Statistica Sinica
doi:http://dx.doi.org/10.5705/ss.20XX.XXX

Supplementary Materials for Generalized scale-change models for recurrent event processes under informative censoring

Gongjun Xu[1], Sy Han Chiou[2], Jun Yan[3], Kieren Marr[4], and Chiung-Yu Huang[5].

[1] *University of Michigan,* [2] *University of Texas at Dallas,* [3] *University of Connecticut,*

[4] *Johns Hopkins University, and* [5] *University of California, San Francisco*

Abstract: The Supplementary materials contains the proof of Theorem 1 in Section S1, additional simulation results in Section S2, and identifiability assumption check for the infection data in Sections S3.

S1. Proof of Theorem 1

S1.1 Asymptotic linearity of $\widehat{\Lambda}_n(t;a)$. We first show the asymptotic linearity of $\widehat{\Lambda}_n(t;a)$, which will be used to establish the asymptotic normality of $n^{1/2}(\widehat{\alpha}_n - \alpha)$ and $n^{1/2}(\widehat{\beta}_n - \beta)$. For any fixed a, let $\mathcal{N}_n(t;a) = \frac{1}{n}\sum_{i=1}^n N_i^*(t,a) = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq t \wedge Y_i^*(a)\}$ and recall that

$$\mathcal{R}_n^{(k)}(t;a) = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} X_i^k I\{t_{ij}^*(a) \leq t \leq Y_i^*(a)\}$$

for $k = 0$ and 1. Define stochastic processes $\mathcal{N}(u; a)$ and $\mathcal{R}^{(k)}(u; a)$ as follows:

$$
\begin{aligned}
\mathcal{N}(u; a) &:= E\left\{N_i(ue^{-X_i^\top a} \wedge Y_i)\right\} \\
&= E\left[E\left\{Z_i \Lambda_0(ue^{X_i^\top(\alpha - a)} \wedge Y_i e^{X_i^\top \alpha}) e^{X_i^\top(\beta-\alpha)} \;\middle|\; Z_i, Y_i, X_i\right\}\right] \\
&= E\left(E\left[\int_0^u Z_i I\{Y_i^*(a) \geq v\} \, \mathrm{d}\Lambda_0\{ve^{X_i(\alpha - a)}\} \times e^{X_i^\top(\beta-\alpha)} \;\middle|\; Z_i, Y_i, X_i\right]\right) \\
&= E\left[\int_0^u Z_i I\{Y_i^*(a) \geq v\} \, \mathrm{d}\Lambda_0\{ve^{X_i(\alpha - a)}\} \times e^{X_i^\top(\beta-\alpha)}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{R}^{(k)}(u; a) &:= E\left[X_i^k \textstyle\sum_{j=1}^{m_i} I\{t_{ij}^*(a) \leq u \leq Y_i^*(a)\}\right] \\
&= E\left(E[X_i^k I\{Y_i^*(a) \geq u\} N(ue^{-X_i^\top a}) \mid Y_i, Z_i, X_i]\right) \\
&= E\left[I(Y_i^*(a) \geq u) Z_i \Lambda_0\{ue^{X_i(\alpha - a)}\} X_i^k e^{X_i^\top(\beta-\alpha)}\right].
\end{aligned}
$$

When a equals the true parameter α, we have $\mathcal{N}(u; \alpha) = \int_0^u E[Z_i I\{Y_i^*(\alpha) \geq v\} e^{X_i^\top(\beta-\alpha)}] \, \mathrm{d}\Lambda_0(v)$ and $\mathcal{R}^{(k)}(u; \alpha) = E[Z_i I\{Y_i^*(\alpha) \geq u\} X_i^k e^{X_i^\top(\beta-\alpha)}] \Lambda_0(u)$. We have

$$
H(t) = \log \Lambda_0(t) = -\int_t^\tau \Lambda_0^{-1}(u) \mathrm{d}\Lambda_0(u) = \int_t^\tau (\mathcal{R}^{(0)})^{-1}(u; \alpha) \mathrm{d}\mathcal{N}(u; \alpha).
$$

This also establishes the mean zero property of the process $M_i^*(t)$ in Equation (4) of the main text.

Let $\Lambda_0(t, a) = \exp\{-\int_t^\tau (\mathcal{R}^{(0)})^{-1}(u; a) \mathrm{d}\mathcal{N}(u; a)\}$. When $a = \alpha$, we have $\Lambda_0(t) = \Lambda_0(t, \alpha)$. By the uniform strong law of large numbers (van der Vaart and Wellner, 1996), we have $\mathcal{N}_n(u, a) \to \mathcal{N}(u, a)$ and $\mathcal{R}_n^{(k)}(u, a) \to \mathcal{R}^{(k)}(u, a)$ a.s. uniformly in a and $u \in [0, \tau]$; furthermore we have $\|\mathcal{N}_n(u, a) - \mathcal{N}(u, a)\| = O_p(n^{-1/2})$ and $\|\mathcal{R}_n^{(k)}(u, a) -$

$\mathcal{R}^{(k)}(u,a)\| = O_p(n^{-1/2})$, where $\|\cdot\|$ denotes the supremum norm. Arguing as in the proof of Theorem 1 in Wang et al. (2001) and Xu et al. (2017), we have

$$\begin{aligned}
\int_t^\tau \frac{\mathrm{d}\mathcal{N}_n(u;a)}{\mathcal{R}_n^{(0)}(u;a)} &= \int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)} - \int_t^\tau \frac{\{\mathcal{R}_n^{(0)}(u;a) - \mathcal{R}^{(0)}(u;a)\}\,\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)^2} \\
&\quad + \int_t^\tau \frac{\mathrm{d}\{\mathcal{N}_n(u;a) - \mathcal{N}(u;a)\}}{\mathcal{R}^{(0)}(u;a)} + o_p(n^{-1/2}) \\
&= \int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)} - \frac{1}{n}\sum_{i=1}^n \eta_i(t;a) + o_p(n^{-1/2}),
\end{aligned}$$

where

$$\eta_i(t;a) = \sum_{j=1}^{m_i} \int_t^\tau \frac{I\{t_{ij}^*(a) \le u \le Y_i^*(a)\}\,\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)^2} - \sum_{j=1}^{m_i} \int_t^\tau \frac{\mathrm{d}I\{t_{ij}^*(a) \le u\}}{\mathcal{R}^{(0)}(u;a)}.$$

