right],$$

$$U = B \times \Phi\left[2\Phi^{-1}(q) + \Phi^{-1}(1-\alpha/2)\right],$$

and q is the proportion of the bootstrap estimates $\{\hat{F}^*(t)_b, b = 1, 2, \ldots, B\}$ that are less than $\hat{F}(t)$.

F Additional Results of the Simulation Studies

F.1 Parameter Estimation v.s. Sample Size

Table F.1. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_0.

n	m	$\mu_{b1}=5$	$\mu_{b2}=4$	$\mu_{b3}=3$	$\sigma^2_{b1}=1$	$\sigma^2_{b2}=1$	$\sigma^2_{b3}=1$	$\rho^b_{12}=0.8$	$\rho^b_{13}=0.8$	$\rho^b_{23}=0.8$
10	15	359.462	374.314	378.424	516.805	533.453	597.773	309.720	289.675	311.681
	30	342.470	343.790	352.317	472.246	495.023	503.878	225.580	236.438	237.803
20	15	249.888	257.263	266.886	373.677	396.473	418.752	151.479	168.434	173.639
	30	230.232	236.228	243.990	335.944	351.598	363.669	124.962	133.352	135.559
30	15	204.209	211.188	215.520	316.638	340.905	336.138	128.798	134.425	141.356
	30	194.144	194.325	194.959	299.082	296.896	292.872	97.158	105.512	110.172

n	m	$\rho^{ab}_1=0.3$	$\rho^{ab}_{21}=0.1$	$\rho^{ab}_{31}=0.1$	$\rho^{ab}_{12}=0.1$	$\rho^{ab}_2=0.3$	$\rho^{ab}_{32}=0.1$	$\rho^{ab}_{13}=0.1$	$\rho^{ab}_{23}=0.1$	$\rho^{ab}_3=0.3$
10	15	362.520	390.217	390.688	406.921	359.368	401.831	399.684	388.667	382.170
	30	342.195	376.332	376.780	375.708	335.164	364.781	381.185	370.214	367.230
20	15	241.739	262.323	257.455	266.883	262.288	271.117	279.115	269.515	255.695
	30	231.929	251.544	239.608	251.092	235.895	256.363	261.803	254.358	242.954
30	15	201.022	214.112	205.828	227.788	197.356	215.606	223.178	219.262	205.130
	30	190.243	202.372	198.693	207.799	188.226	204.507	207.734	209.684	187.273

n	m	$\sigma^2_1=4$	$\sigma^2_2=5$	$\sigma^2_3=6$
10	15	473.280	565.385	708.471
	30	336.107	415.068	505.324
20	15	330.421	415.113	503.831
	30	233.559	290.281	356.259
30	15	262.920	337.809	398.403
	30	195.457	242.525	282.553

F.2 Parameter Estimation v.s. Degradation Curvature

In this section, we evaluate the performance of model estimation under different degradation curvatures. The simulation model used here is the same as in Section F.1, with the addition of time scale transformation using parameter γ_j for $j = 1, 2, 3$, set to $0.7, 1$, and 1.3 for three different scenarios. The sample size is chosen from three options, $n = 10, 20, 30$, with $m = 30$. Figure F.1 displays the average RMSE across all parameters for each scenario. As expected, estimation accuracy improves with larger sample sizes. Interestingly, we also find that the shape of the degradation curve affects the estimation performance. Specifically, the convex degradation trend (i.e., $\gamma_j > 1$) results in the best performance among the three scenarios. This is because a convex degradation path implies a degradation rate that increases with the level of degradation, resulting in a larger value of degradation level for a fixed inspection period compared to concave or linear degradation paths. This can be interpreted as spanning a longer time horizon for the convex degradation scenario, resulting in more data points for parameter estimation and, therefore, lower RMSE values.

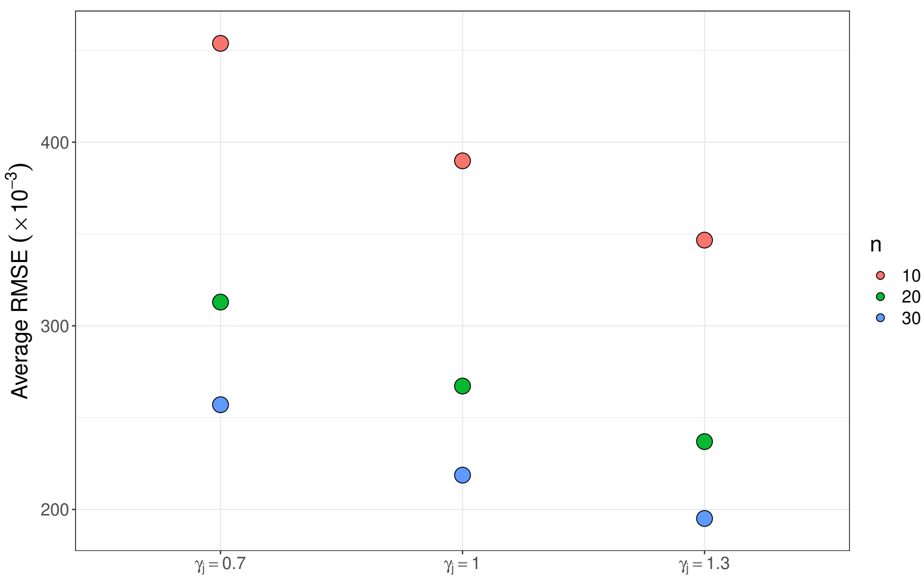


Figure F.1. Average RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes & Different Degradation Shapes for Model M_0.

