Supplement to "A Multi-level Trend-Renewal Process for Modeling Systems with Recurrence Data"

Zhibing Xu[1], Yili Hong[1], William Q. Meeker[2],
Brock E. Osborn[3], and Kati Illouz[3]

[1]Department of Statistics, Virginia Tech, Blacksburg, VA 24061

[2]Department of Statistics, Iowa State University, Ames, IA, 50011

[3]Applied Statistics Laboratory, GE Global Research Center, Niskayuna, NY, 12309

1 Prediction for the Time-Dependent Covariate

To predict future recurrent events for a system with a time-dependent covariate, it is necessary to have a parametric model for the covariate process. we use a linear mixed effects model to describe the dynamic covariate data as an example. In particular,

$$X_i(t_{ik}^x) = (\beta_x + \nu_i)t_{ik}^x + \epsilon_i(t_{ik}^x) \quad i = 1, \cdots, n, \quad k = 1, \cdots, m_i, \tag{1}$$

where β_x is the coefficient of time, ν_i is the random effect, and $\epsilon_i(t_{ik}^x)$ is the error term. We assume that $\nu_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\nu^2)$, and $\epsilon_i(t_{ik}^x) \overset{\text{iid}}{\sim} \text{N}(0, \sigma_x^2)$ is independent of ν_i. The parameters in (1) are denoted by $\boldsymbol{\theta}^x = (\beta_x, \sigma_\nu, \sigma_x)'$. The estimation of $\boldsymbol{\theta}^x$ in the covariate model can be accomplished by using existing software packages (e.g., using the R function `lme`).

We use an approach that is similar to that used by Hong and Meeker (2013) in covariate prediction but for a different of failure-time model. Let $\boldsymbol{t}_i = (t_{i1}^x, \cdots, t_{im_i}^x)'$, $\boldsymbol{t}_{it^*} = (t_{i,m_{i+1}}^x, \cdots, t_{i,m_i+z_i}^x)'$ be the observed time points before τ_i and the predicted time points during $(\tau_i, \tau_i + t^*]$, respectively. Let $\boldsymbol{X}_i(\boldsymbol{t}_i) = [X_i(t_{i1}^x), \cdots, X_i(t_{im_i}^x)]'$, and $\boldsymbol{X}_i(\boldsymbol{t}_{it^*}) = [X_i(t_{i,m_{i+1}}^x), \cdots, X_i(t_{i,m_i+z_i}^x)]'$ be the corresponding time-dependent covariate processes. Here, z_i is the number of predicted time points for system i. The joint distribution of $\boldsymbol{X}_i(\boldsymbol{t}_i)$ and $\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$ can be expressed as

$$\begin{bmatrix} \boldsymbol{X}_i(\boldsymbol{t}_i) \\ \boldsymbol{X}_i(\boldsymbol{t}_{it^*}) \end{bmatrix} \sim \text{N}\left[\begin{pmatrix} \boldsymbol{t}_i \\ \boldsymbol{t}_{it^*} \end{pmatrix}\beta_x, \begin{pmatrix} \boldsymbol{\Sigma}_{i11} & \boldsymbol{\Sigma}_{i12} \\ \boldsymbol{\Sigma}_{i21} & \boldsymbol{\Sigma}_{i22} \end{pmatrix}\right],$$

where $\boldsymbol{\Sigma}_{i11} = \sigma_\nu^2 \boldsymbol{t}_i \boldsymbol{t}_i' + \sigma_x^2 \boldsymbol{I}_{m_i}$, $\boldsymbol{\Sigma}_{i22} = \sigma_\nu^2 \boldsymbol{t}_{it^*} \boldsymbol{t}_{it^*}' + \sigma_x^2 \boldsymbol{I}_{z_i}$, and $\boldsymbol{\Sigma}_{i12} = \sigma_\nu^2 \boldsymbol{t}_i \boldsymbol{t}_{it^*}'$. Here, $\boldsymbol{I}_{m_i}$ and $\boldsymbol{I}_{z_i}$ are $m_i \times m_i$ and $z_i \times z_i$ identity matrices. The conditional distribution of $\boldsymbol{X}_i(\boldsymbol{t}_i)|\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$ is

$$\mathrm{N}\left(\boldsymbol{t}_{it^*}\beta_x + \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}[\boldsymbol{X}_i(\boldsymbol{t}_i) - \boldsymbol{t}_i\beta_x],\ \boldsymbol{\Sigma}_{i22} - \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}\boldsymbol{\Sigma}_{i12}\right). \tag{2}$$

Based on (2), the time-dependent covariate processes can be predicted.