Note that $E(\eta_i) = 0$. This gives an asymptotic i.i.d. representation of $n^{1/2}\{\widehat{\Lambda}_n(a,t) - \Lambda_0(a,t)\}$:

$$n^{1/2}\{\widehat{\Lambda}_n(t,a) - \Lambda_0(t,a)\} = n^{-1/2}\Lambda_0(t,a)\sum_{i=1}^n \eta_i(t,a) + o_p(1). \tag{S1.1}$$

We next show the asymptotic linearity results for $n^{1/2}\{\widehat{\Lambda}_n(t;a) - \widehat{\Lambda}_n(t;\alpha)\}$. Applying the techniques in the proof of Theorem 1 in Ying (1993), for some positive sequence $d_n \to 0$ and $\|a - \alpha\| \le d_n$, we have uniformly in a and $t \in [0,\tau]$,

$$\begin{aligned}
n^{1/2}\left\{\int_t^\tau \frac{\mathrm{d}\mathcal{N}_n(u;a)}{\mathcal{R}_n^{(0)}(u;a)} - \int_t^\tau \frac{\mathrm{d}\mathcal{N}_n(u;\alpha)}{\mathcal{R}_n^{(0)}(u;\alpha)}\right\} &= n^{1/2}\left\{\int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)} - \int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;\alpha)}{\mathcal{R}^{(0)}(u;\alpha)}\right\} \\
&\quad + o_p(n^{1/2}\|a-\alpha\| + 1).
\end{aligned}$$

Furthermore, we have asymptotic approximation

$$n^{1/2}\left\{\int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;a)}{\mathcal{R}^{(0)}(u;a)} - \int_t^\tau \frac{\mathrm{d}\mathcal{N}(u;\alpha)}{\mathcal{R}^{(0)}(u;\alpha)}\right\} = \kappa(t)^\top n^{1/2}(a-\alpha) + o(n^{1/2}\|a-\alpha\| + 1),$$

where $\kappa(t)$ is the corresponding derivative matrix given by

$$\kappa(t)=\int_t^\tau \frac{E\left[XZI\{Y^*(\alpha)\geq u\}\right]}{\mathcal{R}^{(0)}(u;a)}\,\mathrm{d}\{\lambda_0(u)u\}+\int_t^\tau \left.\frac{\partial E\left[ZI\{Y^*(a)\geq u\}\right]\mathcal{R}^{(0)}(u;a)^{-1}}{\partial a}\right|_{a=\alpha}\mathrm{d}\Lambda_0(u).$$

Therefore, we have uniformly for $\|a-\alpha\|\leq d_n\to 0$ and $t\in[0,\tau]$

$$n^{1/2}\left\{\int_t^\tau \frac{\mathrm{d}\mathcal{N}_n(u;a)}{\mathcal{R}_n^{(0)}(u;a)}-\int_t^\tau \frac{\mathrm{d}\mathcal{N}_n(u;\alpha)}{\mathcal{R}_n^{(0)}(u;\alpha)}\right\}=\kappa(t)^\top n^{1/2}(a-\alpha)+o_p(n^{1/2}\|a-\alpha\|+1).$$

This further implies

$$n^{1/2}\left\{\widehat{H}_n(t,a)-\widehat{H}_n(t,\alpha)\right\}=\kappa(t)^\top n^{1/2}(a-\alpha)+o_p(n^{1/2}\|a-\alpha\|+1)$$

and $n^{1/2}\{\widehat{\Lambda}_n(t,a)-\widehat{\Lambda}_n(t,\alpha)\}=\Lambda_0(t,\alpha)\kappa(t)^\top n^{1/2}(a-\alpha)+o_p(n^{1/2}\|a-\alpha\|+1)$, uniformly for $\|a-\alpha\|\leq d_n\to 0$ and $t\in[0,\tau]$.

S1.2 Asymptotic results of $\widehat{\alpha}_n$ and $\widehat{\Lambda}_n(t,\widehat{\alpha}_n)$. We show the asymptotic normality of $n^{1/2}(\widehat{\alpha}_n-\alpha)$. First, a similar discussion as the previous section gives that

$$\begin{aligned}
&n^{1/2}\left\{\int_t^\tau \frac{\mathcal{R}_n^{(1)}(u;a)\,\mathrm{d}\mathcal{N}_n(u;a)}{\mathcal{R}_n^{(0)}(u;a)}-\int_t^\tau \frac{\mathcal{R}_n^{(1)}(u;\alpha)\,\mathrm{d}\mathcal{N}_n(u;\alpha)}{\mathcal{R}_n^{(0)}(u;\alpha)}\right\}\\
=\;&\{\kappa^{(1)}(t)\}^\top n^{1/2}(a-\alpha)+o(n^{1/2}\|a-\alpha\|+1),
\end{aligned}$$

where $\kappa^{(1)}(t)$ is the corresponding derivative matrix given by

$$\begin{aligned}
\kappa^{(1)}(t)\;=\;&\int_t^\tau \frac{E\left[\mathcal{R}^{(1)}(u;a)XZI\{Y^*(\alpha)\geq u\}\right]}{\mathcal{R}^{(0)}(u;a)}\,\mathrm{d}\{\lambda_0(u)u\}\\
&+\int_t^\tau \left.\frac{\partial E\left[ZI\{Y^*(a)\geq u\}\right]\mathcal{R}^{(1)}(u;a)\mathcal{R}^{(0)}(u;a)^{-1}}{\partial a}\right|_{a=\alpha}\mathrm{d}\Lambda_0(u).
\end{aligned}$$

Therefore from the estimating equation, we have

$$n^{1/2}(\widehat{\alpha}_n-\alpha)=J^{-1}n^{1/2}S_n(\alpha)+o_p(1),$$

where J is defined as $J = \kappa^{(1)}(0)$ and

$$S_n(\alpha) = n^{-1}\sum_{i=1}^{n} e_i(\alpha) + o_p(n^{-1/2}),$$

where $e_i(\alpha) = \sum_j \int_0^\tau X_i dM_{ij}^*(\tau)$. By the central limit theorem, this implies that $n^{1/2}(\widehat{\alpha}_n - \alpha)$ converges weakly to a multivariate normal distribution with mean zero and variance $\Sigma_\alpha = J^{-1}E[e_i(\alpha)e_i(\alpha)^\top](J^{-1})^\top$.

