Statistica Sinica **23** (2013), 383-408
doi:http://dx.doi.org/10.5705/ss.2011.151

A FRAILTY MODEL APPROACH FOR REGRESSION ANALYSIS OF BIVARIATE INTERVAL-CENSORED SURVIVAL DATA

Chi-Chung Wen and Yi-Hau Chen

Tamkang University and Academia Sinica

Abstract: Owing to the fact that general semiparametric inference procedures are still underdeveloped for multivariate interval-censored event time data, we propose semiparametric maximum likelihood estimation for the gamma-frailty Cox model under mixed-case interval censoring. We establish the consistency of the semiparametric maximum likelihood estimator (SPMLE) for the model parameters, including the regression coefficients and the cumulative hazard functions in the Cox model, and the variance of the gamma frailty. The SPMLEs of the cumulative hazard functions are shown to have a $n^{1/3}$-rate of convergence, while those of the regression coefficients and the frailty variance have a $n^{1/2}$-rate of convergence; here n denotes the number of study units. The asymptotic normality of the regression coefficients and the frailty variance is also established, with the asymptotic variance given by the inverse of the efficient Fisher information matrix. A profile-likelihood approach is proposed for estimating the asymptotic variance. Based on the self-consistency equations and the contraction principle, we propose a stable and efficient computation algorithm. Simulation results reveal that the large sample theories work quite well in finite samples. We analyze a dataset from an AIDS clinical trial by the proposed methods to assess the effects of the baseline CD4 cell counts on the times to CMV shedding in blood and urine.

Key words and phrases: Correlated data, interval censoring, proportional hazards, self-consistency.

1. Introduction

Data on survival or event time are often subject to censoring due to limitations in the observational process. For example, right censoring occurs when time to the event is beyond the end of observation, while interval censoring occurs when the observation is only made at several examination times, and hence one can only know that the event time lies in some interval bracketed by two examination times. The incomplete nature of censored event time data complicates the subsequent statistical analysis, including event time regression analysis where the covariate effects on the event time are to be assessed. In particular, interval-censored data generally create more difficulties than right-censored data in both theory and computation. For instance, for the Cox model under the "case k" interval censoring where there are k examination times per subject, it has been shown that the maximum likelihood estimator of the regression parameter is asymptotically normal and efficient, but that of the baseline cumulative hazard function has only a $n^{1/3}$-rate of convergence (Huang and Wellner (1997)), slower than the $n^{1/2}$-rate achieved with right-censored data (Andersen and Gill (1982)). Also, in contrast to convenient computation via the partial likelihood (Cox (1972)) under right censoring, the computation of the maximum likelihood estimator for the Cox model with interval censoring may involve a high dimensional Newton-Raphson iteration (Finkelstein (1986)). The monograph by Sun (2006) provides a comprehensive review of the problems on event time analysis with interval-censored data.

The results mentioned above pertain to univariate interval-censored data. The challenges from interval censoring become even more prominent when multivariate event time data are considered, because correlations among multiple event times cause a further complication. Some marginal regression approaches based on working independence have been proposed for the Cox model (Goggins and Finkelstein (2000), Kim and Xue (2002)), the proportional odds model (Chen, Tong, and Sun (2007)), and the additive hazards model (Tong, Chen, and Sun (2008)). It is expected that such methods lose information since correlations among event times are not accounted for. The full-likelihood approaches based on the frailty Cox model have also been developed, but only for the specific "case 1" interval censoring where there is only one examination time for each event per subject (Chen, Tong, and Sun (2009), Wen and Chen (2011)), or for restricted maximum likelihood estimation (Xiang, Ma, and Yau (2011)). A semiparametric maximum likelihood approach under general "mixed-case" interval censoring, where the number of examination times for each event per subject can vary randomly, is still lacking for the analysis of multivariate interval-censored data.

In this work we consider semiparametric maximum likelihood estimation for the gamma-frailty Cox model with bivariate mixed-case interval-censored event time data. Our motivation comes from a dataset from the ACTG 181 clinical trial on HIV-infected patients (Goggins and Finkelstein (2000), Sun (2006)), where the effects of baseline CD4 cell counts on the times to shedding of cytomegalovirus (CMV) in the urine and blood are of interest, and data on CMV shedding times are subject to interval censoring since they are determined only at intermittent clinic visits. We formally establish the consistency of the semiparametric maximum likelihood estimator (SPMLE) for the model parameters, including the regression coefficients and the cumulative hazard functions in the Cox model, and the variance parameter for the gamma frailty. In particular, the SPMLEs of the cumulative hazard functions are shown to have a $n^{1/3}$-rate of convergence, while the SPMLEs of the regression coefficients and the frailty variance parameter have a $n^{1/2}$-rate of convergence; hereinafter n denotes the number of study units. The asymptotic normality of the finite-dimensional parameters, including the regression coefficients and the frailty variance, is also established, with the asymptotic variance given by the inverse of the efficient Fisher information matrix. A profile-likelihood approach is proposed for estimating the asymptotic variance. Based on a set of self-consistency equations and the contraction principle, we propose a stable and efficient computation algorithm for semiparametric maximum likelihood estimation of the gamma-frailty Cox model under general types of interval censoring, extending an earlier version of the algorithm proposed by Wen and Chen (2011) for the "case 1" interval-censored or "current status" data. Simulation results reveal that the large sample theories developed for the SPMLE work quite well in the finite sample setting. We assess the effects of the baseline CD4 cell counts on the times to CMV shedding in blood and urine using the proposed method.

2. The Data and Model

Let T_1 and T_2 denote two possibly correlated failure times in one study unit (e.g. subject or family), and Z_1 and Z_2 the vectors of covariates that may affect T_1 and T_2, respectively. To assess the effects of Z_j on T_j $(j = 1, 2)$, while accounting for correlation between T_1 and T_2, the gamma-frailty Cox model proposed by Vaupel, Manton, and Stallard (1979) may be utilized. Suppose η is a gamma random variable with mean 1 and variance $\gamma > 0$. We consider three types of gamma frailty models that assume that, conditional on (η, Z_1, Z_2), T_1 and T_2 are independent with the marginal cumulative hazard function of T_j $(j = 1, 2)$ given by one of

$$\eta \exp(\beta' Z_j)\Lambda(t), \tag{2.1}$$

$$\eta \exp(\beta' Z_j)\Lambda_j(t), \tag{2.2}$$

$$\eta \exp(\beta_j' Z_j)\Lambda_j(t). \tag{2.3}$$

In these models, β and the β_j's denote vectors of unknown regression parameters, and Λ and the Λ_j's denote functions of unspecified baseline marginal cumulative hazards. Hence the correlation between T_1 and T_2 is accounted for by the shared but unobserved frailty η, with a larger value of γ corresponding to a stronger correlation. Model (2.1) assumes homogeneous baseline hazards as well as covariate effects for T_1 and T_2; model (2.2) assumes homogeneous baseline hazards but heterogeneous covariate effects; model (2.3) assumes heterogeneous baseline hazards as well as covariate effects.

In bivariate interval-censored data, T_j, $j = 1, 2$, is not observed exactly but is known only to occur within some censoring interval $(L_j, R_j]$ with $L_j < R_j$. To define how the pair (L_j, R_j) is generated for each j, we consider the 'mixed-case' interval censoring as defined in Schick and Yu (2000). Let K_j be a random positive integer denoting the number of examination times for T_j in a study unit, and $U_j = \{U^{(j)}_{K_j,l} : l = 1, \ldots, K_j, K_j = 1, 2, \ldots\}$ a triangular array of random examination times with $U^{(j)}_{K_j,1} < \cdots < U^{(j)}_{K_j,K_j}$. Then $L_j = U^{(j)}_{K_j,l-1}$ and $R_j = U^{(j)}_{K_j,l}$ when $T_j \in (U^{(j)}_{K_j,l-1}, U^{(j)}_{K_j,l}]$ for $l = 1, \ldots, K_j + 1$, with $U^{(j)}_{K_j,0} \equiv 0$ and $U^{(j)}_{K_j,K_j+1} \equiv \infty$.

The following assumptions are imposed on the censoring mechanism. (i) $\{K_j, U_j, j = 1, 2\}$ and $\{\eta, T_j, j = 1, 2\}$ are independent conditioned on $\{Z_j, j = 1, 2\}$. (ii) The conditional distribution of $\{K_j, U_j, j = 1, 2\}$ given $\{Z_j, j = 1, 2\}$ does not depend on parameters of interest. Also, assume that η is independent of (Z_1, Z_2). Then, the likelihood under (2.3) for a single observation $O = \{L_j, R_j, Z_j, j = 1, 2\}$ is

$$\begin{aligned}\mathcal{L}(\theta, \Lambda_1, \Lambda_2)(O) &= E_\eta\Big\{ \prod_{j=1}^{2} \Big[\exp(-\eta e^{\beta_j' Z_j} \Lambda_j(L_j)) - \exp(-\eta e^{\beta_j' Z_j} \Lambda_j(R_j)) \Big] \Big\} \\ &= S(L_1, L_2|Z_1, Z_2) - S(L_1, R_2|Z_1, Z_2) \\ &\quad - S(R_1, L_2|Z_1, Z_2) + S(R_1, R_2|Z_1, Z_2), \end{aligned} \tag{2.4}$$

where $\theta = (\beta_1', \beta_2', \gamma)'$, E_η is the expectation with respect to η, and

$$S(t_1, t_2|Z_1, Z_2) = (1 + \gamma e^{\beta_1' Z_1} \Lambda_1(t_1) + \gamma e^{\beta_2' Z_2} \Lambda_2(t_2))^{-1/\gamma}$$

is the unconditional joint survival function of (T_1, T_2) under (2.3). The likelihood under (2.2) can be obtained by setting $\beta_1 = \beta_2$ in (2.4) and that under model (2.1) can be obtained by further setting $\Lambda_1 = \Lambda_2$.

The identifiability of model (2.3), the most complicated model considered in our setup, is established in Appendix A.2.