F.3 Parameter Estimation v.s. Correlation

In this section, we evaluate the performance of model estimation with different correlations. Specifically, we examine three options for both the PC-wise correlation and the initiation-growth correlation, denoted as $\rho_b = 0.1, 0.5, 0.9$ and $\rho_{ab} = 0.1, 0.5, 0.9$ across all processes.

The simulation model is the same as the one used in Section F.1 except for setting $\rho_a = 0.9$. Note that not all combinations of ρ_b and ρ_{ab} are feasible due to the positive definiteness of $\mathbf{\Sigma}$. For simplicity, We provide two sample size options, $n = 30, 60$ under $m = 30$. Similar to Figure 3, Table F.2 presents the average RMSE over the elements in $\hat{\boldsymbol{\mu}}_b$, $\hat{\boldsymbol{\sigma}}_b^2$, $\hat{\boldsymbol{\rho}}_b$, $\hat{\boldsymbol{\rho}}_j^{ab}$, $\hat{\boldsymbol{\rho}}_{jj'}^{ab}$, and $\hat{\boldsymbol{\sigma}}^2$. As expected, the estimation accuracy improves with the increase in sample size. Furthermore, the estimation performance of the parameters associated with $\boldsymbol{\mu}_b$, $\boldsymbol{\sigma}_b^2$, and $\boldsymbol{\sigma}^2$ remains consistent regardless of the values of ρ_b and ρ_{ab}. However, the RMSE of $\hat{\boldsymbol{\rho}}_b$ decreases as ρ_b increases with a fixed ρ_{ab}. An identical pattern happens for $\hat{\boldsymbol{\rho}}_j^{ab}$ and $\hat{\boldsymbol{\rho}}_{jj'}^{ab}$ with the increase of ρ_{ab} given a fixed ρ_b. As we explained in Section 4.1, an intuitive explanation is that the stronger correlation results in more "similar" data, which can largely enhance the estimation accuracy.

Table F.2. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes & Different Correlations for Model M_0.

n	ρ_b	ρ_{ab}	$\hat{\boldsymbol{\mu}}_b$	$\hat{\boldsymbol{\sigma}}_b^2$	$\hat{\boldsymbol{\rho}}_b$	$\hat{\boldsymbol{\rho}}_j^{ab}$	$\hat{\boldsymbol{\rho}}_{jj'}^{ab}$	$\hat{\boldsymbol{\sigma}}^2$
30	0.1	0.1	192.099	301.397	222.328	207.739	207.629	241.891
		0.5	192.847	305.115	223.800	159.001	161.068	242.450
	0.5	0.1	195.392	299.110	178.413	206.851	208.678	242.478
		0.5	194.849	302.242	181.171	162.137	164.794	241.457
	0.9	0.1	196.184	293.784	69.351	205.118	204.360	239.784
		0.5	195.915	294.550	70.464	161.042	161.323	239.966
		0.9	197.229	294.034	67.075	55.094	54.395	239.275
60	0.1	0.1	136.397	211.124	153.692	137.621	138.418	173.192
		0.5	135.022	217.858	151.598	108.135	108.959	172.092
	0.5	0.1	134.377	212.042	124.097	140.092	140.428	171.244
		0.5	134.774	213.864	124.189	107.096	108.884	172.089
	0.9	0.1	134.296	203.630	51.871	134.725	134.701	170.703
		0.5	133.444	207.328	51.384	107.025	107.371	172.667
		0.9	138.375	204.120	50.965	37.619	37.888	172.696

F.4 Parameter Estimation v.s. Process Dimension

In this section, we evaluate the performance of our model estimation with different process dimensions, namely $p = 3, 6, 9$. We maintain the same simulation model as in Section F.1 for $p = 3$. For the other two dimensions, we increase the parameter setting proportionally. The sample size is set to $n = 30, 45, 60$ under $m = 30$. Figure F.2 shows the average RMSE over all parameters for each scenario. As expected, the estimation accuracy improves with increased sample size and process dimensions. Figure F.3 presents the average execution time (in seconds) over 1,000 runs for each scenario. We should owe such a fast performance to the model's tractability. Moreover, the average execution time increases approximately linearly with the number of PCs across different sample sizes, indicating that the EM algorithm reduces the optimization complexity as p increases.