The consistency of $\widehat{H}_n(t, \widehat{\alpha}_n)$ follows from that of $\widehat{\alpha}_n$. From the asymptotic linearity results for $n^{1/2}\{\widehat{H}_n(t, \widehat{\alpha}_n) - \widehat{H}_n(t, \alpha)\}$ and $n^{1/2}\{\widehat{H}_n(t, \alpha) - H(t)\}$ in the preceding section, we have uniformly for $t \in [0, \tau]$

$$\begin{aligned} n^{1/2}\{\widehat{H}_n(t, \widehat{\alpha}_n) - H(t)\} &= n^{1/2}\{\widehat{H}_n(t, \widehat{\alpha}_n) - \widehat{H}_n(t, \alpha)\} + n^{1/2}\{\widehat{H}_n(t, \alpha) - H(t)\} \\ &= n^{-1/2}\textstyle\sum_{i=1}^{n}\{\kappa^{(1)}(t)^\top J^{-1} e_i(\alpha) + \eta_i(t; \alpha)\} + o_p(1). \end{aligned}$$

Applying the functional central limit theorem, we have the weak convergence of $n^{1/2}\{\widehat{H}_n(t, \widehat{\alpha}_n) - H(t)\}$ to a mean-zero Gaussian process for $t \in [0, \tau]$. This also gives the asymptotic normality of the process $n^{1/2}\{\widehat{\Lambda}_n(t, \widehat{\alpha}_n) - \Lambda_0(t)\}$.

S1.3 Asymptotic results of $\widehat{\theta}_n$ From the asymptotic linearity property of $\widehat{\Lambda}_n$ developed in the last section, we have

$$\widehat{\Lambda}_n\{Y_i^*(a)\} - \widehat{\Lambda}_n\{Y_i^*(\alpha)\} = \{1 + o_p(1)\}\Lambda_0\{Y_i^*(\alpha)\}\kappa\{Y_i^*(\alpha)\}^\top Y_i e^{X_i^\top \alpha} X_i^\top (a - \alpha) + o_p(n^{-1/2}).$$

For any sequence $d_n \to 0$, consider r in a neighborhood of θ such that $\|a - \alpha\| \le d_n$

and $\|r-\theta\| \leq d_n$. Thus, the following result holds uniformly for the considered r

$$\begin{aligned} U_n(r,\widehat{\alpha}_n) - U_n(\theta,\alpha) &= \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\}\kappa\{Y_i^*(\alpha)\}^\top Y_i e^{X_i^\top \alpha} X_i^\top(\widehat{\alpha}_n - \alpha) \\ &\quad + \frac{1}{n}\sum_{j=1}^{n} \bar{X}_i^\top \exp(\bar{X}_i^\top\theta)\bar{X}_i^\top(r-\theta) + o_p(n^{-1/2} + \|r-\theta\|) \\ &:= J_1(\widehat{\alpha}_n - \alpha) + J_2(r-\theta) + o_p(n^{-1/2} + \|r-\theta\|), \end{aligned}$$

where J_1 and J_2 are as defined in the above displays. We can further derive the normality of $n^{1/2}U_n(\theta,\alpha)$. In particular, $U_n(\theta,\alpha)$ can be written as

$$\begin{aligned} U_n(\theta,\alpha) &= \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top m_i \widehat{\Lambda}_n^{-1}\{Y_i^*(\alpha)\} - \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top \exp(\bar{X}_i^\top\theta) \\ &= \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top m_i [\widehat{\Lambda}_n^{-1}\{Y_i^*(\alpha)\} - \Lambda_0^{-1}\{Y_i^*(\alpha)\}] \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top [m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} - \exp(\bar{X}_i^\top\theta)]. \end{aligned}$$

From (S1.1), the first part of the above display can be further written as

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top m_i [\widehat{\Lambda}_n^{-1}\{Y_i^*(\alpha)\} - \Lambda_0^{-1}\{Y_i^*(\alpha)\}] \\ = &-\frac{1}{n}\sum_{i=1}^{n} \bar{X}_i^\top m_i \Lambda_0^{-1}\{Y_i^*(\alpha)\} \frac{1}{n}\sum_{j=1}^{n} \eta_j\{Y_i^*(\alpha)\} + o_p(n^{-1/2}) \\ = &-\frac{1}{n}\sum_{i=1}^{n} \int \bar{x}^\top m \Lambda_0^{-1}\{y_i^*(\alpha)\}\eta_i\{y^*(\alpha)\}\,\mathrm{d}V(x,y,m) + o_p(n^{-1/2}), \end{aligned}$$

where $V(x,y,m)$ denotes the joint distribution function of (X,Y,m) (e.g., Proof of (9) in Wang et al., 2001). Therefore,

$$n^{1/2}U_n(\theta,\alpha) = n^{-1/2}\textstyle\sum_{i=1}^{n} d_i(\theta,\alpha) + o_p(1)$$

where

$$d_i(\theta,\alpha) = -\int \bar{x}^\top m\Lambda_0^{-1}\{y_i^*(\alpha)\}\eta_i\{y^*(\alpha)\}\,\mathrm{d}V(x,y,m) + \bar{X}_i^\top[m_i\Lambda_0^{-1}\{Y_i^*(\alpha)\} - \exp(\bar{X}_i^\top\theta)].$$

This gives the normality of $n^{1/2}U_n(\theta,\alpha)$ with asymptotic mean 0 and covariance matrix $E[d_i(\theta,\alpha)d_i(\theta,\alpha)^\top]$.