3. Semiparametric Maximum Likelihood Estimation

In this section we discuss semiparametric maximum likelihood estimation of $(\theta, \Lambda_1, \Lambda_2)$ under (2.3) with bivariate interval-censored data, where $\theta = (\beta_1', \beta_2', \gamma)'$. Semiparametric maximum likelihood estimation under (2.1) and (2.2) can be analogously obtained with slight modifications. Let the observed data $O_1, \ldots, O_n$ be n i.i.d. copies of O with $O_i = \{L_{j,i}, R_{j,i}, Z_{j,i}, j = 1, 2\}$. The likelihood function of $(\theta, \Lambda_1, \Lambda_2)$ based on $\{O_i, i = 1, \ldots, n\}$ is

$$\mathcal{L}_n(\theta, \Lambda_1, \Lambda_2) = \prod_{i=1}^{n} \mathcal{L}(\theta, \Lambda_1, \Lambda_2)(O_i). \tag{3.1}$$

We now establish the existence of the semiparametric maximum likelihood estimator (SPMLE) that maximizes this likelihood. For every fixed θ, we first show that there exist random elements $\widehat{\Lambda}_{1\theta}$ and $\widehat{\Lambda}_{2\theta}$ in the space of right-continuous non-decreasing functions that maximize $\mathcal{L}_n$. Let $L_{j(n)} = \max_i L_{j,i}$ for $j = 1, 2$. From (3.1) it is clear that $\widehat{\Lambda}_{j\theta}(R_{j,k}) = \infty$ for those $R_{j,k} > L_{j(n)}$, and $\widehat{\Lambda}_{j\theta}(L_{j(n)}) < \infty$ as $\mathcal{L}_n$ would be 0 otherwise. By the fact that $\lim_{y\to\infty} \exp(-Cy) \to 0$ for positive constant C, we have that there exist positive constants M_1 and M_2 such that, for each fixed θ,

$$\max_{\Lambda_1(L_{1(n)})\le M_1, \Lambda_2(L_{2(n)})\le M_2} \mathcal{L}_n(\theta, \Lambda_1, \Lambda_2) > \sup_{\Lambda_1(L_{1(n)})>M_1 \text{ or } \Lambda_2(L_{2(n)})>M_2} \mathcal{L}_n(\theta, \Lambda_1, \Lambda_2).$$

The existence of $(\widehat{\Lambda}_{1\theta}, \widehat{\Lambda}_{2\theta})$ thus follows from the continuity of $\mathcal{L}_n$. Further, due to the compactness of the parameter space Θ of θ, the maximizer, say $\widehat{\theta}$, of the continuous function $\theta \mapsto \mathcal{L}_n(\theta, \widehat{\Lambda}_{1\theta}, \widehat{\Lambda}_{2\theta})$ exists. Let $\widehat{\Lambda}_1 = \widehat{\Lambda}_{1\widehat{\theta}}$ and $\widehat{\Lambda}_2 = \widehat{\Lambda}_{2\widehat{\theta}}$, then $(\widehat{\theta}, \widehat{\Lambda}_1, \widehat{\Lambda}_2)$ maximizes $\mathcal{L}_n(\theta, \Lambda_1, \Lambda_2)$.

Since the likelihood function $\mathcal{L}_n$ depends on the baseline cumulative hazards (Λ_1, Λ_2) only through their values at the examination times $\{L_{j,i}, R_{j,i}; i = 1, \ldots, n, j = 1, 2\}$, it is easy to see that the SPMLE of the baseline cumulative hazards is unique only within the class of all right-continuous non-decreasing step functions with possible jumps only at $\{L_{j,i}, R_{j,i}; i = 1, \ldots, n, j = 1, 2\}$. Hence, we need only restrict the search of SPMLE of Λ_1 and Λ_2 within this class of functions.

Our theorems establish asymptotic properties for the proposed SPMLE $\widehat{\zeta} = (\widehat{\theta}, \widehat{\Lambda}_1, \widehat{\Lambda}_2)$ of $\zeta = (\theta, \Lambda_1, \Lambda_2)$.

Theorem 1 (Consistency and rate of convergence)**.** *Under conditions* (C1)−(C6) *in Appendix* A.1*, the SPMLE* $\widehat{\zeta}$ *is consistent; that is,* $\widehat{\theta} \xrightarrow{P} \theta_0$ *and each* $\widehat{\Lambda}_j(t) \xrightarrow{P} \Lambda_{j0}(t)$ *for every* t *in* (τ_1, τ_2)*. The rate of convergence of SPMLE is of order only* $n^{-1/3}$ *under the metric* d^* *defined in* (A.1)*,* $d^*(\widehat{\zeta}, \zeta_0) = O_p(n^{-1/3})$*.*

Theorem 2 (Asymptotic normality)**.** *Under conditions* (C1)−(C6) *in Appendix* A.1*,* $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, I_0^{-1})$*. The asymptotic variance* I_0^{-1} *is the inverse of the efficient Fisher information matrix* I_0*, whose existence is examined in Appendix* A.4.

In theory, the variance estimation of $\widehat{\theta}$ can be obtained by inverting the observed information matrix. However we note that the information matrix does not have a closed form, which makes this approach to variance estimation difficult to implement. One approach is to numerically approximate the observed information matrix by

$$\widehat{I}_{ij} \equiv -n^{-1}\rho_n^{-2} \Big[\log \widetilde{\mathcal{L}}_n(\widehat{\theta} + \rho_n e_i + \rho_n e_j) - \log \widetilde{\mathcal{L}}_n(\widehat{\theta} + \rho_n e_i)$$

$$- \log \widetilde{\mathcal{L}}_n(\widehat{\theta} + \rho_n e_j) + \log \widetilde{\mathcal{L}}_n(\widehat{\theta}) \Big], \tag{3.2}$$

where $\widetilde{\mathcal{L}}_n(\theta) = \sup_{\Lambda_1,\Lambda_2} \mathcal{L}_n(\theta, \Lambda_1, \Lambda_2)$, i.e., $\widetilde{\mathcal{L}}_n(\hat{\theta}) = \mathcal{L}_n(\hat{\theta}, \hat{\Lambda}_1, \hat{\Lambda}_2)$, e_i is a d-dimensional unit vector with the ith element equal to 1, and ρ_n is a tuning constant with an order of $n^{-1/2}$. This method of approximation was used by Wen and Chen (2011), among others.

Remark 1. Although the overall convergence rate for $(\widehat{\theta}, \widehat{\Lambda}_1, \widehat{\Lambda}_2)$ is $O_p(n^{-1/3})$, the convergence rate for $\widehat{\theta}$ achieves the usual parametric rate $O_p(n^{-1/2})$. This extends the results of Huang and Wellner (1997) from the univariate Cox model to the bivariate gamma-frailty Cox model under interval censoring.

4. Computation Algorithm

In this section, we first state the main idea underlying our computation method for the SPMLE, then describe in detail the algorithm. For simplicity, here we only focus on the computation algorithm for the model (2.3). The algorithms for (2.1) and (2.2) can be obtained in a similar way.

For $j = 1, 2$, let $0 = c_{j,0} < c_{j,1} < \cdots < c_{j,n_j} < c_{j,n_j+1} = \infty$ denote the distinct ordered values of the examination times $\{L_{j,i}, R_{j,i}; i = 1, \ldots, n\}$. As discussed in Section 3, to obtain the SPMLE we can simply consider Λ_j a right-continuous non-decreasing step function with possible jumps only at time points $\{c_{j,1}, \ldots, c_{j,n_j}\}$. In this case Λ_j can be represented by $\Lambda_j(t) = \sum_{l:c_{j,l} \le t} v_{j,l}$, where $v_j = (v_{j,1}, \ldots, v_{j,n_j})'$ is a vector of nonnegative parameters with $v_{j,l}$ representing the jump size of Λ_j at $c_{j,l}$. Then in terms of the parameters (θ, v_1, v_2), the logarithm of the likelihood $\mathcal{L}_n$ can be written as

$$\ell(\theta, v_1, v_2) = \sum_{i=1}^{n} \log \{A_{LL,i} - A_{LR,i} - A_{RL,i} + A_{RR,i}\} (\theta, v_1, v_2),$$

where

$$A_{LL,i}(\theta, v_1, v_2) = \Big(1 + \gamma e^{\beta_1' Z_{1,i}} \Big\{ \sum_{l:c_{1,l} \le L_{1,i}} v_{1,l} \Big\} + \gamma e^{\beta_2' Z_{2,i}} \Big\{ \sum_{l:c_{2,l} \le L_{2,i}} v_{2,l} \Big\} \Big)^{-1/\gamma},$$

$$A_{LR,i}(\theta, v_1, v_2) = \Big(1 + \gamma e^{\beta_1' Z_{1,i}} \Big\{ \sum_{l:c_{1,l} \le L_{1,i}} v_{1,l} \Big\} + \gamma e^{\beta_2' Z_{2,i}} \Big\{ \sum_{l:c_{2,l} \le R_{2,i}} v_{2,l} \Big\} \Big)^{-1/\gamma},$$

$$A_{RL,i}(\theta, v_1, v_2) = \Big(1 + \gamma e^{\beta_1' Z_{1,i}} \Big\{ \sum_{l:c_{1,l} \le R_{1,i}} v_{1,l} \Big\} + \gamma e^{\beta_2' Z_{2,i}} \Big\{ \sum_{l:c_{2,l} \le L_{2,i}} v_{2,l} \Big\} \Big)^{-1/\gamma},$$

$$A_{RR,i}(\theta, v_1, v_2) = \Big(1 + \gamma e^{\beta_1' Z_{1,i}} \Big\{ \sum_{l:c_{1,l} \le R_{1,i}} v_{1,l} \Big\} + \gamma e^{\beta_2' Z_{2,i}} \Big\{ \sum_{l:c_{2,l} \le R_{2,i}} v_{2,l} \Big\} \Big)^{-1/\gamma},$$

with $v_{1,0} = v_{2,0} = 0$, and $v_{1,n_1+1} = v_{2,n_2+1} = \infty$.

For $1 \le k \le n_j, j = 1, 2$, the partial derivative of ℓ with respect to $v_{j,k}$ takes the form

$$\frac{\partial \ell}{\partial v_{j,k}}(\theta, v_1, v_2) = a_{j,k}(\theta, v_1, v_2) - b_{j,k}(\theta, v_1, v_2),$$

where $a_{j,k}(\theta, v_1, v_2)$ and $b_{j,k}(\theta, v_1, v_2)$ are positive functions given by

$$a_{j,k}(\theta, v_1, v_2) = \sum_{i: R_{j,i} \ge c_{j,k}} \frac{e^{\beta_j' Z_{j,i}} \{A_{RL,i}^{1+\gamma} - A_{RR,i}^{1+\gamma}\}(\theta, v_1, v_2)}{\{A_{LL,i} - A_{LR,i} - A_{RL,i} + A_{RR,i}\}(\theta, v_1, v_2)},$$

$$b_{j,k}(\theta, v_1, v_2) = \sum_{i: L_{j,i} \ge c_{j,k}} \frac{e^{\beta_j' Z_{j,i}} \{A_{LL,i}^{1+\gamma} - A_{LR,i}^{1+\gamma}\}(\theta, v_1, v_2)}{\{A_{LL,i} - A_{LR,i} - A_{RL,i} + A_{RR,i}\}(\theta, v_1, v_2)}.$$

A necessary condition for $(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)$ to be the maximizer is $(\partial/\partial v_{j,k})\ell_n(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) = a_{j,k}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) - b_{j,k}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) = 0$, which leads to the "self-consistency" equations

$$\widehat{v}_{j,k} = \widehat{v}_{j,k} \frac{a_{j,k}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) + M_0}{b_{j,k}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) + M_0}, \ k = 1, \ldots, n_j, \ j = 1, 2, \tag{4.1}$$

where $M_0 \ge 0$ is a chosen constant whose rationale will be given later. Let $D = (\{D_{1,1}, \ldots, D_{1,n_1}\}', \{D_{2,1}, \ldots, D_{2,n_2}\}')$, where

$$D_{j,k} \equiv D_{j,k}(\theta, v_1, v_2) = v_{j,k} \frac{a_{j,k}(\theta, v_1, v_2) + M_0}{b_{j,k}(\theta, v_1, v_2) + M_0}, \ k = 1, \ldots, n_j, j = 1, 2.$$

Then, we have $(\widehat{v}_1, \widehat{v}_2) = D(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)$, i.e., $(\widehat{v}_1, \widehat{v}_2)$ is a fixed point of $D(\widehat{\theta}, \cdot, \cdot)$, which motivates the computational approach.