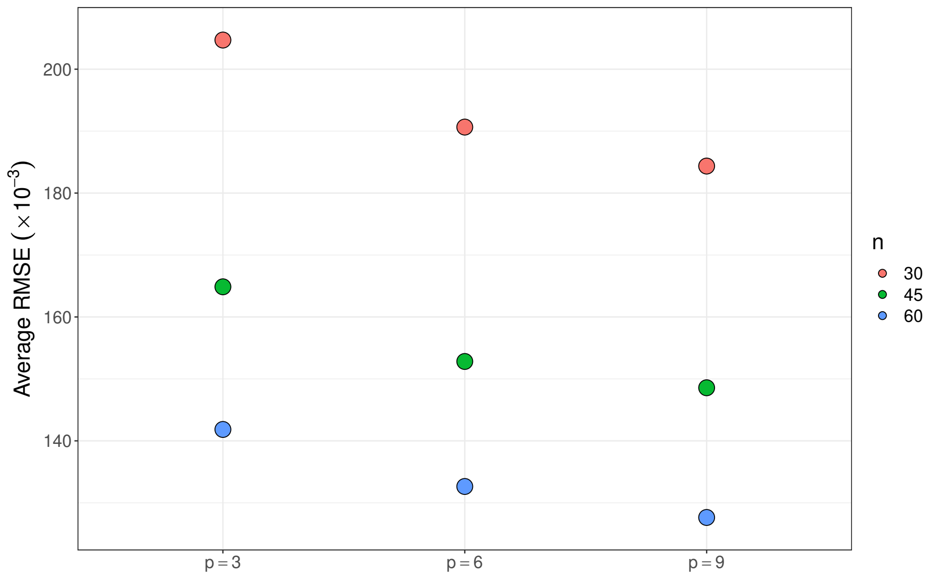


Figure F.2. Average RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes & Different Process Dimensions for Model M_0.

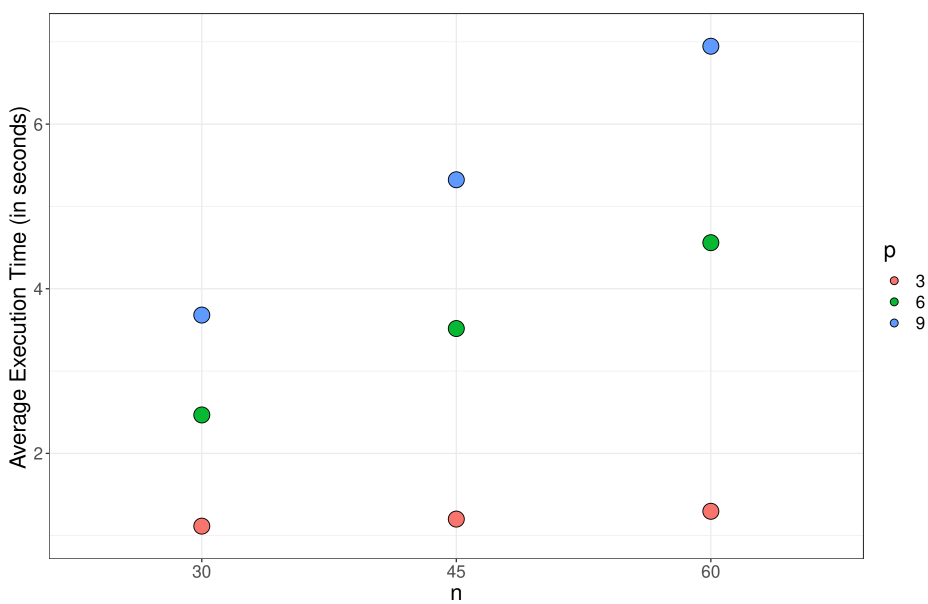


Figure F.3. Average Execution Time (in seconds) versus Various Sample Sizes & Different Process Dimensions for Model M_0. The CPU utilized was Apple M1 Max.

F.5 Parameter Estimation v.s. Sample Size for Models M_2-M_4

Tables F.3-F.5 report the RMSE of each EM estimator versus various sample sizes for models M_2, M_3, and M_4, respectively. Figures F.4-F.6 present the result in dot plots.

Table F.3. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_2.

n	m	$\mu_{b1}=5$	$\mu_{b2}=4$	$\mu_{b3}=3$	$\sigma^2_{b1}=1$	$\sigma^2_{b2}=1$	$\sigma^2_{b3}=1$
10	15	360.160	364.414	372.807	538.000	571.864	591.195
	30	337.767	338.629	352.016	496.023	499.533	495.752
20	15	252.453	251.089	262.098	390.422	412.515	404.823
	30	236.565	238.566	241.661	359.499	354.143	354.721
30	15	201.990	201.570	212.674	319.143	340.954	347.944
	30	190.541	193.984	197.665	276.825	288.780	303.815
n	m	$\rho^b_{12}=0.2$	$\rho^b_{13}=0.5$	$\rho^b_{23}=0.8$	$\sigma^2_1=4$	$\sigma^2_2=5$	$\sigma^2_3=6$
10	15	480.987	399.091	312.531	464.433	592.667	717.696
	30	403.805	346.623	243.560	337.117	426.715	498.320
20	15	309.295	275.455	178.541	334.129	425.471	494.984
	30	261.167	226.606	135.089	236.299	294.711	357.622
30	15	244.629	203.792	140.049	270.397	342.539	403.960
	30	215.181	176.020	108.819	194.711	242.391	286.069