Together with the result that $n^{1/2}(\widehat{\alpha}_n - \alpha) = J^{-1}n^{-1/2}\sum_{i=1}^n e_i(\alpha) + o_p(1)$, we have

$$n^{1/2}(\widehat{\theta} - \theta) = n^{-1/2}J_2^{-1}\textstyle\sum_{i=1}^n\{d_i(\theta,\alpha) - J_1J^{-1}e_i(\alpha)\} + o_p(1)$$

and it converges weakly to a multivariate normal distribution with mean zero and variance

$$\Sigma_\theta = J_2^{-1}E[\{d_i(\theta,\alpha) - J_1J^{-1}e_i(\alpha)\}\{d_i(\theta,\alpha) - J_1J^{-1}e_i(\alpha)\}^\top](J_2^{-1})^\top.$$

In particular, $\{n^{1/2}(\widehat{\alpha}_n - \alpha), n^{1/2}(\widehat{\theta} - \theta)\}$ jointly converges weakly to a multivariate normal distribution with mean zero and covariance matrix induced by the above linear asymptotic approximations that

$$\Sigma(\alpha,\theta) = \begin{pmatrix} \Sigma_\alpha & \mathrm{Cov}(\alpha,\theta) \\ \mathrm{Cov}(\alpha,\theta) & \Sigma_\theta \end{pmatrix},$$

where $\mathrm{Cov}(\alpha,\theta) = \mathrm{Cov}[J^{-1}e_i(\alpha), J_2^{-1}\{d_i(\theta,\alpha) - J_1J^{-1}e_i(\alpha)\}]$. As a consequence, we have the asymptotic normality of $\{n^{1/2}(\widehat{\alpha}_n - \alpha), n^{1/2}(\widehat{\beta} - \beta)\}$ with mean zero and covriance matrix

$$\Sigma(\alpha,\beta) = \begin{pmatrix} I_p & 0 & 0_p \\ I_p & 0 & I_p \end{pmatrix}\Sigma(\alpha,\theta)\begin{pmatrix} I_p & 0 & 0_p \\ I_p & 0 & I_p \end{pmatrix}^\top.$$

S2. Additional simulation results

Table 1 presents the simulation results under an accelerated rate model. For this scenario, regardless of the choice of the initial value, the proposed estimator is fairly unbiased, with the proposed variance estimator in close agreement to the empirical counterpart. The corresponding empirical coverage probabilities are reasonably close to the nominal level of 95%. On the other hand, the estimator of Sun and Su (2008) requires the initial value to be specified at the true value, in which case it yields moderate bias when the noninformative censoring assumption is violated.

When the data generating model reduces to the accelerated means model ($\alpha = \beta$), in addition to the proposed estimator and the conventional estimator of Sun and Su (2008), we also report the result from the estimator of Xu et al. (2017). Table 2 shows the proposed estimator outperforms the conventional estimator of Sun and Su (2008) under both the noninformative and informative censoring setting. The estimator of Xu et al. (2017) also yields small biases. The standard errors of the estimator of Xu et al. (2017) were obtained from an efficient resampling variance estimator that shares the similar spirit as the proposed version. Comparing to the proposed estimator, the estimator of Xu et al. (2017) has a slightly larger bias but smaller standard errors. This biases-efficiency trade-off comes from the additional model assumption on $\alpha = \beta$ that is imposed in Xu et al. (2017). Nonetheless, both of these estimators accommodate informative censoring and yield comparable coverage probabilities. Figure 1 presents the estimates and the empirical pointwise 95% confidence intervals for the baseline cumulative rate function with $n = 200$. The averages of the estimated baseline cumulative rate function are indistinguishable from the truth for the cases considered.

We extended the simulation studies to explore scenarios with different variance configuration for the subject-specific latent variable. Specifically, we considered scenarios with Z generated from gamma distributions with mean 1 and variance 0.5 or 1. Since the proposed estimator performs well regardless of the choice of the initial value, we only present the results that use the zero vector as the initial in Table 3. The standard errors were estimated through the proposed resampling approach with 200 bootstrap size. For all scenarios considered, our estimator remains virtually unbiased, with estimated standard errors reasonably close to the empirical standard errors. The magnitude of the estimated standard errors seem to increase with the variance of the frailty variable, but the empirical coverage rates remain close to the nominal level in all scenarios.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Figure 1 about here.]

A small-scale simulation study was conducted to examine the performance of the proposed method under scenarios where a smaller number of events are observed. We considered the simulation settings described in the manuscript, but set $\tau = 10$ and generated the censoring time from an exponential distribution with mean $7e^{-X_1}/Z$. Simulation results with $\alpha = \beta = (-1, 1)^\top$ are presented in Table 4. Under this setting, the average number of observed recurrent event was about 0.6 per subject, which is smaller than what we observed in the two transplant cohorts. The proposed estimators are virtually unbiased with only a few exceptions with $n = 200$ and the bias the exception cases quickly diminishes as the sample size increases. The estimated standard errors are reasonably close to the empirical standard error, with better agreement with $n = 400$. The coverage rates are closer to the nominal level of 95% when $n = 400$. These results suggest that our model gives valid inference even when the average number of observed recurrent event is small.

[Table 4 about here.]