Since the parameter γ is restricted to be positive, for stability in computation we use the reparametrization $\gamma^* = \log \gamma$. With a slight abuse of notation, we denote by θ the parameter set $(\beta_1', \beta_2', \gamma^*)'$. Using the Newton-Raphson method and the self-consistency equations in (4.1), we propose the following procedure to iteratively compute the SPMLE $(\widehat{\theta}, \widehat{\Lambda}_1(t), \widehat{\Lambda}_2(t))$:

Step 1. Choose an initial value $(\theta^{(1)}, v_1^{(1)}, v_2^{(1)}) \in \mathbf{R}^d \times (0, \infty)^{n_1} \times (0, \infty)^{n_2}$.

Step 2. Update each current estimate $(\theta^{(k)}, v_1^{(k)}, v_2^{(k)})$ for $k \ge 1$.

Step 2.1. Update $\theta^{(k)}$ to $\theta^{(k+1)}$ by

$$\theta^{(k+1)} = \theta^{(k)} - \ell_{\theta\theta}^{-1}(\theta^{(k)}, v_1^{(k)}, v_2^{(k)}) \ell_\theta(\theta^{(k)}, v_1^{(k)}, v_2^{(k)}),$$

where ℓ_θ and $\ell_{\theta\theta}$ are the first and second derivatives of ℓ with respect to θ, respectively.

Step 2.2. Update $(v_1^{(k)}, v_2^{(k)})$ to $(v_1^{(k+1)}, v_2^{(k+1)})$ by

$$(v_1^{(k+1)}, v_2^{(k+1)}) = D(\theta^{(k+1)}, v_1^{(k)}, v_2^{(k)}). \tag{4.2}$$

Step 3. If the updated estimate $(\theta^{(k+1)}, v_1^{(k+1)}, v_2^{(k+1)})$ is close to $(\theta^{(k)}, v_1^{(k)}, v_2^{(k)})$, then stop the procedure and let $(\widehat{\theta}, \widehat{\Lambda}_1(t), \widehat{\Lambda}_2(t)) = (\theta^{(k)}, \sum_{l:c_{1,l}\leq t} v_{1,l}^{(k)}, \sum_{l:c_{2,l}\leq t} v_{2,l}^{(k)})$; otherwise, return to Step 2.

In the procedure, Step 2.1 is essentially the one-step Newton-Raphson method for maximizing $\ell(\theta, v_1, v_2)$ with v_1 and v_2 being fixed, and Step 2.2 is based on (4.1).

It is worth noting that the logarithm of the likelihood function $\mathcal{L}_n$ is not concave in Λ, hence the existing algorithms for nonparametric maximum likelihood estimation with interval-censored data, such as the iterative convex minorant (ICM) algorithm and its variants (Groeneboom and Wellner (1992), Huang (1996), Wellner and Zhan (1997)), cannot be applied to the problem we consider.

We now describe the rationale of the self-consistency equations (4.1) leading to the proposed algorithm. From the facts that both $a_{j,k}$ and $b_{j,k}$ are positive functions, and $\partial\ell/\partial v_{j,k} = a_{j,k} - b_{j,k} = 0$, $\partial^2\ell/\partial {v_{j,k}}^2 = \partial a_{j,k}/\partial v_{j,k} - \partial b_{j,k}/\partial v_{j,k} < 0$ at $(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)$, we can choose a constant $M_0 \geq 0$ large enough such that

$$\left|\frac{\partial D_{j,k}}{\partial v_{j,k}}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)\right| = \left|1 + \widehat{v}_{j,k}\frac{\frac{\partial a_{j,k}}{\partial v_{j,k}}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) - \frac{\partial b_{j,k}}{\partial v_{j,k}}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)}{b_{j,k}(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2) + M_0}\right| \in (0,1) \tag{4.3}$$

for all $k = 1, \ldots, n_j$, $j = 1, 2$. By the Mean Value Theorem and the continuity of $\partial D_{j,k}/\partial v_{j,k}$, we know from (4.3) that there exists $0 < b_0 < 1$ such that $|D_{j,k}(\theta, u_1, u_2) - D_{j,k}(\theta, v_1, v_2)| \leq b_0|u_{j,k} - v_{j,k}|$ for $k = 1, \ldots, n_j$, $j = 1, 2$, and $(\theta, u_1, u_2), (\theta, v_1, v_2)$ near $(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)$. If $\tilde{d}\{(u_1, u_2), (v_1, v_2)\} = \sum_{j=1}^{2}\sum_{k=1}^{n_j} |u_{j,k} - v_{j,k}|$, then $\tilde{d}$ is a metric satisfying $\tilde{d}\{D(\theta, u_1, u_2), D(\theta, v_1, v_2)\} \leq b_0\tilde{d}\{(u_1, u_2), (v_1, v_2)\}$ for all (θ, u_1, u_2) and (θ, v_1, v_2) near $(\widehat{\theta}, \widehat{v}_1, \widehat{v}_2)$. This implies that the system of simultaneous equations (4.2) forms a locally contractive iteration, and hence converges by the contraction principle (see, for example, Rudin (1973, p.220).

Although a sufficiently large M_0 may theoretically be needed in order to satisfy the condition (4.3) for local convergence, a large M_0 may adversely slow down the convergence. To solve the dilemma, we may start with $M_0 = 0$ for convenience. If the algorithm has shown a trend towards convergence during early iterations, then we fix $M_0 = 0$ throughout; otherwise, we increase M_0 to a larger value in later iterations to ensure convergence. In all of our simulations and data analysis, we found that the convenience choice of $M_0 = 0$ leads to convergent results, and using a variety of values of M_0 leads to the same convergent solution. Therefore, the specification of M_0 seems not to be a sensitive issue in computation.

The self-consistency equations may have multiple solutions as the information loss due to censoring and missing data (if any) becomes heavier (see Wellner and Zhan (1997) for the examples of this issue with doubly- and interval-censored data). However, our simulation studies have shown that the proposed algorithm is not sensitive to initial values for θ and (v_1, v_2), various different choices of the initial values usually lead to the same solution. Also, multiple initial values can be used to check whether the algorithm is trapped in a local maximum or not.

Remark 2. Note that $\widehat{v}_{j,k}$ is kept at 0 if its initial value is 0. Hence, the proposed algorithm always starts with non-zero initial values for the jump sizes (v_1, v_2) to avoid stopping prematurely at a zero point.

Remark 3. It is not necessary that all the intervals $\{L_{j,i}, R_{j,i}; i = 1, \ldots, n, j = 1, 2\}$ be given mass by the SPMLE. For example, in (2.2) or (2.3) , where separate cumulative hazards are specified for T_1 and T_2, the theory developed for the univariate interval-censored data (e.g., Hudgens (2005)) can be applied to know which intervals contain no mass for Λ_j. For these intervals, the proposed algorithm does result in a value of $\widehat{v}_{j,k}$ very close to 0 ($< 10^{-8}$) in our numerical study under (2.2) or (2.3).

5. Simulation Studies

We report on the assessment of the numerical performances of the proposed SPMLE and the adequacy of the normal approximation. All the computation was done on an ordinary PC with MATLAB. We conducted 400 replications in each setting of each simulation study.

In the first simulation study, two related survival times T_1 and T_2 were simulated from the frailty Cox model (2.2) for each individual, where $\Lambda_1(t) = 0.8t^{0.8}$, $\Lambda_2(t) = 0.8t^{1.2}$, and η followed a gamma distribution with mean 1 and variance γ_0. For each T_j, $j = 1, 2$, $K_j = 2$ examination time points $U_{2,1}^{(j)} < U_{2,2}^{(j)}$ were generated as the order statistics of a random sample of size 2 from $Unif(0, 1.5)$, and the censoring interval $(L_j, R_j]$ was just $(0, U_{2,1}^{(j)}]$ if $T_j < U_{2,1}^{(j)}$, and was set to $(U_{2,2}^{(j)}, \infty)$ if $T_j > U_{2,2}^{(j)}$; otherwise, $(L_j, R_j] = (U_{2,1}^{(j)}, U_{2,2}^{(j)}]$. The covariates Z_j, $j = 1, 2$, were Bernoulli with a success probability of 0.5. The number of subjects was $n = 200$ or 400. Various combinations of values for (β_0, γ_0), with $\beta_0 = 0, 0.5$, or 1, and $\gamma_0 = 0.4$ or 1.2, were considered.

Table 1. Results for the first simulation study under model (2.2). SDp: the average of standard error estimates; CP: the coverage probability of the 95% confidence interval.

			$n = 200$					$n = 400$				
γ_0	β_0	Parameter	Bias	SD	MSE	SDp	CP	Bias	SD	MSE	SDp	CP
0.4	0	β	0.001	0.181	0.033	0.170	93.75	-0.011	0.122	0.015	0.118	94.50
		γ	0.045	0.221	0.051	0.238	97.75	0.008	0.154	0.024	0.157	96.00
	0.5	β	0.022	0.168	0.029	0.171	96.25	-0.003	0.124	0.016	0.117	92.25
		γ	0.039	0.202	0.042	0.220	98.25	0.012	0.142	0.020	0.145	96.25
	1	β	0.049	0.195	0.041	0.192	96.25	0.016	0.132	0.018	0.129	94.00
		γ	0.035	0.193	0.038	0.214	98.00	0.014	0.137	0.019	0.140	96.75
1.2	0	β	0.009	0.212	0.045	0.208	95.00	0.006	0.146	0.021	0.143	95.25
		γ	0.130	0.407	0.183	0.408	96.00	0.044	0.288	0.085	0.268	94.50
	0.5	β	0.038	0.213	0.047	0.211	94.50	0.021	0.144	0.021	0.144	95.00
		γ	0.137	0.386	0.168	0.385	95.50	0.052	0.265	0.073	0.252	94.00
	1	β	0.083	0.239	0.064	0.234	96.00	0.043	0.153	0.025	0.157	96.25
		γ	0.137	0.372	0.157	0.376	96.50	0.061	0.254	0.068	0.245	94.50

In the second simulation study, we considered the frailty Cox model (2.3) under the same scenario as in the first simulation study, except that here Z_1 was Bernoulli with success probability 0.5, and Z_2 was $Unif(-1, 1)$. The covariate effects were given as $\beta_{10} = 0, 0.5$ or 1 and $\beta_{20} = 0.5$. For each simulated sample, we applied the algorithm described in Section 4 and formula (3.2) to calculate the SPMLE $\hat{\theta}$ and the estimated information matrix for variance estimation. The ρ_n in (3.2) was set to $n^{-1/2}$. The algorithm was declared convergent when the change in any parameter estimate at successive iterations was less than 10^{-7}.

In Tables 1 and 2, these results are shown: "Bias", the average of $\widehat{\theta} - \theta_0$ over replications; "SD", the simulation standard deviation of the estimates; "MSE", the simulation mean squared error of the estimates; "SDp", the average of the standard error estimates; "CP", the coverage probability of the 95% confidence intervals obtained by normal approximation. It is seen that the proposed SPMLE is essentially unbiased, and the proposed standard error estimates are quite close to the simulation standard deviations. Also, the coverage probabilities of the 95% confidence intervals match the nominal value well, implying that Theorem 2 works well for the proposed SPMLE in the finite-sample settings considered. The Q-Q plots in Figure 1 also confirm the adequacy of the normal approximation theory, where we depict the standardized SPMLE $(n\widehat{I})^{1/2}(\widehat{\theta} - \theta_0)$ versus the standard normal variate, based on the simulation scenario under model (2.3) with $\beta_{10} = \beta_{20} = 0.5$, $\gamma_0 = 1.2$, and $n = 400$.