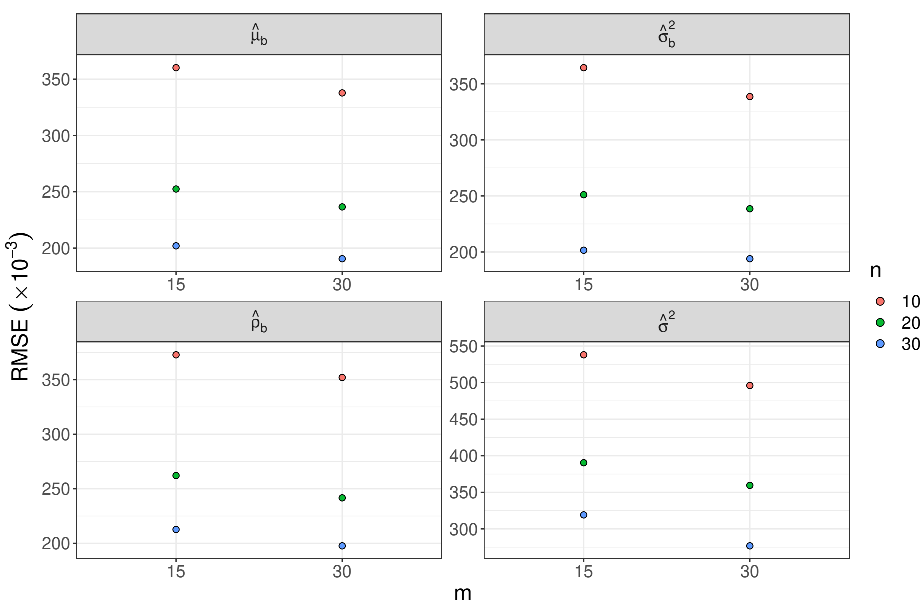


Figure F.4. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_2.

Table F.4. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_3.

n	m	$\mu_b = 5$	$\sigma_b^2 = 0.8$	$\rho_{ab} = 0.3$	$\sigma^2 = 4$
10	15	331.902	457.343	420.690	481.095
	30	310.153	399.560	373.319	333.234
20	15	228.935	321.202	264.827	343.880
	30	219.074	287.086	239.177	241.143
30	15	188.163	262.644	210.879	282.759
	30	177.502	235.491	192.328	193.293

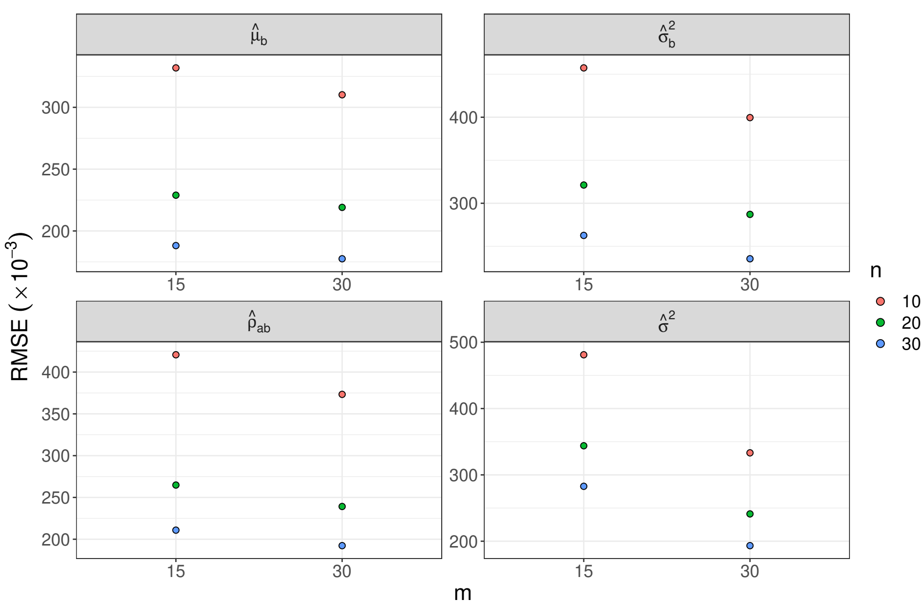


Figure F.5. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_3.

Table F.5. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_4.

n	m	$\mu_{b1} = 5$	$\mu_{b2} = 4$	$\mu_{b3} = 3$	$\sigma_b^2 = 0.3$
10	15	866.676	707.980	544.569	145.050
	30	856.739	692.006	526.826	143.221
20	15	625.505	504.186	394.991	108.452
	30	616.987	499.773	380.737	106.654
30	15	527.456	433.197	327.585	97.005
	30	527.685	417.587	324.924	95.920

n	m	$\rho_{ab} = 0.5$	$\sigma_1^2 = 4$	$\sigma_2^2 = 5$	$\sigma_3^2 = 6$
10	15	315.525	551.582	599.043	704.054
	30	297.286	334.369	420.084	502.618
20	15	210.385	313.642	410.793	536.728
	30	203.555	225.932	301.844	368.702
30	15	150.133	285.624	321.885	392.115
	30	150.834	190.286	227.986	290.951