Lastly, we also considered situations when the recurrent events were generated from a non-Poisson process given the frailty by altering the interarrival time distribution from exponential and adding a second latent variable to the rate function. In particular, we generated the recurrent events from the inversion algorithm for nonstationary renewal processes (e.g., Gerhardt and Nelson, 2009). Let G be the distribution of the scaled interarrival time such that the resulting renewal process has rate 1. For example, G can be unit exponential (for Poisson process) or uniform over $(0, 2)$ (for a non-Poisson process). Let $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u)\,\mathrm{d}u$, where

$$\lambda_{ij}(t) = Z_i \zeta_{ij} \lambda_0(te^{\alpha_1 X_1 + \alpha_2 X_2}) e^{\beta_1 X_1 + \beta_2 X_2},$$

$\lambda_0(t) = 1/(5+t)$, and ζ_{ij}'s are additional latent variables, independent of Z_i, following a gamma distribution with mean 1 and variance 0.5. If ζ_{ij}'s are dependent within each i, the process allows for association of the arrival time of the recurrent events within the a subject. The data generation algorithm is outlined as follows:

Step 1. Initialize $\epsilon_0 = 0$.

Step 2. Generate a random number η from G.

Step 3. Set $t_{ij} = \Lambda_{ij}^{-1}(\epsilon_j)$, where $\epsilon_j = \epsilon_{j-1} + \eta$, for $j \geq 1$, and $\Lambda_{ij}^{-1}(\cdot)$ is the inverse of $\Lambda_{ij}(\cdot)$.

Step 4. Repeat Steps 2 and 3 until $t_{i,m_i+1} \equiv t_{i,j+1} > Y_i$.

The sequence $\{t_{ij}, j = 1, \ldots, m_i\}$ is then the observed recurrent event times for the ith subject. In this algorithm, when $\zeta_{ij} \equiv 1$ and G is unit exponential, the process reduces to the conditional Poisson process given Z_i; otherwise, the process is not Poisson conditional on Z_i. In the simulation, we set G to be uniform over $(0, 2)$. We used the same configurations used in the manuscript to generate X_i, Z_i, and the censoring time. The average number of the observed recurrent events in this simulation ranges from 2.6 to 5.1.

[Table 5 about here.]

Table 5 summarizes the results based on 1000 replications. For all scenarios, our estimator yields small bias, with estimated standard errors reasonably close to the empirical standard errors. The empirical coverage probabilities are somewhat lower than the nominal level under the accelerated rate model, with $\alpha = (-1, -1)^\top, \beta = (0, 0)^\top$ when $n = 200$. As n increases to 400, however, the empirical coverage probabilities become closer to the anticipated level of 95%. Overall, our estimator remains satisfactory under the non-Poisson settings.

S3. Assessment of the identifiability assumption for the infection datasets

Under the Weibull model, $\log\{\Lambda_0(t)\}$ is linear in $\log(t)$. This motivates a check for model identifiability by assessing whether $\log\{\hat{\Lambda}_0(t)\}$ and $\log(t)$ form a linear pattern. Under the Weibull model, Model (1) reduces to the special case of the Cox-type model considered by Wang et al. (2001), and $\hat{\Lambda}_0(t)$ can be obtained using their conditional likelihood method; Figure 2 shows $\log\{\widehat{\Lambda}_0(t)\}$ versus $\log(t)$ for the kidney transplant cohort and the HSCT cohort. In both cases, the linearity does not seem to be appropriate. A formal test would be of future interest.

[Figure 2 about here.]

References

Gerhardt, I. and Nelson, B. L. (2009), "Transforming renewal processes for simulation of nonstationary arrival processes," *INFORMS Journal on Computing*, 21, 630–640.

Sun, L. and Su, B. (2008), "A class of accelerated means regression models for recurrent event data," *Lifetime Data Analysis*, 14, 357–375.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence*, Springer.

Wang, M.-C., Qin, J., and Chiang, C.-T. (2001), "Analyzing recurrent event data with informative censoring," *Journal of the American Statistical Association*, 96, 1057–1065.

Xu, G., Chiou, S. H., Huang, C.-Y., Wang, M.-C., and Yan, J. (2017), "Joint scale-change models for recurrent events and failure time," *Journal of the American Statistical Association*, 112, 794–805.

Ying, Z. (1993), "A large sample study of rank estimation for censored regression data," *The Annals of Statistics*, 21, 76–99.

Table 1: Simulation results with $\alpha = (-1, -1)^\top$ and $\beta = (0, 0)^\top$. Columns without an asterisk ($*$) present results using zero vector as initial value; Columns with an asterisk present results using true value as initial value; Bias is the empirical bias; ESE is the empirical standard error; ASE is the average standard error; CP is the empirical coverage probability (%) of 95% confidence intervals.

		Proposed						Sun and Su (2008)					
n		Bias	ESE	ASE	CP	Bias*	ESE*	Bias	ESE	ASE	CP	Bias*	ESE*
							$Z = 1$						
200	α_1	−0.024	0.369	0.350	92.5	−0.018	0.375	0.754	0.388	0.127	11.3	−0.012	0.206
	α_2	−0.025	0.336	0.310	93.4	0.030	0.335	0.736	0.364	0.136	12.3	−0.017	0.207
	β_1	−0.013	0.236	0.229	94.1	0.013	0.244	0.470	0.260	0.110	19.5	−0.014	0.160
	β_2	−0.017	0.217	0.208	93.7	0.023	0.215	0.468	0.242	0.109	17.3	−0.010	0.166
400	α_1	−0.007	0.257	0.249	93.6	0.012	0.232	0.688	0.460	0.172	24.4	−0.011	0.189
	α_2	0.001	0.222	0.222	94.3	−0.003	0.216	0.619	0.417	0.160	23.4	0.004	0.192
	β_1	−0.007	0.170	0.158	94.5	0.012	0.156	0.429	0.293	0.123	28.3	−0.002	0.133
	β_2	−0.004	0.148	0.147	94.1	−0.002	0.148	0.395	0.273	0.118	26.5	−0.009	0.136
							$Z \sim$ Gamma(4, 4)						
200	α_1	−0.011	0.365	0.354	92.3	−0.005	0.368	0.736	0.341	0.119	8.3	−0.106	0.267
	α_2	−0.011	0.331	0.315	92.4	0.023	0.331	0.705	0.360	0.126	11.9	−0.085	0.236
	β_1	−0.012	0.233	0.233	94.0	0.010	0.253	0.420	0.250	0.117	25.2	−0.091	0.207
	β_2	−0.010	0.226	0.221	92.8	0.009	0.226	0.442	0.247	0.121	20.8	0.096	0.188
400	α_1	−0.009	0.260	0.255	93.1	−0.026	0.266	0.718	0.369	0.145	14.1	−0.137	0.230
	α_2	−0.010	0.235	0.223	93.6	0.014	0.223	0.627	0.387	0.149	19.0	−0.086	0.205
	β_1	−0.003	0.173	0.168	94.8	−0.000	0.172	0.403	0.253	0.115	26.4	−0.080	0.166
	β_2	−0.006	0.154	0.154	94.8	0.007	0.149	0.387	0.256	0.119	26.8	−0.088	0.153