In the simulation study under model (2.3) with $(\beta_{10}, \beta_{20}, \gamma_0) = (0.5, 0.5, 1.2)$, the average CPU time per replication (in seconds) for implementing the proposed

Table 2. Results for the second simulation study under model (2.3). SDp: the average of standard error estimates; CP: the coverage probability of the 95% confidence interval.

			$n = 200$					$n = 400$				
γ_0	(β_{10}, β_{20})	Parameter	Bias	SD	MSE	SDp	CP	Bias	SD	MSE	SDp	CP
0.4	$(0, 0.5)$	β_1	-0.006	0.230	0.053	0.238	96.50	-0.009	0.171	0.029	0.166	93.50
		β_2	0.017	0.242	0.059	0.220	92.50	0.025	0.155	0.025	0.152	96.00
		γ	0.032	0.222	0.050	0.238	97.75	0.014	0.151	0.023	0.159	96.25
	$(0.5, 0.5)$	β_1	0.016	0.224	0.051	0.237	97.00	-0.002	0.179	0.032	0.163	92.75
		β_2	0.017	0.241	0.059	0.219	91.50	0.026	0.156	0.025	0.151	95.50
		γ	0.032	0.215	0.047	0.229	98.25	0.011	0.153	0.023	0.152	94.75
	$(1, 0.5)$	β_1	0.043	0.252	0.065	0.257	96.50	0.019	0.183	0.034	0.174	94.75
		β_2	0.020	0.241	0.058	0.220	93.25	0.025	0.155	0.025	0.151	95.25
		γ	0.036	0.219	0.049	0.228	96.50	0.008	0.151	0.023	0.149	95.75
1.2	$(0, 0.5)$	β_1	0.011	0.312	0.098	0.292	94.25	0.010	0.215	0.046	0.202	93.00
		β_2	0.018	0.269	0.073	0.268	95.75	0.027	0.189	0.036	0.183	94.75
		γ	0.121	0.414	0.186	0.410	96.25	0.059	0.288	0.087	0.272	94.75
	$(0.5, 0.5)$	β_1	0.043	0.311	0.099	0.295	95.00	0.024	0.205	0.043	0.202	96.25
		β_2	0.016	0.269	0.073	0.266	95.50	0.027	0.189	0.037	0.182	95.50
		γ	0.131	0.405	0.181	0.400	96.50	0.056	0.265	0.073	0.263	96.00
	$(1, 0.5)$	β_1	0.097	0.337	0.123	0.319	93.75	0.045	0.214	0.048	0.215	95.50
		β_2	0.014	0.272	0.074	0.267	95.25	0.030	0.188	0.036	0.182	94.50
		γ	0.148	0.403	0.184	0.399	96.50	0.068	0.265	0.075	0.261	95.25

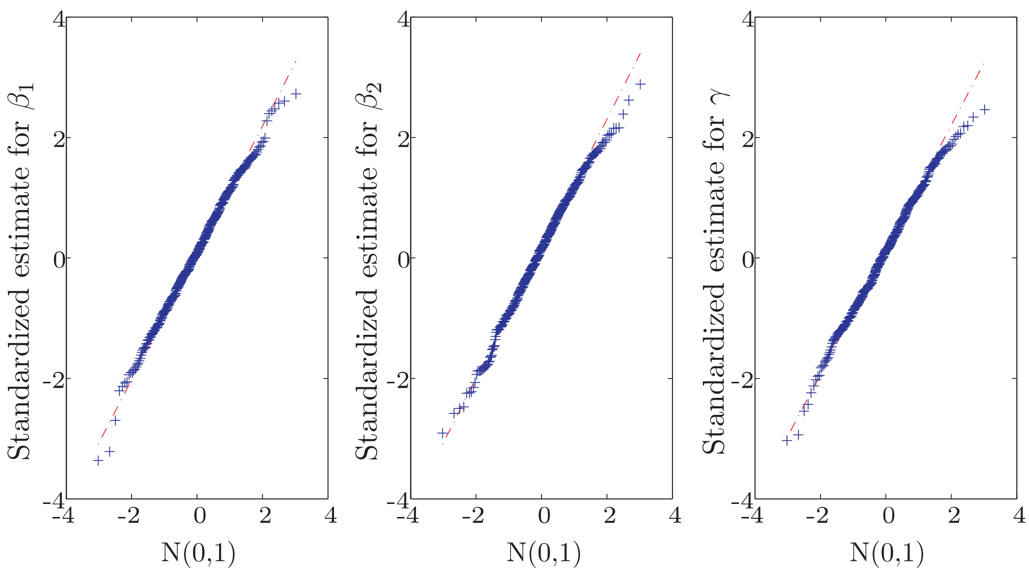

Figure 1. Q-Q plots of standardized estimates versus the standard normal distribution under the simulation scenario for model (2.3) with $\beta_{10} = \beta_{20} = 0.5, \gamma_0 = 1.2$, and $n = 400$.

Table 3. Analysis of CMV shedding data.

Model	Parameter	Estimate	Standard Error	p-value	log likelihood
(2.1)	β	0.8326	0.1851	0.0000	-459.4535
	γ	0.0003	0.2114	-	
(2.2)	β	1.3617	0.2996	0.0000	-397.9190
	γ	1.4597	0.5244	-	
(2.3)	β_1	1.3962	0.4743	0.0032	-397.9144
	β_2	1.3490	0.3272	0.0000	
	γ	1.4559	0.5256	-	

procedure for both point and interval estimation, was 18.3 ($n = 100$), 50.2 ($n = 200$), and 269.3 ($n = 400$). The CPU time for the case of $\gamma_0 = 0.4$ was similar.

6. Data Analysis

We applied the proposed inference procedures to the ACTG 181 data where the effects of baseline CD4 cell counts on the times to shedding of cytomegalovirus (CMV) in the urine and blood are of interest. In this study, the presence of CMV shedding was determined from the urine and blood samples collected at clinical visits for each patient. The sample collection times differed from patient to patient, resulting in "mixed-case" interval censoring for blood and urine CMV shedding times bracketed by the last negative and first positive lab test dates. There also were left- and right-censored CMV shedding times owing to shedding that occurred before the first visit or had not started by the last visit.

We fit the gamma-frailty Cox models (2.1), (2.2), and (2.3) to the CMV shedding time data using the computation algorithm proposed in Section 4. Following Goggins and Finkelstein (2000), the covariate $Z_1 = Z_2$ is binary with a value 1 if the number of baseline CD4 counts is less than 75 cells/μl and with a value 0 otherwise. Results from all the three models, shown in Table 3, imply that the baseline CD4 counts do have a significant effect on the CMV shedding times in either blood and urine; patients with baseline CD4 cell counts below 75 (cells/μl) have significantly higher risk of CMV shedding in blood or urine than those with baseline CD4 cell counts above 75 (cells/μl). The models (2.2) and (2.3) seem to produce remarkably better fit than the model (2.1), suggesting that the CMV shedding times in blood and urine may have different baseline survival functions. This can also be confirmed from Figure 2, where the estimated baseline survival functions for CMV shedding times in blood and urine are depicted. We note from Table 3 that models (2.2) and (2.3) fit the data equally well. Figure 3 shows the estimated marginal survival functions for CMV shedding times in blood and urine under the model (2.2).

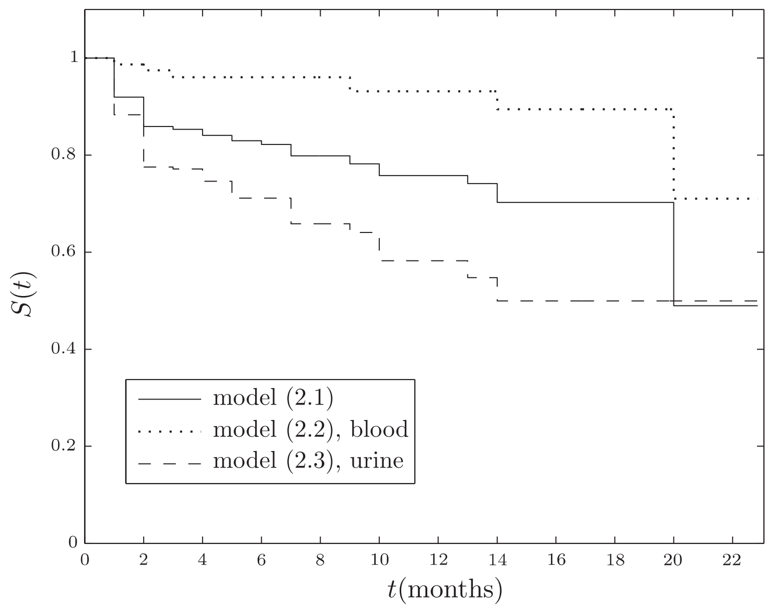

Figure 2. Estimates of marginal baseline survival functions for blood and urine CMV shedding.

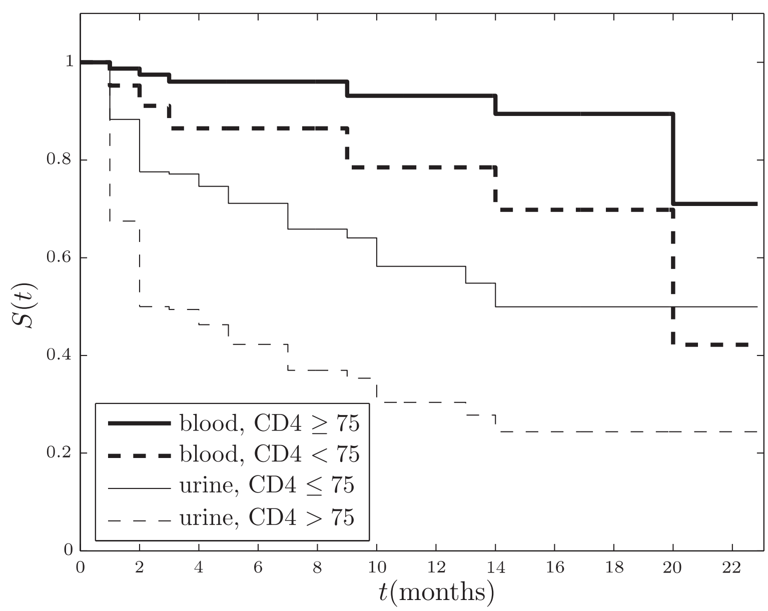

Figure 3. Estimates of marginal survival functions for CMV shedding time data under model (2.2).

7. Concluding Remarks

Interval censoring creates substantial challenges for subsequent statistical analysis in both theory and computation. This is particularly so for bivariate or multivariate event time data. In the literature, analysis procedures for multivariate interval censored data are mainly limited to parametric or non-likelihood based methods; general theories and computation methods for semiparametric inference are still underdeveloped.