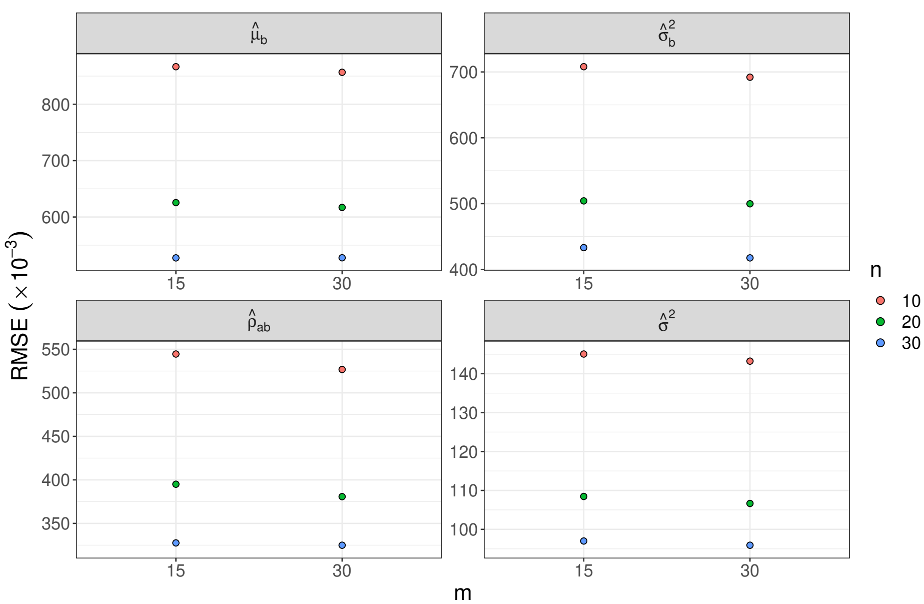


Figure F.6. RMSE ($\times 10^{-3}$) of the EM Estimators versus Various Sample Sizes for Model M_4.

G Additional Result of the Case Studies

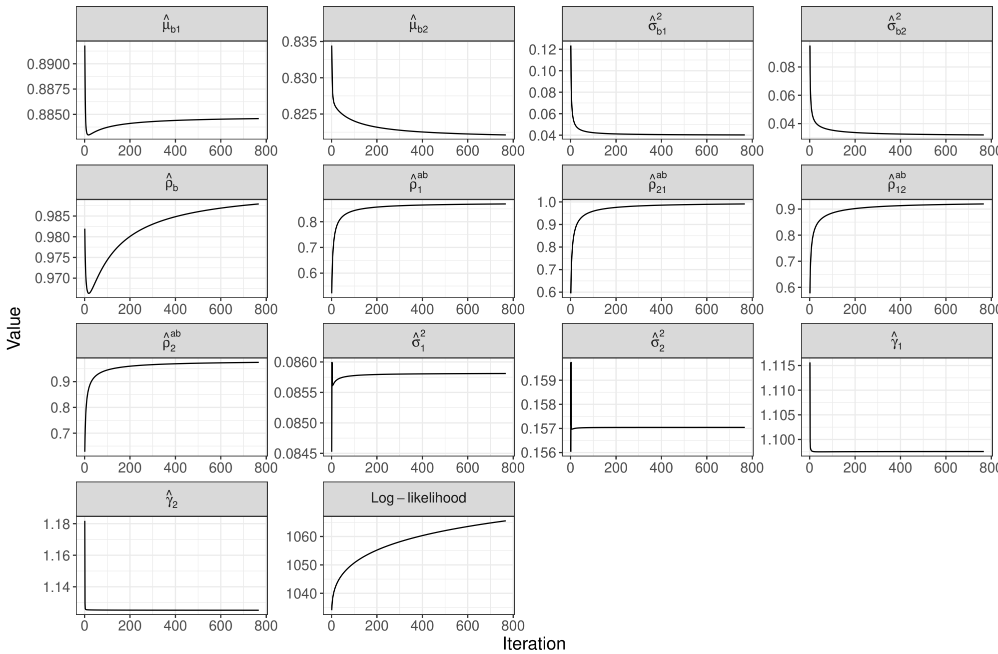


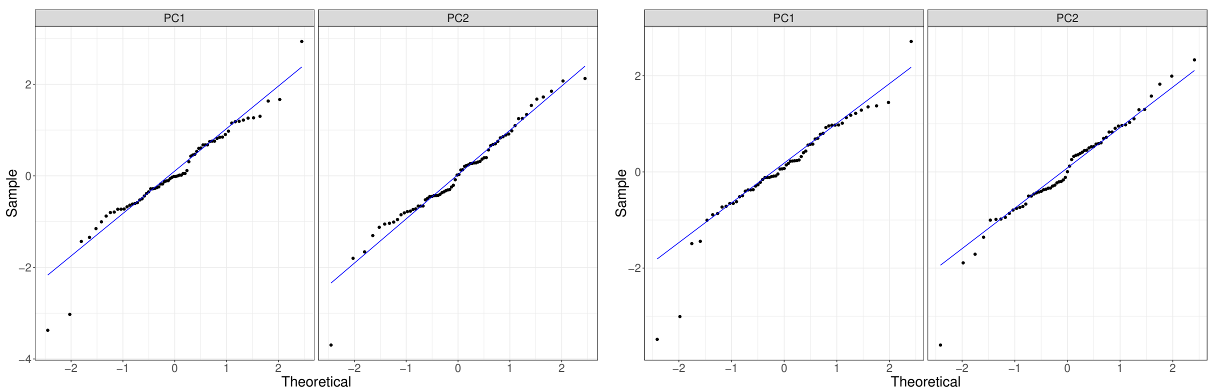
Figure G.1. Parameter Estimates over Iteration for Model M_0 regarding the Transceiver Degradation Data.