Table 2: Simulation results with $\alpha = \beta = (-1, 1)^\top$. Columns without an asterisk (*) present results using zero vector as initial value; Columns with an asterisk present results using true value as initial value; Bias is the empirical bias; ESE is the empirical standard error; ASE is the average standard error; CP is the empirical coverage probability (%) of 95% confidence intervals.

		Proposed						Sun and Su (2008)						Xu et al. (2017)			
n		Bias	ESE	ASE	CP	Bias*	ESE*	Bias	ESE	ASE	CP	Bias*	ESE*	Bias	ESE	ASE	CP
								$Z = 1$									
200	α_1	−0.016	0.333	0.329	93.7	−0.008	0.384	0.911	0.132	0.092	0.9	0.008	0.213	0.046	0.255	0.242	93.1
	α_2	−0.008	0.316	0.303	93.1	−0.011	0.320	−0.790	0.091	0.086	1.4	0.002	0.202	−0.011	0.231	0.220	93.8
	β_1	−0.011	0.230	0.226	94.0	0.018	0.243	0.616	0.099	0.107	1.8	0.005	0.180				
	β_2	−0.004	0.215	0.211	94.4	0.008	0.205	−0.524	0.123	0.104	2.7	0.010	0.171				
400	α_1	0.003	0.244	0.235	93.8	−0.008	0.253	0.927	0.128	0.131	1.3	−0.009	0.154	0.023	0.192	0.187	93.9
	α_2	−0.006	0.220	0.216	94.0	−0.016	0.238	−0.773	0.110	0.118	1.1	0.005	0.154	−0.010	0.175	0.166	94.1
	β_1	−0.002	0.165	0.160	94.2	0.008	0.162	0.628	0.086	0.111	1.7	0.003	0.127				
	β_2	0.001	0.156	0.151	94.8	−0.001	0.157	−0.523	0.112	0.104	2.4	0.003	0.124				
								$Z \sim \text{Gamma}(4, 4)$									
200	α_1	0.001	0.358	0.338	92.2	−0.047	0.416	0.925	0.136	0.091	0.3	−0.084	0.219	0.038	0.269	0.246	93.9
	α_2	0.003	0.325	0.308	92.9	−0.042	0.345	−0.790	0.093	0.089	1.1	0.087	0.198	−0.012	0.241	0.221	94.3
	β_1	−0.016	0.241	0.238	94.3	0.000	0.260	0.603	0.105	0.117	1.5	0.107	0.193				
	β_2	0.005	0.222	0.217	94.6	−0.007	0.230	−0.514	0.138	0.118	2.9	−0.091	0.179				
400	α_1	−0.003	0.253	0.248	94.2	−0.013	0.256	0.935	0.123	0.120	0.6	−0.136	0.183	0.021	0.200	0.182	93.5
	α_2	−0.002	0.227	0.222	94.4	−0.018	0.227	−0.765	0.118	0.116	1.2	0.097	0.161	−0.009	0.182	0.175	94.3
	β_1	−0.009	0.168	0.167	94.8	0.002	0.177	0.607	0.094	0.112	1.2	0.103	0.150				
	β_2	0.001	0.161	0.156	94.7	−0.006	0.162	−0.505	0.120	0.111	2.8	−0.083	0.136				

Table 3: Additional simulation results with different variance for the subject-specified frailty variable. Bias is the empirical bias; ESE is the empirical standard error; ASE is the average standard error; CP is the empirical coverage probability (%) of 95% confidence intervals.

		$Z \sim$ Gamma(1, 1)				$Z \sim$ Gamma(2, 2)			
n		Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
		$\alpha = (-1, -1)^\top, \beta = (1, 1)^\top$							
200	α_1	−0.007	0.357	0.334	92.4	−0.001	0.314	0.309	93.2
	α_2	0.012	0.320	0.299	91.8	−0.006	0.282	0.270	93.4
	β_1	−0.006	0.343	0.308	93.4	0.026	0.291	0.276	92.9
	β_2	0.004	0.317	0.306	94.2	0.009	0.271	0.264	92.7
400	α_1	−0.011	0.225	0.223	95.0	−0.010	0.236	0.227	93.4
	α_2	−0.005	0.207	0.199	95.0	−0.011	0.206	0.199	93.5
	β_1	−0.004	0.250	0.238	93.6	−0.000	0.179	0.176	95.8
	β_2	−0.003	0.251	0.232	94.0	0.001	0.180	0.175	94.0
		$\alpha = (0, 0)^\top, \beta = (-1, -1)^\top$							
200	α_1	−0.003	0.177	0.162	91.4	−0.019	0.173	0.160	92.0
	α_2	0.003	0.163	0.158	93.0	0.013	0.158	0.143	91.6
	β_1	0.020	0.214	0.198	91.8	−0.005	0.166	0.159	92.8
	β_2	0.027	0.201	0.189	92.6	0.016	0.161	0.156	93.6
400	α_1	0.002	0.111	0.107	92.8	−0.014	0.114	0.103	92.6
	α_2	0.007	0.106	0.102	94.6	0.001	0.102	0.099	93.6
	β_1	0.014	0.155	0.147	92.2	−0.003	0.126	0.119	94.7
	β_2	0.022	0.144	0.136	93.4	0.006	0.115	0.111	94.6
		$\alpha = (-1, -1)^\top, \beta = (0, 0)^\top$							
200	α_1	0.010	0.413	0.393	91.4	0.007	0.401	0.387	91.6
	α_2	0.009	0.355	0.337	92.0	−0.005	0.359	0.335	92.0
	β_1	0.018	0.320	0.342	95.4	0.013	0.271	0.277	95.2
	β_2	0.009	0.269	0.285	95.2	−0.002	0.259	0.243	94.2
400	α_1	−0.014	0.298	0.273	92.8	−0.003	0.278	0.268	93.8
	α_2	−0.001	0.248	0.235	92.6	−0.007	0.234	0.231	92.6
	β_1	−0.003	0.197	0.196	94.1	−0.003	0.200	0.190	94.4
	β_2	0.001	0.185	0.182	94.4	0.001	0.168	0.164	94.1
		$\alpha = (-1, 1)^\top, \beta = (-1, 1)^\top$							
200	α_1	−0.021	0.397	0.375	91.1	−0.028	0.374	0.355	91.4
	α_2	0.011	0.355	0.334	91.8	0.008	0.322	0.316	95.0
	β_1	−0.005	0.283	0.267	92.0	0.007	0.257	0.247	94.8
	β_2	−0.003	0.258	0.252	92.8	0.005	0.236	0.229	93.8
400	α_1	−0.003	0.265	0.252	95.4	−0.016	0.245	0.245	94.4
	α_2	0.011	0.236	0.228	94.0	0.010	0.222	0.225	94.6
	β_1	−0.013	0.202	0.191	93.0	−0.006	0.181	0.177	94.0
	β_2	0.010	0.182	0.177	93.6	0.007	0.164	0.161	95.0