In this work, we have developed a semiparametric maximum likelihood inference procedure for bivariate interval-censored data based on the gamma-frailty Cox model. In the proposed procedure, the baseline cumulative hazard functions are estimated as non-decreasing right-continuous step functions, with potential jumps at the examination times bracketing the event times. We do not impose a specific form such as piecewise linear on the baseline cumulative hazard functions. We have established large sample theories for the proposed SPMLE. As in the case of univariate interval censoring, the SPMLE for the regression coefficients can achieve the $n^{1/2}$-rate of convergence, and is asymptotically normal with the asymptotic variance given by the inverse of the efficient Fisher information matrix, while the SPMLE of the cumulative hazard functions can achieve only a $n^{1/3}$-rate of convergence. A computation algorithm utilizing the self-consistency equations and contraction principle is proposed, which provides a stable and efficient tool for implementing the proposed SPMLE. Note that the theories and computation method are proposed under a very general "mixed-case" interval censoring, which includes the "case 1" and "case k" interval censoring as special cases.

The proposed inference framework extends those in Chang, Wen, and Wu (2007) and Wen and Chen (2011) from current status data to general mixed-case interval censoring. The extension involves major difficulties. First, in obtaining asymptotic theories including the consistency, rate of convergence and asymptotic normality, we need to pay attention to the distribution of the random number K_j of the examination times for event T_j $(j = 1, 2)$, as well as the distribution of the random examination times $U_j = \{U_{K_j,l} : l = 1, \ldots, K_j\}$, $j = 1, 2$. We employ a technique of characterizing U_j by a triangular array of random variables, and apply some existing empirical process theories. Second, in computation, although we have applied a computation algorithm similar to that in Wen and Chen (2011), the self-consistency equations involved in the current work are more complicated than those under current status data.

As commented by a referee, the shared frailty model has some limitations. For example, dependence and non-proportionality are confounded in the shared gamma frailty model (Elbers and Ridder (1982)). A natural extension of the shared frailty model to allow separate parameters for the association and non-proportionality is to consider the correlated frailty model proposed by Yashin, Vaupel, and Iachine (1995). In this work, we focus exclusively on the shared gamma frailty model owing to the fact that it has a long tradition in modeling multivariate survival times. In particular, we address the inference and computation of the shared frailty model under general interval censoring that has not been well addressed in the literature, even though the shared gamma frailty has been widely studied under right censoring. In fact, the proposed idea of using self-consistency in computation should work in the correlated gamma frailty model, and our theoretical results for the shared gamma frailty model may be extended to the correlated gamma frailty model after suitable modification. Details for such an extension, however, go beyond the scope of this work, and will be studied in another work.

In addition, it is possible to extend our proposal to multivariate interval-censored data with general frailty distribution. Extensions to regression models more general than the Cox model, such as the semiparametric transformation models (Zeng and Lin (2007)), also deserve further research.

Acknowledgement

We are grateful for the very helpful comments from an associate editor and two referees. This research was supported by grants 100-2118-M-032-011 (CCW) and 98-2118-M-001-016-MY3 (YHC) from the National Science Council of Taiwan.

Appendix

We use the notation $\mathbb{P}_n, P_0$, and P for the expectations taken under the empirical distribution, the true underlying distribution, and a given model, respectively. Let Ω be the class of right-continuous non-decreasing functions that are bounded at $t = \tau$, the termination of the study period. Define the metric d^* on the parameter space $\Theta \times \Omega \times \Omega$ as

$$d^*(\zeta, \tilde{\zeta}) = \{\|\theta - \tilde{\theta}\|^2 + \|\Lambda_1 - \tilde{\Lambda}_1\|_1^2 + \|\Lambda_2 - \tilde{\Lambda}_2\|_2^2\}^{1/2}, \tag{A.1}$$

where $\|\cdot\|$ is the Euclidean norm, $\|\Lambda_j\|_j^2 = \int \sum_{k_j=1}^{\infty} \sum_{l=1}^{k_j} f_{K_j,l}(k_j, u)\Lambda_j^2(u)du$, and $f_{K_j,l}(k_j, u)$ denotes the density of $(K_j, U_{K_j,l}^{(j)})$. For simplicity the proofs are presented under the simpler setting where the distribution of $\{K_j, U_j, j = 1, 2\}$ is independent of $\{Z_j, j = 1, 2\}$, although the proposed method can allow the dependent case.

A.1. Regularity conditions

Throughout the proofs of the proposition and theorems, we require regularity conditions. (C1) The distribution of Z_j, $j=1,2$, is not concentrated on any proper subspace of $\mathbb{R}^{d_j}$ and has a bounded support. (C2) There exists a positive ξ such that $P(U^{(j)}_{K_j,l}-U^{(j)}_{K_j,l-1}\geq\xi)=1$ for $l=2,\ldots,K_j, j=1,2$. (C3) Given $K_j, j=1,2$, each $U^{(j)}_{K_j,l}, l=1,\ldots,K_j$, has a continuous density; the union of the support for conditional distribution $U^{(j)}_{K_j,l}$ given K_j, $l=1,\ldots,K_j, j=1,2$, is an interval $[\tau_1,\tau_2]$ with $0<\tau_1<\tau_2<\infty$. (C4) The true parameter is $\zeta_0=(\theta_0,\Lambda_{10},\Lambda_{20})$, where $\theta_0=(\beta'_{10},\beta'_{20},\gamma_0)'$ is an interior point of its parameter space with dimension d; Λ_{j0} is continuously differentiable and satisfies $M^{-1}<\Lambda_{j0}(\tau_1)<\Lambda_{j0}(\tau_2)<M$. (C5) If $m(\zeta)=\log\mathcal{L}^*(\zeta)$ with $\mathcal{L}^*$ given in (A.2), for any ζ near ζ_0, $P_0(m(\zeta)-m(\zeta_0))\preceq -d^*(\zeta,\zeta_0)^2$, where $\preceq$ means smaller than, up to a constant. (C6) There exist t_1^* and t_2^* in (τ_1,τ_2) for which there are $d+2$ different values of $(\delta_1,\delta_2,z_1,z_2)$ such that if

$$\left(u_1'\frac{\partial}{\partial\theta}+u_2\frac{\partial}{\partial y_1}+u_3\frac{\partial}{\partial y_2}\right)\bigg|_{(\theta_1,y_1,y_2)=(\theta_0,\Lambda_{10}(t_1^*),\Lambda_{20}(t_2^*))}\log(1+\delta_1\gamma e^{\beta_1'z_1}y_1+\delta_2\gamma e^{\beta_2'z_2}y_2)^{-1/\gamma}=0$$

for each of these $d+2$ values, then $u_1=u_2=u_3=0$. Here $(\delta_1,\delta_2)=(1,0),(0,1)$ or $(1,1)$ and z_j is in the support of Z_j.

Remark 4. Conditions (C1)$-$(C5) have been similarly made in the context of univariate interval censoring studies (Huang and Wellner (1997), Zeng, Cai, and Yu Shen (2006), Ma (2010)). In particular, (C2) rules out accurately observed failure times and makes the number of monitoring times K_j bounded. Condition (C6) is also similarly made with multivariate "case 1" interval-censored data under the gamma-frailty model (Chang, Wen, and Wu (2007)), and is needed for both the identifiability of the parameters and the invertibility of the efficient Fisher information.

Remark 5. In practice, (C6) can be verified numerically. We illustrate it by assuming Z_1 and Z_2 are binary and univariate ($d=3$). Let $G:\mathbb{R}^5\mapsto\mathbb{R}^5$ be the function whose components are of the form

$$(\theta,y_1,y_2)\mapsto(1+\delta_1\gamma e^{\beta_1z_1}y_1+\delta_2\gamma e^{\beta_2z_2}y_2)^{-1/\gamma},$$

for $(\delta_1,\delta_2,z_1,z_2)=(1,0,0,0),(1,0,1,0),(0,1,0,0),(0,1,0,1)$, and $(1,1,0,0)$. We can validate (C6) by showing that the Jacobian determinant of G, J_G, at $(\theta_0, \Lambda_{10}(t_1^*),\Lambda_{20}(t_2^*))$ is not zero for some t_1^* and t_2^* in $[\tau_1,\tau_2]$. For example, consider the model with $\theta_0 = (1, 0.5, 1.2)', \Lambda_{10}(t) = 0.8t^{0.8}$, and $\Lambda_{20}(t) = 0.8t^{1.2}$, one of the parameter settings in our simulations. Choosing $t_1^* = t_2^* = 1$ we have $J_G = -0.0094$, showing (C6) is satisfied for this model.

A.2. The identifiability

We discuss the identifiability of the model parameters $\zeta = (\theta, \Lambda_1, \Lambda_2)$ under model (2.3). Rewrite the likelihood (2.4) as

$$\mathcal{L}^*(\zeta) = \sum_{l_1=1}^{K_1+1} \sum_{l_2=1}^{K_2+1} \Delta_{K_1,l_1}^{(1)} \Delta_{K_2,l_2}^{(2)} \{Q_{K_1,K_2,l_1-1,l_2-1} - Q_{K_1,K_2,l_1-1,l_2} \\ -Q_{K_1,K_2,l_1,l_2-1} + Q_{K_1,K_2,l_1,l_2}\}, \tag{A.2}$$

where $\Delta_{K_j,l}^{(j)} = I\{U_{K_j,l-1}^{(j)} < T_j \le U_{K_j,l}^{(j)}\}$ for $l = 1, \ldots, K_j + 1$, $j = 1, 2$, and $Q_{K_1,K_2,s,t} = (1 + \gamma e^{\beta_1' Z_1} \Lambda_1(U_{K_1,s}^{(1)}) + \gamma e^{\beta_2' Z_2} \Lambda_2(U_{K_2,t}^{(2)}))^{-1/\gamma}$.

Suppose $\mathcal{L}^*(\zeta) = \mathcal{L}^*(\zeta_0)$ with probability 1. We first claim that, to establish the identifiability, it suffices to show that $\theta = \theta_0$ and $\Lambda_j(t_j^*) = \Lambda_{j0}(t_j^*)$ for $j = 1, 2$ and some t_1^*, $t_2^* \in [\tau_1, \tau_2]$. To see this, consider $\Delta_{K_1,K_1+1}^{(1)} = \Delta_{K_2,K_2+1}^{(2)} = 1$ with $U_{K_1,K_1}^{(1)} = t_1^*$ or $U_{K_2,K_2}^{(2)} = t_2^*$, so that one of

$$(1 + \gamma e^{\beta_1' Z_1} \Lambda_1(t_1^*) + \gamma e^{\beta_2' Z_2} \Lambda_2(U_{K_2,K_2}^{(2)}))^{-1/\gamma} \\ = (1 + \gamma_0 e^{\beta_{10}' Z_1} \Lambda_{10}(t_1^*) + \gamma_0 e^{\beta_{20}' Z_2} \Lambda_{20}(U_{K_2,K_2}^{(2)}))^{-1/\gamma_0},$$

$$(1 + \gamma e^{\beta_1' Z_1} \Lambda_1(U_{K_1,K_1}^{(1)}) + \gamma e^{\beta_2' Z_2} \Lambda_2(t_2^*))^{-1/\gamma} \\ = (1 + \gamma_0 e^{\beta_{10}' Z_1} \Lambda_{10}(U_{K_1,K_1}^{(1)}) + \gamma_0 e^{\beta_{20}' Z_2} \Lambda_{20}(t_2^*))^{-1/\gamma_0},$$

holds. The claim follows by noting that both sides of the two displays are monotone in $U_{K_2,K_2}^{(2)}$ and $U_{K_1,K_1}^{(1)}$, respectively.