Table G.1. IRLED Degradation Data.

	Aging Time (×10 Hours)										
Unit	0	0.6	1.2	2.4	4.8	9.6	15.6	23.0	32.4	47.9	63.5
PC1											
1	0	4.3	6.5	7.8	13.0	21.7	33.0	42.1	49.9	59.9	78.6
2	0	1.4	2.7	5.0	7.8	14.5	23.3	29.0	43.3	59.8	77.4
3	0	3.4	4.6	7.8	13.0	16.8	26.8	34.1	41.5	67.0	65.5
4	0	3.2	4.3	5.8	9.9	15.2	20.3	26.2	33.6	39.5	53.2
5	0	3.7	5.6	8.0	12.8	16.0	23.7	26.7	38.4	49.2	47.2
6	0	0.1	0.4	0.7	2.0	3.5	6.6	12.2	18.8	32.3	47.0
7	0	0.8	1.7	2.8	4.6	7.9	12.4	20.2	24.8	32.5	45.5
PC2											
1	0	2.3	3.7	5.6	8.8	13.7	17.2	24.8	29.1	42.9	45.3
2	0	0.2	0.4	0.9	2.4	4.5	7.1	13.4	21.2	30.7	41.7
3	0	0.5	0.9	1.9	3.5	5.9	10.0	14.4	22.0	26.0	31.8
4	0	2.6	4.4	6.0	8.7	14.6	16.8	17.9	23.2	27.0	31.3
5	0	4.3	5.8	9.5	10.2	13.8	20.6	19.7	25.3	33.4	27.9
6	0	3.6	4.7	6.2	9.1	11.7	13.8	14.5	15.5	23.1	24.0
7	0	0.5	0.9	1.4	3.3	5.0	6.1	9.9	13.2	17.0	20.7

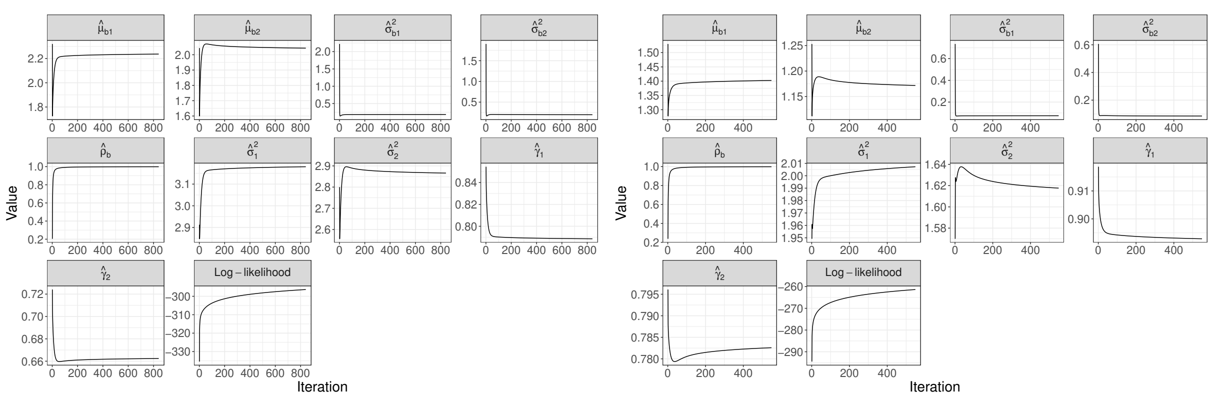
Table G.2. Parameter Point Estimates and Standard Errors (shown in parentheses), along with Log-likelihood for Various Models regarding the IRLED Degradation Data.

	$\hat{\mu}_{b1}$	$\hat{\mu}_{b2}$	$\hat{\sigma}^2_{b1}$	$\hat{\sigma}^2_{b2}$	$\hat{\rho}^b_{12}$	$\hat{\sigma}^2_1$	$\hat{\sigma}^2_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	Log-likelihood
M_2	2.237 (0.443)	2.063 (0.454)	0.183 (0.163)	0.197 (0.210)	0.999 (0.487)	3.181 (0.674)	2.866 (0.677)	0.789 (0.042)	0.662 (0.048)	−296.332
	$\hat{\mu}_{a1}$	$\hat{\mu}_{a2}$	$\hat{\sigma}^2_{a1}$	$\hat{\sigma}^2_{a2}$	$\hat{\rho}^a_{12}$	$\hat{\mu}_{b1}$	$\hat{\mu}_{b2}$	$\hat{\sigma}^2_{b1}$	$\hat{\sigma}^2_{b2}$	$\hat{\rho}^b_{12}$
M_1	2.414 (0.556)	2.000 (0.497)	2.631 (1.033)	2.673 (1.057)	0.192 (0.408)	1.402 (0.298)	1.172 (0.309)	0.074 (0.067)	0.085 (0.106)	0.999 (0.439)
	$\hat{\sigma}^2_1$	$\hat{\sigma}^2_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	Log-likelihood					
	2.007 (0.463)	1.618 (0.389)	0.893 (0.047)	0.783 (0.053)	−286.920					



(a) Model M_2. (b) Model M_1.