Table 4: Additional simulation results with small average number of observed recurrent event. Bias is the empirical bias; ESE is the empirical standard error; ASE is the average standard error; CP is the empirical coverage probability (%) of 95% confidence intervals.

		$Z \sim$ Gamma(1, 1)				$Z \sim$ Gamma(4, 4)			
n		Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
200	α_1	-0.065	0.635	0.626	89.0	-0.080	0.691	0.679	87.3
	α_2	0.060	0.575	0.535	90.9	0.060	0.591	0.581	88.9
	β_1	-0.008	0.382	0.404	93.1	-0.013	0.472	0.514	92.6
	β_2	-0.002	0.317	0.337	93.4	-0.009	0.399	0.397	94.4
400	α_1	-0.050	0.460	0.440	91.7	-0.043	0.484	0.441	90.0
	α_2	0.060	0.398	0.388	92.0	0.034	0.402	0.392	92.3
	β_1	-0.003	0.257	0.266	95.1	0.021	0.334	0.308	94.4
	β_2	0.007	0.207	0.218	95.2	-0.007	0.242	0.251	96.0

Table 5: Additional simulation results with recurrent events generated from a non-Poisson process. Bias is the empirical bias; ESE is the empirical standard error; ASE is the average standard error; CP is the empirical coverage probability (%) of 95% confidence intervals.

		$Z = 1$				$Z \sim$ Gamma$(4, 4)$			
n		Bias	ESE	ASE	CP	Bias	ESE	ASE	CP
				$\alpha = (-1, -1)^\top, \beta = (1, 1)^\top$					
200	α_1	0.018	0.447	0.442	90.8	0.035	0.442	0.440	90.9
	α_2	0.038	0.380	0.369	91.0	0.022	0.378	0.377	91.6
	β_1	0.019	0.162	0.176	95.1	0.016	0.191	0.210	95.2
	β_2	0.028	0.142	0.156	95.3	0.020	0.182	0.198	95.5
400	α_1	0.015	0.323	0.321	91.6	0.022	0.326	0.329	92.0
	α_2	0.031	0.255	0.245	92.5	0.012	0.271	0.278	93.9
	β_1	0.011	0.106	0.121	95.8	0.010	0.140	0.151	95.4
	β_2	0.021	0.092	0.106	95.1	0.017	0.130	0.139	95.3
				$\alpha = (0, 0)^\top, \beta = (-1, -1)^\top$					
200	α_1	−0.003	0.251	0.253	95.1	−0.007	0.270	0.263	95.3
	α_2	−0.009	0.240	0.239	95.2	−0.011	0.246	0.245	95.1
	β_1	−0.016	0.130	0.125	93.8	−0.023	0.153	0.144	93.1
	β_2	−0.020	0.122	0.120	93.5	−0.020	0.147	0.138	93.2
400	α_1	−0.009	0.179	0.177	95.5	−0.012	0.183	0.183	95.4
	α_2	−0.010	0.164	0.165	95.4	−0.008	0.170	0.173	95.5
	β_1	−0.017	0.091	0.088	94.6	−0.019	0.104	0.101	93.8
	β_2	−0.020	0.086	0.085	94.9	−0.018	0.100	0.099	94.9
				$\alpha = (-1, -1)^\top, \beta = (0, 0)^\top$					
200	α_1	0.067	0.431	0.403	88.8	0.051	0.437	0.411	88.1
	α_2	0.061	0.350	0.335	89.7	0.043	0.396	0.353	87.4
	β_1	−0.026	0.220	0.254	96.0	−0.050	0.241	0.265	95.4
	β_2	−0.001	0.184	0.205	95.3	−0.014	0.195	0.226	96.8
400	α_1	0.050	0.312	0.298	91.1	0.038	0.324	0.323	91.8
	α_2	0.038	0.268	0.265	91.7	0.043	0.285	0.273	90.6
	β_1	−0.022	0.177	0.177	95.2	−0.046	0.193	0.198	95.8
	β_2	−0.012	0.133	0.138	95.9	−0.024	0.150	0.159	95.4
				$\alpha = (-1, 1)^\top, \beta = (-1, 1)^\top$					
200	α_1	0.008	0.318	0.320	92.8	0.001	0.327	0.329	93.6
	α_2	−0.013	0.285	0.283	92.5	−0.002	0.297	0.294	92.7
	β_1	−0.010	0.163	0.162	93.1	−0.019	0.172	0.178	95.8
	β_2	−0.001	0.136	0.138	95.1	−0.002	0.146	0.153	96.2
400	α_1	−0.012	0.220	0.232	95.1	0.001	0.239	0.240	95.3
	α_2	−0.005	0.191	0.205	95.5	0.002	0.203	0.213	95.2
	β_1	−0.016	0.106	0.112	96.9	−0.009	0.116	0.122	94.9
	β_2	−0.001	0.094	0.099	94.9	0.005	0.107	0.108	95.6