Now examine possible cases in the identity $\mathcal{L}^*(\zeta) = \mathcal{L}^*(\zeta_0)$: $\Delta_{K_1,1}^{(1)} = \Delta_{K_2,K_2+1}^{(2)} = 1$ with $(U_{K_1,1}^{(1)}, U_{K_2,K2}^{(2)}) = (t_1^*, t_2^*)$; $\Delta_{K_1,K_1+1}^{(1)} = \Delta_{K_2,1}^{(2)} = 1$ with $(U_{K_1,K_1}^{(1)}, U_{K_2,1}^{(2)}) = (t_1^*, t_2^*)$; and $\Delta_{K_1,K_1+1}^{(1)} = \Delta_{K_2,K_2+1}^{(2)} = 1$ with $(U_{K_1,K_1}^{(1)}, U_{K_2,K2}^{(2)}) = (t_1^*, t_2^*)$. We thus have

$$(1 + \delta_1 \gamma e^{\beta_1' z_1} \Lambda_1(t_1^*) + \delta_2 \gamma e^{\beta_2' z_2} \Lambda_2(t_2^*))^{-1/\gamma} \\ = (1 + \delta_1 \gamma_0 e^{\beta_{10}' z_1} \Lambda_{10}(t_1^*) + \delta_2 \gamma_0 e^{\beta_{20}' z_2} \Lambda_{20}(t_2^*))^{-1/\gamma_0}$$

for $(\delta_1, \delta_2) = (1, 0), (0, 1), (1, 1)$, and all (z_1, z_2) in the support of (Z_1, Z_2). Taking specifically the $d+2$ different values of $(\delta_1, \delta_2, z_1, z_2)$ in (C6) for the above display, the Inverse Function Theorem and the claim made in the last paragraph imply the identifiability of the model parameters.

A.3. Proof of Theorem 1 (Consistency and rate of convergence)

Consistency. We apply Wald's theorem (van der Vaart (1998, p.48)). Take $w(\zeta) = \log\{[\mathcal{L}^*(\zeta)+\mathcal{L}^*(\zeta_0)]/[2\mathcal{L}^*(\zeta_0)]\}$. By compactness of the parameter sets of θ and Λ_j, $j = 1, 2$, $w(\zeta)$ is uniformly bounded. Also, $\zeta \mapsto w(\zeta)(Y)$ is continuous at ζ, relative to the product of the Euclidean and two weak topologies, for every $Y = \{U^{(j)}_{K_j,l}, \Delta^{(j)}_{K_j,l}, Z_j, l = 1, \ldots, K_j, j = 1, 2\}$ such that $U^{(j)}_{K_j,1}, \ldots, U^{(j)}_{K_j,K_j}$ are continuous points of Λ_j for $j = 1, 2$. In fact, by (C3), $\zeta \mapsto w(\zeta)(Y)$ is continuous at ζ for almost every Y and every given (Λ_1, Λ_2). Because $\widehat{\zeta}$ is the SPMLE, $\mathbb{P}_n w(\widehat{\zeta}) = \mathbb{P}_n\{(1/2)\log(\mathcal{L}^*(\widehat{\zeta})/\mathcal{L}^*(\zeta_0)) + (1/2)\log 1\} \geq 0 = \mathbb{P}_n w(\zeta_0)$. On the other hand, by the concavity of $g(u) \equiv \log((u + 1)/2)$ and Jensen's Inequality, we have $P_0 w(\zeta) = P_0 g\{\mathcal{L}^*(\zeta)/\mathcal{L}^*(\zeta_0)\} \leq g(P_0\{\mathcal{L}^*(\zeta)/\mathcal{L}^*(\zeta_0)\}) = 0$ for any ζ, and equality holds only if $\theta = \theta_0$ and $\Lambda_j = \Lambda_{j0}$ on (τ_1, τ_2) from the identifiability of the parameters. Therefore, it follows directly from Wald's theorem that $\widehat{\theta} \xrightarrow{P} \theta_0$ and $\widehat{\Lambda}_j(t) \xrightarrow{P} \Lambda_{j0}(t)$ for every $\tau_1 < t < \tau_2$ and $j = 1, 2$.

Rate of convergence. Before obtaining the rate of convergence of the SPMLE, we require some definitions from van der Vaart (1998). Given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \leq f \leq u$. An ε-bracket in $L_2(P) = \{f : Pf^2 < \infty\}$ is a bracket $[l, u]$ with $P(u - l)^2 < \varepsilon^2$. For a subclass $\mathcal{C}$ of $L^2(P)$, the bracketing number $N_{[\,]}(\varepsilon, \mathcal{C}, L_2(P))$ is the minimum number of ε-bracket needed to cover $\mathcal{C}$.

With the established consistency, we can restrict θ to $\mathcal{N}_0$, a neighborhood of θ_0, and Λ_j to $\Omega_0 = \{\Lambda \in \Omega | M^{-1} \leq \Lambda(\tau_1) \leq \Lambda(\tau_2) \leq M\}$. Let $\Psi = \{m(\zeta)|\zeta \in \mathcal{N}_0 \times \Omega_0^2\}$, where $m(\zeta) = \log \mathcal{L}^*(\zeta)$. It is easy to see that each element in Ψ is uniformly bounded and satisfies $P_0(m(\zeta)-m(\zeta_0))^2 \preceq d^*(\zeta, \zeta_0)^2$. By Lemma 1 below, the bracketing integral $J_{[\,]}(\delta, \Psi, L_2(P))$, defined as $\int_0^\delta (\log N_{[\,]}(\varepsilon, \Psi, L_2(P)))^{1/2} d\varepsilon$, is of order $O(\delta^{1/2})$. Consequently, Lemma 19.36 of van der Vaart (1998) gives

$$P^* \sup_{d^*(\zeta,\zeta_0)<\delta} |\sqrt{n}(\mathbb{P}_n - P_0)(m(\zeta) - m(\zeta_0))| \preceq \delta^{1/2}(1 + \frac{\delta^{1/2}}{\delta\sqrt{n}}),$$

where P^* is the outer expectation. According to Theorem 3.2.5 of van der Vaart and Wellner (1996), this, together with (C5), implies $d^*(\zeta, \zeta_0) = O_p(n^{-1/3})$.

Lemma A.1. $\log N_{[\,]}(\varepsilon, \Psi, L_2(P)) = O(1/\varepsilon)$.

Proof. First consider the functions in Ψ for a fixed θ. Given the ε-brackets $\Lambda_j^{L_j} \leq \Lambda_j \leq \Lambda_j^{U_j}$, it is easy to get a bracket (m^L, m^U) for $m(\zeta)$ with

$$m^L \equiv \log E_\eta \Big\{ \prod_{j=1}^{2} \Big[\sum_{l=1}^{K_j} \Delta^{(j)}_{K_j,l} \Big\{ \exp[-\eta e^{\beta_j' Z_j} \Lambda_j^{U_j}(U^{(j)}_{K_j,l-1})]$$

$$- \exp[-\eta e^{\beta_j' Z_j} \Lambda_j^{L_j}(U_{K_j,l}^{(j)})]\Big\}\Big]\Big\},$$

$$m^U \equiv \log E_\eta \Big\{ \prod_{j=1}^{2} \Big[\sum_{l=1}^{K_j} \Delta_{K_j,l}^{(j)} \Big\{ \exp[-\eta e^{\beta_j' Z_j} \Lambda_j^{L_j}(U_{K_j,l-1}^{(j)})] - \exp[-\eta e^{\beta_j' Z_j} \Lambda_j^{U_j}(U_{K_j,l}^{(j)})]\Big\}\Big]\Big\}.$$

Due to (C2), we can choose ε small enough that m^L is well-defined. Further, by the Mean Value Theorem, we have

$$|m^L - m^U|^2 \preceq \sum_{j=1}^{2} \sum_{l=1}^{K_j} \Delta_{K_j,l}^{(j)} \{(\Lambda_j^{U_j} - \Lambda_j^{L_j})^2 (U_{K_j,l-1}^{(j)}) + (\Lambda_j^{U_j} - \Lambda_j^{L_j})^2 (U_{K_j,l}^{(j)})\}.$$

Thus brackets for Λ_j of $\|\cdot\|_j$ -size ε can translate into brackets for $m(\zeta)$ of $L_2(P)$-size proportional to ε. By Example 19.11 of van der Vaart (1998), we can cover the set of all Λ_j by $\exp(C/\varepsilon)$ brackets of size ε for some constant C. Allow θ to vary freely as well; because θ is finite-dimensional and $(\partial/\partial\theta)m(\zeta)(Y)$ is uniformly bounded in (ζ, Y), this increases the entropy only slightly. Lemma 1 is thus proved.

A.4. Efficient information

Here we derive the efficient score for θ and establish the invertibility of the efficient Fisher information.

Efficient score. Denote the score function for θ by $m_0(\zeta)$. The score functions for Λ_1 and Λ_2 are

$$m_1(\zeta)[h_1] = e^{\beta_1' Z_1} \Big\{ \sum_{l_1=2}^{K_1+1} \sum_{l_2=1}^{K_2+1} A_{K_1,K_2,l_1,l_2} h_1(U_{K_1,l_1-1}^{(1)}) + \sum_{l_1=1}^{K_1} \sum_{l_2=1}^{K_2+1} B_{K_1,K_2,l_1,l_2} h_1(U_{K_1,l_1}^{(1)}) \Big\},$$

$$m_2(\zeta)[h_2] = e^{\beta_2' Z_2} \Big\{ \sum_{l_1=1}^{K_1+1} \sum_{l_2=2}^{K_2+1} A_{K_1,K_2,l_1,l_2} h_2(U_{K_2,l_2-1}^{(2)}) + \sum_{l_1=1}^{K_1+1} \sum_{l_2=1}^{K_2} B_{K_1,K_2,l_1,l_2} h_2(U_{K_2,l_2}^{(2)}) \Big\},$$

respectively, where

$$A_{K_1,K_2,l_1,l_2} = \Delta_{K_1,l_1}^{(1)} \Delta_{K_2,l_2}^{(2)} \frac{(-Q_{K_1,K_2,l_1-1,l_2-1}^{1+\gamma} + Q_{K_1,K_2,l_1-1,l_2}^{1+\gamma})}{L^*(\zeta)},$$

$$B_{K_1,K_2,l_1,l_2} = \Delta^{(1)}_{K_1,l_1}\Delta^{(2)}_{K_2,l_2}\frac{(-Q^{1+\gamma}_{K_1,K_2,l_1,l_2-1} + Q^{1+\gamma}_{K_1,K_2,l_1,l_2})}{L^*(\zeta)},$$

and h_1 and h_2 are any functions in $L_2(P)$, where $L_2(P) = \{h : \int_{\tau_1}^{\tau_2} h^2(x)dx < \infty\}$.