The derivation of (2) is given as follows. Let $X_i(t_{ij})$ and $X_i(t_{ik})$ denote two random variables of the time-dependent covariate.It is easy to show that $\mathbf{E}[X_i(t_{ij})] = t_{ij}\beta_x$, $\mathbf{E}[X_i(t_{ik})] = t_{ik}\beta_x$, $\mathrm{Var}[X_i(t_{ij})] = \sigma_\nu^2 t_{ij}^2$, $\mathrm{Var}[X_i(t_{ik})] = \sigma_\nu^2 t_{ik}^2$, and

$$\begin{aligned}
\mathrm{Cov}[X_i(t_{ij}), X_i(t_{ij})] =& \mathrm{Cov}[t_{ij}(\beta_x + \nu_i) + \epsilon_i(t_{ij}), t_{ik}(\beta_x + \nu_i) + \epsilon_i(t_{ik})] \\
=& \mathrm{Cov}[t_{ij}\nu_i, t_{ik}\nu_i] \\
=& \sigma_\nu^2 t_{ij} t_{ik}.
\end{aligned}$$

Then, we can easily obtain the variance and covariance expressions for $\boldsymbol{X}_i(\boldsymbol{t}_i)$ and $\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$: $\boldsymbol{\Sigma}_{i11} = \sigma_\nu^2 \boldsymbol{t}_i \boldsymbol{t}_i' + \sigma_x^2 \boldsymbol{I}_{m_i}$, $\boldsymbol{\Sigma}_{i22} = \sigma_\nu^2 \boldsymbol{t}_{it^*} \boldsymbol{t}_{it^*}' + \sigma_x^2 \boldsymbol{I}_{z_i}$, and $\boldsymbol{\Sigma}_{i12} = \sigma_\nu^2 \boldsymbol{t}_i \boldsymbol{t}_{it^*}'$. Based on the joint distribution of $\boldsymbol{X}_i(\boldsymbol{t}_i)$ and $\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$, we can obtain the conditional distribution of $\boldsymbol{X}_i(\boldsymbol{t}_i)|\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$:

$$\mathrm{N}\left(\boldsymbol{t}_{it^*}\beta_x + \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}[\boldsymbol{X}_i(\boldsymbol{t}_i) - \boldsymbol{t}_i\beta_x],\ \boldsymbol{\Sigma}_{i22} - \boldsymbol{\Sigma}_{i21}\boldsymbol{\Sigma}_{i11}^{-1}\boldsymbol{\Sigma}_{i12}\right).$$

2 Subsystem Event Simulations

Because the model for component events depends on the history of subsystem events, the simulation of subsystem events is needed in the prediction of component events. Let $\varsigma_i = \tau_i + t^*$ be the prediction ending time of system i, $\widehat{F}^{s\star}$ be the estimate of renewal distribution function $F^{s\star}$, $\widehat{\Lambda}_i^\star$ be the estimate of $\Lambda_i^\star$, and $\widehat{\Lambda}_i^{\star-1}(\cdot)$ be the corresponding inverse function given $\widehat{\boldsymbol{\theta}}^s$ and $\widehat{\boldsymbol{\theta}}^x$. Here, $\widehat{\boldsymbol{\theta}}^s$ and $\widehat{\boldsymbol{\theta}}^x$ are ML estimates of $\boldsymbol{\theta}^s$ and $\boldsymbol{\theta}^x$, respectively. Based on the definition of the TRP model, the gaps between two consecutive transformed subsystem event times follow distribution $F^{s\star}$. That is, $\Lambda_i^\star(t_{i,j+1}^s) - \Lambda_i^\star(t_{ij}^s) \overset{\text{iid}}{\sim} F^{s\star}$, where $i = 1, \cdots, n$ and $j = 1, 2, \cdots$. The subsystem events can be simulated as follows.