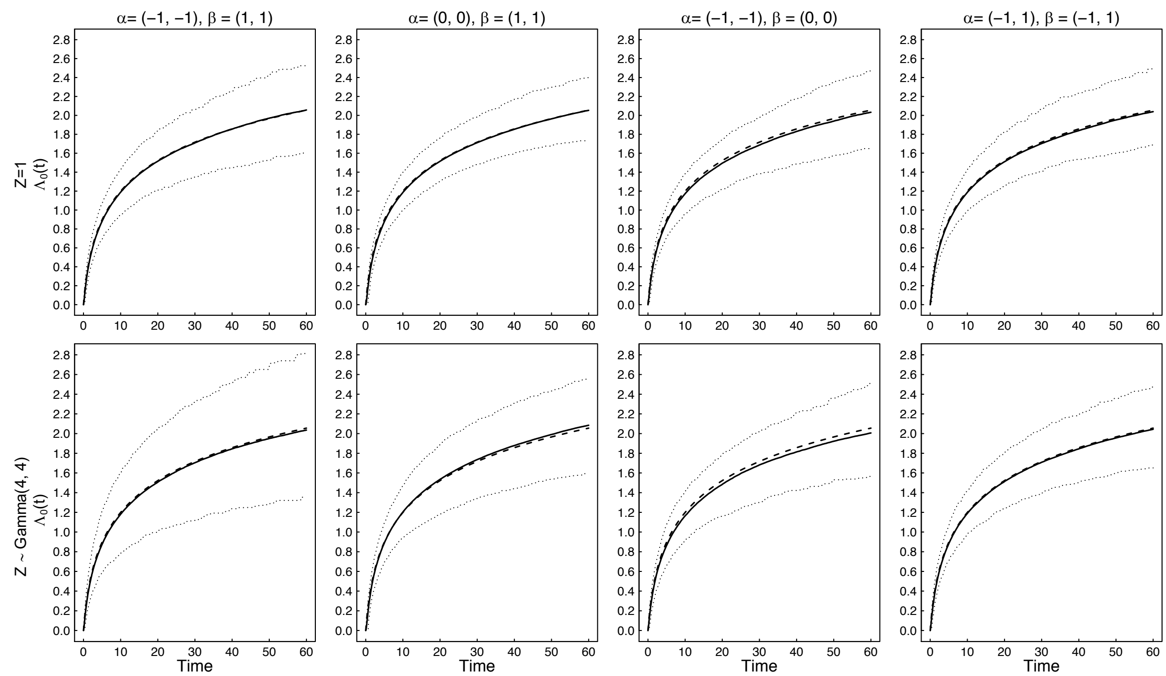
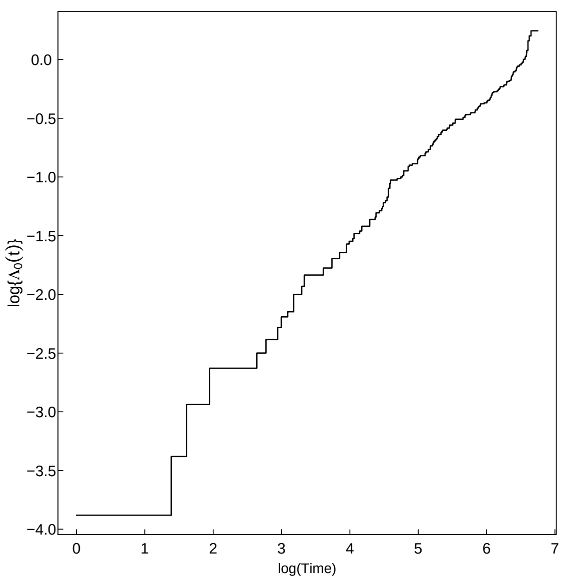
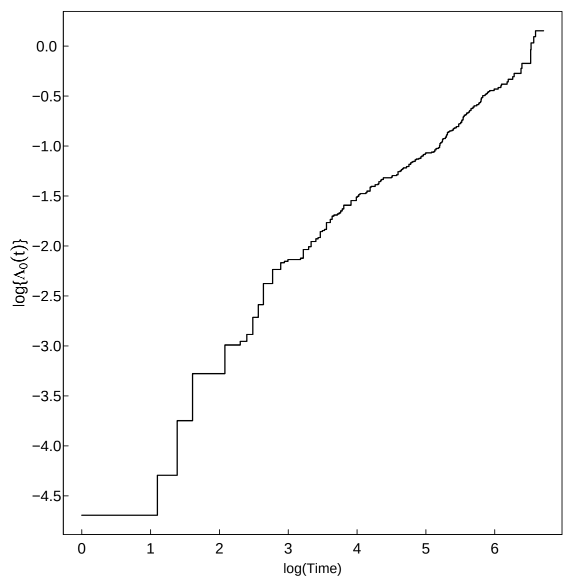


Figure 1: Plots of estimated $\widehat{\Lambda}_0(t)$ with the empirical pointwise 95% conference intervals for $n = 200$. Solid lines (—) present the empirical average of the estimated $\widehat{\Lambda}_0(t)$; Dashed lines (- - -) present the true curve, $\Lambda_0(t) = 0.5 \log(1 + t)$; Dotted lines (·····) present the empirical pointwise confidence intervals obtained from the 2.5th and the 97.5th quantile of the estimated $\widehat{\Lambda}_0(t)$.



(a) Kidney transplant cohort.



(b) HSCT cohort.

Figure 2: Plot of the estimated log baseline rate function versus log event time.