The efficient score for θ is defined as $m^*(\zeta) = m_0(\zeta) - m_1(\zeta)[\mathbf{h}_1^*] - m_2(\zeta)[\mathbf{h}_2^*]$, where $\mathbf{h}_1^*$ and $\mathbf{h}_2^*$ are d-vector functions satisfying

$$P[(m_0(\zeta) - m_1(\zeta)[\mathbf{h}_1^*] - m_2(\zeta)[\mathbf{h}_2^*])(m_1(\zeta)[h_1] + m_2(\zeta)[h_2])] = 0 \qquad \text{(A.3)}$$

for any h_1 and h_2 in $L_2(P)$. Here and in the sequel it should be understood that all operators on $\mathbf{h}_1^*$ or $\mathbf{h}_2^*$ are applied in a componentwise manner. We now establish the existence of $(\mathbf{h}_1^*, \mathbf{h}_2^*)$.

Consider $h_2 = 0$ in (A.3) to get

$$P(m_0(\zeta)m_1(\zeta)[h_1]) = P((m_1(\zeta)[\mathbf{h}_1^*] + m_2(\zeta)[\mathbf{h}_2^*])m_1(\zeta)[h_1]),$$

which can be written as

$$\begin{aligned}&\int h_1(x)c_1(x)dx\\ &\quad= \int h_1(x)\left[\mathbf{h}_1^*(x)a_1(x) + \int\{b_{11}(x,y)\mathbf{h}_1^*(y) + b_{12}(x,y)\mathbf{h}_2^*(y)\}dy\right]dx,\end{aligned}$$

where

$$\begin{aligned}c_1(x)=&\sum_{k_1=1}^{\infty}\sum_{l_1=2}^{k_1+1} f_{K_1,l_1-1}(k_1,x)E\Big[\sum_{l_2=1}^{K_2+1} m_0(\zeta)e^{\beta_1'Z_1}A_{K_1,K_2,l_1,l_2}|K_1=k_1,U^{(1)}_{K_1,l_1-1}=x\Big]\\ &+\sum_{k_1=1}^{\infty}\sum_{l_1=1}^{k_1} f_{K_1,l_1}(k_1,x)E\Big[\sum_{l_2=1}^{K_2+1} m_0(\zeta)e^{\beta_1'Z_1}B_{K_1,K_2,l_1,l_2}|K_1=k_1,U^{(1)}_{K_1,l_1}=x\Big],\end{aligned}$$

$$\begin{aligned}a_1(x) =& \sum_{k_1=1}^{\infty}\sum_{l_1=2}^{k_1+1} f_{K_1,l_1-1}(k_1,x)E\Big[\sum_{l_2=1}^{K_2+1} e^{2\beta_1'Z_1}A^2_{K_1,K_2,l_1,l_2}|K_1 = k_1, U^{(1)}_{K_1,l_1-1} = x\Big]\\ &+ \sum_{k_1=1}^{\infty}\sum_{l_1=1}^{k_1} f_{K_1,l_1}(k_1,x)E\Big[\sum_{l_2=1}^{K_2+1} e^{2\beta_1'Z_1}B^2_{K_1,K_2,l_1,l_2}|K_1 = k_1, U^{(1)}_{K_1,l_1} = x\Big],\end{aligned}$$

$$\begin{aligned}b_{11}(x,y)=&\sum_{k_1=1}^{\infty}\sum_{l_1=2}^{k_1}\Big\{f_{K_1,l_1-1,l_1}(k_1,x,y)\\ &E\Big[\sum_{l_2=1}^{K_2+1} e^{2\beta_1'Z_1}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|K_1=k_1,U^{(1)}_{K_1,l_1-1}=x,U^{(1)}_{K_1,l_1}=y\Big]\Big\}\end{aligned}$$

$$+\sum_{k_1=1}^{\infty}\sum_{l_1=2}^{k_1}\Big\{f_{K_1,l_1-1,l_1}(k_1,y,x)$$
$$E\Big[\sum_{l_2=1}^{K_2+1}e^{2\beta_1'Z_1}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|K_1=k_1,U^{(1)}_{K_1,l_1-1}=y,U^{(1)}_{K_1,l_1}=x\Big]\Big\},$$

$$b_{12}(x,y)=\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=2}^{k_1+1}\sum_{l_2=2}^{k_2+1}\Big\{f_{K_1,K_2,l_1-1,l_2-1}(k_1,k_2,x,y)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}A^2_{K_1,K_2,l_1,l_2}|K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1-1}=x,U^{(2)}_{K_1,l_2-1}=y\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=2}^{k_1+1}\sum_{l_2=1}^{k_2}\Big\{f_{K_1,K_2,l_1-1,l_2}(k_1,k_2,x,y)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|$$
$$K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1-1}=x,U^{(2)}_{K_1,l_2}=y\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=1}^{k_1}\sum_{l_2=2}^{k_2+1}\Big\{f_{K_1,K_2,l_1,l_2-1}(k_1,k_2,x,y)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}B_{K_1,K_2,l_1,l_2}A_{K_1,K_2,l_1,l_2}|$$
$$K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1}=x,U^{(2)}_{K_1,l_2-1}=y\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=1}^{k_1}\sum_{l_2=1}^{k_2}\Big\{f_{K_1,K_2,l_1,l_2}(k_1,k_2,x,y)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}B^2_{K_1,K_2,l_1,l_2}|K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1}=x,U^{(2)}_{K_1,l_2}=y\Big]\Big\}.$$

Here $f_{K_j,s}, f_{K_j,s,t}$, and $f_{K_1,K_2,s,t}$ denote the densities of $(K_j,U^{(j)}_{K_j,s}), (K_j,U^{(j)}_{K_j,s}, U^{(j)}_{K_j,t})$ and $(K_1,K_2,U^{(1)}_{K_1,s},U^{(2)}_{K_2,t})$, respectively. Therefore

$$c_1(x)=\mathbf{h}_1^*(x)a_1(x)+\int\{b_{11}(x,y)\mathbf{h}_1^*(y)+b_{12}(x,y)\mathbf{h}_2^*(y)\}dy.$$

Similar arguments by considering $h_1=0$ in (A.3) give

$$c_2(x)=\mathbf{h}_2^*(x)a_2(x)+\int\{b_{21}(x,y)\mathbf{h}_1^*(y)+b_{22}(x,y)\mathbf{h}_2^*(y)\}dy,$$

with

$$c_2(x)=\sum_{k_2=1}^{\infty}\sum_{l_2=2}^{k_2+1}f_{K_2,l_2-1}(k_2,x)E\Big[\sum_{l_1=1}^{K_1+1}m_0(\zeta)e^{\beta_2'Z_2}A_{K_1,K_2,l_1,l_2}|K_2=k_2,U^{(2)}_{K_2,l_2-1}=x\Big]$$
$$+\sum_{k_2=1}^{\infty}\sum_{l_2=1}^{k_2}f_{K_2,l_2}(k_2,x)E\Big[\sum_{l_1=1}^{K_1+1}m_0(\zeta)e^{\beta_2'Z_2}B_{K_1,K_2,l_1,l_2}|K_2=k_2,U^{(2)}_{K_2,l_2}=x\Big],$$

$$a_2(x)=\sum_{k_2=1}^{\infty}\sum_{l_2=2}^{k_2+1}f_{K_2,l_2-1}(k_2,x)E\Big[\sum_{l_1=1}^{K_1+1}e^{2\beta_2'Z_2}A^2_{K_1,K_2,l_1,l_2}|K_2=k_2,U^{(2)}_{K_2,l_2-1}=x\Big]$$
$$+\sum_{k_2=1}^{\infty}\sum_{l_2=1}^{k_2}f_{K_2,l_2}(k_2,x)E[\sum_{l_1=1}^{K_1+1}e^{2\beta_2'Z_2}B^2_{K_1,K_2,l_1,l_2}|K_2=k_2,U^{(2)}_{K_2,l_2}=x],$$

$$b_{22}(x,y)=\sum_{k_2=1}^{\infty}\sum_{l_2=2}^{k_2}\Big\{f_{K_2,l_2-1,l_2}(k_2,x,y)E\Big[\sum_{l_1=1}^{K_1+1}e^{2\beta_2'Z_2}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|$$
$$K_2=k_2,U^{(2)}_{K_2,l_2-1}=x,U^{(2)}_{K_2,l_2}=y\Big]\Big\}$$
$$+\sum_{k_2=1}^{\infty}\sum_{l_2=2}^{k_2}\Big\{f_{K_2,l_2-1,l_2}(k_2,y,x)E\Big[\sum_{l_1=1}^{K_1+1}e^{2\beta_2'Z_2}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|$$
$$K_2=k_2,U^{(2)}_{K_2,l_2-1}=y,U^{(2)}_{K_2,l_2}=x\Big]\Big\},$$

$$b_{21}(x,y)=\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=2}^{k_1+1}\sum_{l_2=2}^{k_2+1}\Big\{f_{K_1,K_2,l_1-1,l_2-1}(k_1,k_2,y,x)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}A^2_{K_1,K_2,l_1,l_2}|K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1-1}=y,U^{(2)}_{K_1,l_2-1}=x\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=1}^{k_1}\sum_{l_2=2}^{k_2+1}\Big\{f_{K_1,K_2,l_1,l_2-1}(k_1,k_2,y,x)$$
$$E\Big[e^{\beta_1'Z_1+\beta_2'Z_2}A_{K_1,K_2,l_1,l_2}B_{K_1,K_2,l_1,l_2}|$$
$$K_1=k_1,K_2=k_2,U^{(1)}_{K_1,l_1}=y,U^{(2)}_{K_1,l_2-1}=x\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=2}^{k_1+1}\sum_{l_2=1}^{k_2}\Big\{f_{K_1,K_2,l_1-1,l_2}(k_1,k_2,y,x)$$

$$E\Big[e^{\beta_1' Z_1+\beta_2' Z_2} B_{K_1,K_2,l_1,l_2} A_{K_1,K_2,l_1,l_2}|$$
$$K_1=k_1, K_2=k_2, U^{(1)}_{K_1,l_1-1}=y, U^{(2)}_{K_1,l_2}=x\Big]\Big\}$$
$$+\sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty}\sum_{l_1=1}^{k_1}\sum_{l_2=1}^{k_2}\Big\{f_{K_1,K_2,l_1,l_2}(k_1,k_2,y,x)$$
$$E\Big[e^{\beta_1' Z_1+\beta_2' Z_2} B^2_{K_1,K_2,l_1,l_2}|K_1=k_1, K_2=k_2, U^{(1)}_{K_1,l_1}=y, U^{(2)}_{K_1,l_2}=x\Big]\Big\}.$$

Because $a_j>0$, we can write

$$\mathcal{B}\begin{bmatrix}\mathbf{h}_1^*\\ \mathbf{h}_2^*\end{bmatrix}\equiv\begin{bmatrix}I_{11}+B_{11} & B_{12}\\ B_{21} & I_{22}+B_{22}\end{bmatrix}\begin{bmatrix}\mathbf{h}_1^*\\ \mathbf{h}_2^*\end{bmatrix}=\begin{bmatrix}(\frac{c_1}{a_1})\\ (\frac{c_2}{a_2})\end{bmatrix},$$

where I_{jj} is the identity operator and $B_{ij}(\mathbf{h}_j^*)=\int b_{ij}(x,y)\mathbf{h}_j^*(y)dy/a_j(x)$, $i,j=1,2$. Because each B_{ij} is a compact operator on $L_2(P)$, by Theorem 4.25 in Rudin (1973) it suffices to show the operator $\mathcal{B}$ is one-to-one to establish its invertibility. Now suppose $\mathcal{B}([h_1,h_2]')=0$. By reversing the above derivation with $c_1(x)=c_2(x)=0$, we can obtain $P(m_1(\zeta)[h_1]+m_2(\zeta)[h_2])^2=0$, and hence $m_1(\zeta)[h_1]+m_2(\zeta)[h_2]=0$ a.s. Setting $\Delta^{(1)}_{K_1,K_1+1}=\Delta^{(2)}_{K_2,K_2+1}=1$ in $m_1(\zeta)[h_1]+m_2(\zeta)[h_2]=0$, we conclude that $e^{\beta_1' Z_1}h_1(U^{(1)}_{K_1,K_1})+e^{\beta_2' Z_2}h_2(U^{(2)}_{K_2,K_2})=0$, which implies $h_1=h_2=0$ by the non-degeneracy of Z_1 and Z_2. Therefore $\mathcal{B}$ is invertible and hence the existence of $(\mathbf{h}_1^*,\mathbf{h}_2^*)$ is established.