Algorithm S1

1. Simulate a realization of $\boldsymbol{X}_i(\boldsymbol{t}_{it^*})$, the ith time-dependent covariate process, based on $\widehat{\boldsymbol{\theta}}^x$ using the conditional distribution (2).

2. Compute $\widehat{\Lambda}_i^\star(\varsigma_i)$ as the prediction ending time for unit i.

3. Generate a sequence of random variables U_{ij} from distribution $\widehat{F}^{s\star}$ and obtain the sequence of simulated event times in a transformed time scale, $T^*_{ij} = \widehat{\Lambda}^\star_i[t^s_{i,N_{is}(\tau_i)}]+\sum_{k=1}^{j} U_{ik}$, $j = 1, \cdots, C^s_i$, until $T^*_{i,C^s_i+1} > \widehat{\Lambda}^\star_i(\varsigma_i)$. Here, $T^*_{ij}, j = 1, \cdots, C^s_i$ are the event times in the transformed time scale according to the RP($F^{s\star}$) model. Then, C^s_i is the random number of simulated subsystem events for unit i.

4. Compute the simulated subsystem event times $T^s_{ij} = \widehat{\Lambda}^{\star-1}_i(T^*_{ij}),\ j = 1, \cdots, C^s_i$.

5. Repeat steps 1-4 for each system i, where $i = 1, \cdots, n$.

Note that in step 3, the time of the first simulated subsystem event T^s_{i1} should be larger than τ_i, because the simulation is conditioned on the history. Otherwise it needs to be re-simulated.

3 Prediction Interval Computing Algorithm

Algorithm S2

1. Simulate $\widehat{\boldsymbol{\theta}}^{x*}$, $\widehat{\boldsymbol{\theta}}^{s*}$, $\widehat{\boldsymbol{\theta}}^{c*}$, and $\widehat{\upsilon}^*$ from $\mathrm{N}(\widehat{\boldsymbol{\theta}}^x, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}}^x})$, $\mathrm{N}(\widehat{\boldsymbol{\theta}}^s, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}}^s})$, $\mathrm{N}(\widehat{\boldsymbol{\theta}}^c, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}}^c})$ and $\mathrm{N}(\widehat{\upsilon}, \widehat{\sigma}^2_{\widehat{\upsilon}})$, respectively.

2. Replace $\widehat{\boldsymbol{\theta}}^x$ by $\widehat{\boldsymbol{\theta}}^{x*}$, $\widehat{\boldsymbol{\theta}}^s$ by $\widehat{\boldsymbol{\theta}}^{s*}$, $\widehat{\boldsymbol{\theta}}^c$ by $\widehat{\boldsymbol{\theta}}^{c*}$, and $\widehat{\upsilon}$ by $\widehat{\upsilon}^*$, and repeat steps 1-7 in **Algorithm 2** to obtain $\widehat{N}^*_c(t^*; \widehat{\boldsymbol{\theta}}^{c*}, \widehat{\boldsymbol{\theta}}^{s*}, \widehat{\boldsymbol{\theta}}^{x*})$.

3. Repeat steps 1-2 B times to obtain $\widehat{N}^{*(b)}_c(t^*; \widehat{\boldsymbol{\theta}}^{c*}, \widehat{\boldsymbol{\theta}}^{s*}, \widehat{\boldsymbol{\theta}}^{x*})$, where $b = 1, \cdots, B$.

4. The $100(1-\alpha)\%$ PI for N_c is the $(\alpha/2, 1-\alpha/2)$ quantile of the B ordered values of $\widehat{N}^{*(b)}_c(t^*; \widehat{\boldsymbol{\theta}}^{c*}, \widehat{\boldsymbol{\theta}}^{s*}, \widehat{\boldsymbol{\theta}}^{x*})$.

References

Hong, Y. and W. Q. Meeker (2013). Field-failure predictions based on failure-time data with dynamic covariate information. *Technometrics 55*, 135–149.