Figure G.2. Normal Q-Q Plots for Various Models regarding the IRLED Degradation Data.



(a) Model M_2. (b) Model M_1.

Figure G.3. Parameter Estimates over Iteration for Various Models regarding the IRLED Degradation Data.

H List of Notations

Indices:	
n	number of units ($i \in \{1, ..., n\}$)
p	number of performance characteristics ($j \in \{1, ..., p\}$)
m_i	number of measurements ($k \in \{1, ..., m_i\}$)
Random Variables:	
$\boldsymbol{Y}_i(t)$	degradation values on unit i at time t, where $\boldsymbol{Y}_i(t) = (Y_{i1}(t), Y_{i2}(t), \ldots, Y_{ip}(t))'$ with $\boldsymbol{Y}_i(0)$ representing the initial levels
$\tilde{\mathbf{Y}}_i(t, 0)$	degradation increments from time 0 to t, where $\tilde{\mathbf{Y}}_i(t, 0) = \boldsymbol{Y}_i(t) - \boldsymbol{Y}_i(0)$
$\boldsymbol{b}_i$	degradation rates, where $\boldsymbol{b}_i = (b_{i1}, b_{i2}, \ldots, b_{ip})'$
$\boldsymbol{\mathcal{B}}_p(t)$	the standard p-dimensional Brownian motion (Wiener process)
Data:	
t_{ik}	the elapsed time when the k-th measurement on unit i is taken
$\tilde{\mathsf{t}}_{ik}$	measurement time interval with $\tilde{\mathsf{t}}_{ik} = t_{ik} - t_{i,k-1}$
τ_{ijk}	transformed time interval replacing $\tilde{\mathsf{t}}_{ik}$ with $\tau_{ijk} = \Lambda_j(t_{ik}) - \Lambda_j(t_{i,k-1})$ if time scale transformation is applicable
y_{ijk}	degradation measurement of process j on unit i at time point k
$\tilde{\mathsf{y}}_{ijk}$	degradation increment with $\tilde{\mathsf{y}}_{ijk} = y_{ijk} - y_{ij,k-1}$
$\tilde{\mathbf{y}}_i$	collection of all the degradation increments on unit i, where $\tilde{\mathbf{y}}_{ik} = (\tilde{\mathsf{y}}_{i1k}, \tilde{\mathsf{y}}_{i2k}, \ldots, \tilde{\mathsf{y}}_{ipk})'$ and $\tilde{\mathbf{y}}_i = \begin{pmatrix} \tilde{\mathbf{y}}'_{i1} \\ \tilde{\mathbf{y}}'_{i2} \\ \vdots \\ \tilde{\mathbf{y}}'_{im_i} \end{pmatrix} = \begin{pmatrix} \tilde{\mathsf{y}}_{i11} & \tilde{\mathsf{y}}_{i21} & \cdots & \tilde{\mathsf{y}}_{ip1} \\ \tilde{\mathsf{y}}_{i12} & \tilde{\mathsf{y}}_{i22} & \cdots & \tilde{\mathsf{y}}_{ip2} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathsf{y}}_{i1m_i} & \tilde{\mathsf{y}}_{i2m_i} & \cdots & \tilde{\mathsf{y}}_{ipm_i} \end{pmatrix}_{m_i \times p}$
$\tilde{\mathbf{t}}_i$	collection of all the time intervals on unit i, where $\tilde{\mathbf{t}}_i = (\tilde{\mathsf{t}}_{i1}, \tilde{\mathsf{t}}_{i2}, \ldots, \tilde{\mathsf{t}}_{im_i})'$
$\mathbb{Y}_0$	collection of all the initial degradation levels, where $\boldsymbol{y}_{i0} = (y_{i10}, y_{i20}, \ldots, y_{ip0})'$ and $\mathbb{Y}_0 = \begin{pmatrix} \boldsymbol{y}'_{10} \\ \boldsymbol{y}'_{20} \\ \vdots \\ \boldsymbol{y}'_{n0} \end{pmatrix} = \begin{pmatrix} y_{110} & y_{120} & \cdots & y_{1p0} \\ y_{210} & y_{220} & \cdots & y_{2p0} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n10} & y_{n20} & \cdots & y_{np0} \end{pmatrix}_{n \times p}$
$\mathbb{D}$	collection of degradation increments and time intervals, where $\mathbb{D} = \{\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \ldots, \tilde{\mathbf{y}}_n, \tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2, \ldots, \tilde{\mathbf{t}}_n\}$
$\boldsymbol{\mathcal{D}}$	vector of all the failure thresholds with $\boldsymbol{\mathcal{D}} = (\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_p)'$