Invertibility of efficient Fisher information. We now prove that the efficient Fisher information I_0, defined as $P_0(m^*(\zeta_0)m^*(\zeta_0)')$, is positive definite. Let $v\in\mathbb{R}^d$. Since $v'I_0v=P_0(v'm^*(\zeta_0))^2\geq 0$, it suffices to show that $v'I_0v=0$ implies $v=0$.

Suppose $v'I_0v=0$, then $v'm^*(\zeta_0)=0$ a.s. Consider $\Delta^{(1)}_{K_1,1}=\Delta^{(2)}_{K_2,K_2+1}=1$ with $(U^{(1)}_{K_1,1},U^{(2)}_{K_2,K2})=(t_1^*,t_2^*)$, $\Delta^{(1)}_{K_1,K_1+1}=\Delta^{(2)}_{K_2,1}=1$ with $(U^{(1)}_{K_1,K_1},U^{(2)}_{K_2,1})=(t_1^*,t_2^*)$, and $\Delta^{(1)}_{K_1,K_1+1}=\Delta^{(2)}_{K_2,K_2+1}=1$ with $(U^{(1)}_{K_1,K_1},U^{(2)}_{K_2,K2})=(t_1^*,t_2^*)$ in $v'm^*(\zeta_0)=0$. Some algebra concludes that

$$v'\Big(\frac{\partial}{\partial\theta}-\mathbf{h}_1^*(t_1^*)\frac{\partial}{\partial y_1}-\mathbf{h}_2^*(t_2^*)\frac{\partial}{\partial y_2}\Big)\Big|_{(\theta_1,y_1,y_2)=(\theta_0,\Lambda_{10}(t_1^*),\Lambda_{20}(t_2^*))}\log(1+\delta_1\gamma e^{\beta_1' Z_1}y_1+\delta_2\gamma e^{\beta_2' Z_2}y_2)^{-1/\gamma}=0$$

for $(\delta_1,\delta_2)=(1,0),(0,1),(1,1)$ and almost every (Z_1,Z_2). Using condition (C6), we know $v=0$. This completes the proof.

A.5. Asymptotic normality

Define

$$m_{01}(\zeta)[h_1] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_0(\theta, \Lambda_{1\varepsilon}, \Lambda_2),$$
$$m_{02}(\zeta)[h_2] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_0(\theta, \Lambda_1, \Lambda_{2\varepsilon}),$$
$$m_{11}(\zeta)[\tilde{h}_1, h_1] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_1(\theta, \Lambda_{1\varepsilon}, \Lambda_2)[\tilde{h}_1],$$
$$m_{12}(\zeta)[\tilde{h}_1, h_2] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_1(\theta, \Lambda_1, \Lambda_{2\varepsilon})[\tilde{h}_1],$$
$$m_{21}(\zeta)[\tilde{h}_2, h_1] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_2(\theta, \Lambda_{1\varepsilon}, \Lambda_2)[\tilde{h}_2],$$
$$m_{22}(\zeta)[\tilde{h}_2, h_2] = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} m_2(\theta, \Lambda_1, \Lambda_{2\varepsilon})[\tilde{h}_2],$$

where $(\partial/\partial\varepsilon)|_{\varepsilon=0}\Lambda_{j\varepsilon} = h_j, j = 1, 2$. We first verify that

$$\sqrt{n}P_0 m^*(\theta_0, \widehat{\Lambda}_1, \widehat{\Lambda}_2) = o_P(1). \tag{A.4}$$

Apply a Taylor expansion to $m^*(\theta_0, \Lambda_1, \Lambda_2)(Y)$ at the point $(\Lambda_{10}(U^{(1)}_{K_1,1}), \ldots, \Lambda_{10}(U^{(1)}_{K_1,K_1}), \Lambda_{20}(U^{(2)}_{K_2,1}), \ldots, \Lambda_{20}(U^{(2)}_{K_2,K_2}))$ to get

$$\begin{aligned} P_0 m^*(\theta_0, \Lambda_1, \Lambda_2) = {} & P_0 m^*(\zeta_0) + P_0 \{m_{01}(\zeta_0)[\Lambda_1 - \Lambda_{10}] + m_{02}(\zeta_0)[\Lambda_2 - \Lambda_{20}] \\ & - m_{11}(\zeta_0)[\mathbf{h}_1^*, \Lambda_1 - \Lambda_{10}] - m_{12}(\zeta_0)[\mathbf{h}_1^*, \Lambda_2 - \Lambda_{20}] \\ & - m_{21}(\zeta_0)[\mathbf{h}_2^*, \Lambda_1 - \Lambda_{10}] - m_{22}(\zeta_0)[\mathbf{h}_2^*, \Lambda_2 - \Lambda_{20}]\} \\ & + O_p(\sum_{j=1}^{2} \|\Lambda_j - \Lambda_{j0}\|_j^2). \end{aligned}$$

Using the facts that $P_0 m^*(\zeta_0) = 0$, $P(m_0(\zeta)m_j(\zeta)[h_j]) = -P(m_{0j}(\zeta)[h_j])$ for $j = 1, 2$, $P(m_i(\zeta)[\tilde{h}_i]m_j(\zeta)[h_j]) = -P(m_{ij}(\zeta)[\tilde{h}_i, h_j])$ for $i, j = 1, 2$, (A.3), and the rate of convergence of $\widehat{\Lambda}_j$, we have $P_0 m^*(\theta_0, \widehat{\Lambda}_1, \widehat{\Lambda}_2) = O_P(n^{-2/3})$, which implies (A.4).

It is known from Example 19.11 of van der Vaart (1998) that the class of uniformly bounded functions of bounded variations is a Donsker class. Applying Theorem 2.10.6 of van der Vaart and Wellner (1996), it can be verified that $\{m^*(\zeta)|\zeta \in \mathcal{N}_0 \times \Omega_0 \times \Omega_0\}$ is a uniformly bounded Donsker class; the proof is technical and hence omitted here. Combining this with the consistency of $\widehat{\zeta}$ leads to $\sqrt{n}(\mathbb{P}_n - P_0)(m^*(\widehat{\zeta}) - m^*(\zeta_0)) = o_P(1)$. Adding (A.4) to this display and using the fact that $P_0 m^*(\zeta_0) = \mathbb{P}_n m^*(\widehat{\zeta}) = 0$, it is seen that

$$-\sqrt{n}P_0(m^*(\widehat{\zeta}) - m^*(\theta_0, \widehat{\Lambda}_1, \widehat{\Lambda}_2)) = \sqrt{n}\mathbb{P}_n m^*(\zeta_0) + o_P(1).$$

By the Mean Value Theorem, there exists $\tilde{\theta}$ lying between $\widehat{\theta}$ and θ_0 such that

$$-\sqrt{n}P_0(\frac{\partial}{\partial\theta} m^*(\tilde{\theta}, \widehat{\Lambda}_1, \widehat{\Lambda}_2))(\widehat{\theta} - \theta_0) = \sqrt{n}\mathbb{P}_n m^*(\zeta_0) + o_P(1).$$

By the consistency of $\widehat{\zeta}$ and the fact that $P_0[-\frac{\partial}{\partial\theta} m^*(\zeta_0)] = P_0[m^*(\zeta_0)m^*(\zeta_0)'] = I_0$, we have

$$\sqrt{n}(\widehat{\theta} - \theta_0) = I_0^{-1}\sqrt{n}\mathbb{P}_n m^*(\zeta_0) + o_P(1) \xrightarrow{d} N(0, I_0^{-1}).$$

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Chang, I. S., Wen, C. C. and Wu, Y. J. (2007). A profile likelihood theory for the correlated gamma-frailty model with current status family data. *Statist. Sinica* **17**, 1023-1046.

Chen, M., Tong, X. and Sun, J., (2007). The proportional odds model for multivariate interval-censored failure time data. *Statist. Medicine* **26**, 5147-5161.

Chen, M., Tong, X. and Sun, J., (2009). A frailty model approach for regression analysis of multivariate current status data. *Statist. Medicine* **28**, 3424-3436.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard model. *Rev. Econom. Stud.* **49**, 403-409.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845-854.

Goggins, W. B. and Finkelstein, D. M. (2000). A proportional hazards model for multivariate interval-censored failure time data. *Biometrics* **56**, 940-943.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Boston.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, 540-568.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (Edited by D. Lin and T. Fleming), 123-169. Springer-Verlag, New York.

Hudgens, M. G. (2005). On nonparametric maximum likelihood estimation with interval censoring and left truncation. *J. Roy. Statist. Soc. Ser. B* **67**, 573-587.

Kim, M. Y. and Xue, X. (2002). The analysis of multivariate interval-censored survival data. *Statist. Medicine* **21**, 3715-3726.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statist. Sinica* **17**, 1165-1181.

Rudin, W. (1973). *Functional Analysis.* McGraw-Hill, New York.

Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.* **27**, 45-55.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer-Verlag, New York.

Tong, X., Chen, M. and Sun, J. (2008). Regression analysis of multivariate interval-censored failure time data with application to tumorigenicity experiments. *Biometrical J.* **50,** 364-374.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press, New York.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439-54.

Wen, C. C. and Chen, Y. H. (2011). Nonparametric maximum likelihood analysis of clustered current status data with the gamma frailty Cox model. *Comput. Statist. Data Anal.* **55**, 1053-1060.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the semiparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92**, 945-959.

Xiang, L., Ma, X. and Yau, K. W. K. (2011). Mixture cure model with random effects for clustered interval-censored survival data. *Statist. Medicine* **30**, 995-1006.

Yashin, A., Vaupel, J. and Iachine, I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies* **5**, 145-159.

Zeng, D., Cai, J. and Yu Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statist. Sinica* **16**, 287-302.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. Roy. Statist. Soc. Ser. B* **69**, 507-564.

Department of Mathematics, Tamkang University, New Taipei 25137, Taiwan.

E-mail: ccwen@math.tku.edu.tw

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

E-mail: yhchen@stat.sinica.edu.tw

(Received July 2011; accepted December 2011)