Parameters:	
$\boldsymbol{\mu}_a$	mean vector of the initial levels, where $\boldsymbol{\mu}_a = (\mu_{a1}, \mu_{a2}, \ldots, \mu_{ap})'$
$\boldsymbol{\mu}_b$	mean vector of the degradation rates, where $\boldsymbol{\mu}_b = (\mu_{b1}, \mu_{b2}, \ldots, \mu_{bp})'$
$\boldsymbol{\sigma}_a^2$	variance vector of the initial levels, where $\boldsymbol{\sigma}_a^2 = (\sigma_{a1}^2, \sigma_{a2}^2, \ldots, \sigma_{ap}^2)'$
$\boldsymbol{\rho}_a$	correlation vector of the initial levels, where $\boldsymbol{\rho}_a = (\rho_{12}^a, \rho_{13}^a, \ldots, \rho_{p-1,p}^a)'$
$\boldsymbol{\sigma}_b^2$	variance vector of the degradation rates, where $\boldsymbol{\sigma}_b^2 = (\sigma_{b1}^2, \sigma_{b2}^2, \ldots, \sigma_{bp}^2)'$
$\boldsymbol{\rho}_b$	correlation vector of the degradation rates, where $\boldsymbol{\rho}_b = (\rho_{12}^b, \rho_{13}^b, \ldots, \rho_{p-1,p}^b)'$
$\boldsymbol{\rho}_{ab}$	cross-correlation vector, where $\boldsymbol{\rho}_{ab} = (\rho_1^{ab}, \rho_2^{ab}, \ldots, \rho_p^{ab}, \rho_{12}^{ab}, \rho_{13}^{ab}, \ldots, \rho_{p-1,p}^{ab}, \rho_{21}^{ab}, \rho_{31}^{ab}, \ldots, \rho_{p,p-1}^{ab})'$
$\boldsymbol{\sigma}^2$	variance vector of the standard Brownian motion, where $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)'$
$\boldsymbol{\gamma}$	vector of the time-scale transformation parameters, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$
$\boldsymbol{\Sigma}_a$	variance-covariance matrix of the initial levels, where $\boldsymbol{\Sigma}_a = \begin{pmatrix} \sigma_{a1}^2 & \sigma_{12}^a & \cdots & \sigma_{1p}^a \\ \sigma_{12}^a & \sigma_{a2}^2 & \cdots & \sigma_{2p}^a \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p}^a & \sigma_{2p}^a & \cdots & \sigma_{ap}^2 \end{pmatrix}$
$\boldsymbol{\Sigma}_b$	variance-covariance matrix of the degradation rates, where $\boldsymbol{\Sigma}_b = \begin{pmatrix} \sigma_{b1}^2 & \sigma_{12}^b & \cdots & \sigma_{1p}^b \\ \sigma_{12}^b & \sigma_{b2}^2 & \cdots & \sigma_{2p}^b \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p}^b & \sigma_{2p}^b & \cdots & \sigma_{bp}^2 \end{pmatrix}$
$\boldsymbol{\Sigma}_{ab}$	cross-covariance matrix, where $\boldsymbol{\Sigma}_{ab} = \begin{pmatrix} \sigma_1^{ab} & \sigma_{12}^{ab} & \cdots & \sigma_{1p}^{ab} \\ \sigma_{21}^{ab} & \sigma_2^{ab} & \cdots & \sigma_{2p}^{ab} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}^{ab} & \sigma_{p2}^{ab} & \cdots & \sigma_p^{ab} \end{pmatrix}$
$\boldsymbol{D}$	diagonal matrix of $\boldsymbol{\sigma}^2$, where $\boldsymbol{D} = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2\right)$
$\boldsymbol{\theta}$	collection of all the unknown parameters, where $\boldsymbol{\theta} = \{\boldsymbol{\mu}_a, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_{ab}, \boldsymbol{D}\}$ or $\boldsymbol{\theta} = (\boldsymbol{\mu}_a', \boldsymbol{\mu}_b', \boldsymbol{\sigma}_a^{2'}, \boldsymbol{\rho}_a', \boldsymbol{\sigma}_b^{2'}, \boldsymbol{\rho}_b', \boldsymbol{\rho}_{ab}', \boldsymbol{\sigma}^{2'})'$ with $\boldsymbol{\gamma}$ adding to it if applicable
Others:	
a_i	the common random initial level for model M_4
b_i	the common random degradation rate for model M_4
σ_a^2	the variance of a_i for model M_4
σ_b^2	the variance of b_i for model M_4
ρ_{ab}	the correlation between a_i and b_i for model M_4

I Description of Data & Codes

The repository (https://github.com/hnasu-code/multi_wiener) contains the following files:

1. "run_Table5.R": The main R file to reproduce Table 5 in the paper.

2. "func_start.R": The R function that finds the helpful starting point for the EM algorithm.

3. "func_M0.R": The R function that performs the EM algorithm for model M0.

4. "func_M2.R": The R function that performs the EM algorithm for model M2.

5. "func_M3.R": The R function that performs the EM algorithm for model M3.

6. "func_M4.R": The R function that performs the EM algorithm for model M4.

7. "IRLED.csv": The CSV file storing the raw IRLED degradation Data.

Instructions: To execute the code, ensure that all the files mentioned above are placed in a common folder. Then, open "run_Table5.R" in either RStudio or R terminal and install all the relevant packages as claimed at the beginning. Run the code line by line under each code section in terms of a corresponding candidate model to obtain the results, which include the log-likelihood and AIC values.

References

Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer.

Li, H., Pan, D., and Chen, C. L. P. (2015). Reliability modeling and life estimation using an expectation maximization based Wiener degradation model for momentum wheels. *IEEE Transactions on Cybernetics*, 45(5):969–977.

Rencher, A. C. and Schaalje, G. B. (2008). *Linear models in statistics*. John Wiley & Sons.